Accessing the Impact of Information and Communication Technology (ICT) On the Performance of Banks in Ghana: A Case Study of Ghana Commercial Bank Limited

Abukari Abdul Aziz Danaa
Dept. of Computer Science
Tamale Polytechnic
Tamale, Ghana.

Enoch Dugbanor Sappor
GT Bank
Head Office
Accra, Ghana.

Mumin Diyawu
Dept. of Computer Science
Tamale Polytechnic
Tamale, Ghana.

Abstract: The study sought to examine the impact of Information and Communication Technology (ICT) on the performance of Ghanaian banks with a case study at Ghana Commercial Bank Limited. Information and Communication Technology (ICT) is at the centre of this global change curve. The application of ICT concepts, techniques, policies and implementation strategies to banking services has become a subject of fundamental importance and concerns to all banks and indeed a prerequisite for local and global competitiveness. The main objectives of the study were to assess the impact of ICT on the financial performance of Ghana Commercial Bank Limited (GCB), to find the contribution of ICT based products to customer satisfaction, to examine how ICT has helped reduce operational cost of GCB and to determine the effect of information technology on the liquidity of GCB. Qualitative data was collected using purposive sampling technique. SPSS was used in the descriptive analysis of one-hundred (100) administered interview guide and open-ended questionnaires. The study revealed that ICT Capital has accelerated growth in the bank. Also, ICT facilitated the absorption of high and medium skilled labour. This has a positive effect on the labour output at the bank. The study findings revealed that the bank's use of ICT also enabled the bank to offer a broad variety of services to customers, coordinate branch activities, and meet up with changes in government regulations and policies. The study observed by concluding that the bank's use of ICT has had an impact on the performance of the bank by way of increasing the bank profitability and also reducing the operational cost of the bank.

Keywords: Information; Performance; Financial; Administered; Operational

1. INTRODUCTION

In recent times, Information and Communication Technology (ICT) has in particular brought a complete paradigm shift on bank's customer service delivery and financial performance in the banking industry. In a bid to catch up with global development, improve the quality of customer service delivery, reduce transaction cost and increase shareholders wealth, banks have invested heavily in ICT, and have widely adopted ICT software, hardware and networks for delivering a wide range of value added products and services. The ICT development has a significant effect on development of more flexible and user friendly banking services. Today, ICT has become the heart of banking sector, while banking industry is the heart of every robust economy. If it collapses so will the economy. It is absolutely evident from the current financial downturn in most of the European banks' crises, and in turn. ICT has created a new infrastructure for the world economy to become truly global and also provided the users of new technology a competitive advantage over their rivals. Services like electronic banking and ATMs have become the main technology driven revolution in conducting banking financial transactions. Banks have made huge investments in telecommunication and electronic systems, users have also been validated to accept electronic banking system as useful and easy to use (Adesina and Ayo, 2010). Transactions worth billions of dollars can only take place in seconds in the electronic circuit throughout the globe by pressing a single button. The Ghana economy yearns for serious development. Irrespective of the fact that she is blessed with massive natural resources such as Gold, Oil, Cocoa and Timber, she remains one of the poorest countries in the world [1]. Banks today are looked upon as stated above as vehicle of change for a much needed economic growth and development as it behoves on banks to develop the most effective means of delivering effective, efficient and quality service that will help drive our much awaited economic success. These can be made possible if and only if ICT is properly put to use in the Ghanaian banking sector.

It is believed that the application of ICT in banking has the tendency to stimulate economic growth by bringing borrowers and lenders of financial resource together more efficiently than if they had to relate directly with one another. ICT has been acknowledged as the backbone of banks in the financial sector as it promotes and facilitates the performance (Financial and Operational) of banks in the country. These therefore call for a pre-requisite need to embrace information technology. It is in view of this that the research work attempts to examine the

impact of Information and Communication and Technology (ICT) on the performance (operational and financial) of banks in Ghana.

2. STATEMENT OF THE PROBLEM

Almost all banks worldwide are now using ICT at the various levels of operations. They are also offering ICT based products and services as part of the service delivery to customers. In general, ICT has been integrated into banking systems so as to enhance service delivery, reduce operational cost and improve efficiency in banks. This is expected to have appreciably positive effects on bank's productivity, general operations, market share and services delivery to customers.

Despite the use of information technology in banking sectors today with its numerous objectives, observation has shown that, not all the objectives have been realized and felt by banks and users. However customers have expressed their disgust through customer feedback forums, about the frequent ATM breakdowns, breakdown in network, the inability to access funds at other branches, referrals to branches due to breakdown, identity theft and high charges on visa transactions thereby disadvantaging shareholders, potential investors among other users. In light of this that the researcher deems it fit and appropriate to examine the impact of ICT on the performance (Financial and Operational) of Ghana Commercial Bank Limited.

3. OBJECTIVES OF THE STUDY

The general objective of this study was to examine the impact of Information and Communication and Technology (ICT) on the performance (Financial and Operational) of Ghanaian Banks.

Specifically this study aims;

1. To assess the impact of ICT on the financial performance of Ghana Commercial Bank Limited (GCB).
2. To find the contribution of ICT based products to customer satisfaction.
3. To examine how ICT has helped reduce operational cost of GCB.
4. To determine the effect of information technology on the liquidity of GCB.

4. SIGNIFICANCE OF THE STUDY

The study is significant because the adoption of ICT might be instrumental in helping improve the performance base of the banking industry. This work would be of great assistance to the management of the banks under study and other players in the Ghanaian banking industry to improve upon their ICT deployment to enhance performance.

The extent of utilizing information technology in banks will be measured by the volume of investment in equipment, and the volume of investment in the software and the use of Internet Banking, Phone Internet, the number of ATM, the use of Cyber Branch and Banking via SMS.

It would also contribute to existing literature by identifying the major barriers to the adoption of this innovation on banking operation in Ghana and suggest how to address them.

Finally, the study has served as a source of reference by building upon previous researches and as well as enriching the scientific knowledge on the subject. It will also serve as basis for future research relating to ICT by researchers and students

5. METHODOLOGY

In research there are two basic approaches, these are qualitative and quantitative. The quantitative research approach makes use of statistics and numbers which are mostly presented in figures whiles qualitative approach relies on describing an event with the use of words. A research approach chosen should be done according to the research questions in that particular situation since each approach has its own merit and demerit and how empirical data is collected and analysed [14]. Additionally, the degree of focus on either contemporary or historical event as well as the type of questions asked should be the main basis on which a research approach should be chosen. In conducting this study a comparison of both quantitative and qualitative research approaches was made and the quantitative research approach was used for the study.

5.2 Study Area

The research focused on Ghana Commercial Bank Limited. Ghana Commercial Bank established some decades ago has remained one of the greatest pillars in the banking sector of Ghana.

The bank has a large customer base and branches in the country. The services of the bank continue to be felt in every community in the country. With regard to its large deposit base, there is ever-growing pressure on its service delivery across the country. Thus the need to introduce ICT has become paramount. Ghana Commercial Bank Limited like every responsive organization is dynamic to changing trends in business transactions.

Apart from the traditional bank products such as savings, current accounts, fixed term deposits which were sourced through traditional practices such as going to the bank premises to transact business, technology has changed the practice of this bank significantly. Now Ghana Commercial Bank Limited has products or services such as the Ready Cash, MasterCard, Smart-Pay, Tansflow and SMS banking, and E-Loans, technologically-driven products or services which are affecting the way the bank carries out its business. Personal banking is now mobile, thanks to wireless technology and cellular phones, which were introduced into the system a couple of years ago. An individual can now access his or her account any time anywhere. Investors can now buy and sell stocks and other securities using their mobile phones and Personal Computers without even facing their bankers or brokers.

5.3 Research Design

This study made used of survey method on which data was collected once across the population through sampling. The study used descriptive research methods to assess the impact of Information, Communication and Technology (ICT) on the performance (Operational and Financial) of banks. The rationale behind the use of the descriptive research method helps the researcher to obtain first hand data from respondents in order to describe what exists. This technique gave a valuable insight into the question under study. The research used a case study approach and due to the descriptive nature of the study

qualitative technique was considered. The qualitative research was based on in-depth interview with respondents.

5.4 Population and Sampling

Staffs of Ghana Commercial Bank Limited were used as the population for this study. This research adopted sampling, which according to Newman (2004) is a set of cases a researcher selects from a larger pool and generalizes to the population. Using sampling in this way reduces cost and makes it possible to gather useful information quickly and where the samples are properly selected accurate result are guaranteed. Due to the focus of this study, 100 questionnaires were administered to 100 staffs from five different branches of the bank, a total of 20 respondents from each branch. The selected branches were Head office, Circle branch, Legon branch, Madina branch and Osu branch

It was impractical to collect data on the whole population, considering the size, as well as the time, available to the researcher, hence there was a need to select a sample to represent the whole population. Simple random sampling was used to select the branches at Accra Metropolis that the questionnaires were administered to, by assigning numbers and choosing without replacement. Purposive sampling technique was employed to select the staffs. This category was purposively selected because they are knowledgeable in the study setting and the subject area study. The sample size covered 33% of the target population of the study (thus, 100 out of 300). Thus 100 questionnaires were given out to help sample views of staffs of GCB

5.5 Data Collection Instruments

In order to gain access to contemporary data, a case study was carried out with the help of an interview guide. Thus, the interview guide was employed to collect primary data.
One-on-one in depth-open-ended interview was conducted with all the study participants. This help in generating relevant qualitative data to shed more light on issues relating to the impact of information, communication and technology on the performance of Ghanaian Banks in Ghana already contained in the unstructured interview guide. This technique allowed the study participants to share more freely their experiences in the ICT performance and also offers the advantage of allowing the researcher to uncover previously unforeseen information. This method or instrument also allowed the researcher to probe for clarification and better understanding, especially on issues about paraphrasing, reflection and laddering to explore the issues in depth.

5.6 Data Management and Analysis

Data management is a challenging, integral, and vital part of qualitative research. Data handling and management has been identified as necessary for facilitating the coherence of a research project.
To achieve the research objective, information received through the interviews were first recorded, transcribed before they were typed out.
Besides, the researcher planned from the outset to sort, summarize, analyze, and store project data, including their process of working with the data through the iterative process.

The iterative progress of qualitative research means that data management and data analysis are integral to each other.
For better analysis, the data was arranged under themes which were developed from the literature and the prevailing issues from respondents taken into cognizance the study objectives.

5.7 Ethical Issues

The research contributed to the knowledge of ICT in an organisation and the impact on performance. It maintains utmost confidentiality about the respondents. The research ensured that all respondents were given a free will to participate and contribute voluntarily to the study. In addition, the researcher made sure that necessary research authorities are consulted and permission granted and due explanations given to the respondents before commencement of the study.

6. LITERATURE REVIEW

The literature explains that information and communication technology (ICT) related to banks and expanded to include both optimistic and pessimistic views as well as different opinions of changes and forces driving them. The literature shows that ICT generates many benefits and is a key to develop the concern employees in organizations (Venkatesh & Davis, 2000). There is a significant resistance to accepting and using computer resources in banks [3]. A lot of banks experience struggle and even more failure in moving ICT into practice even with spending billions of dollars (Hill et. al, 2002). The problems appear to be worst in developing countries [10]. The aim of every bank is by using information and communication technology is to improve employee performance, and this performance efficiency is only achieved when ICT is installed correctly. The slow ICT diffusion in banks could be from the poor infrastructure, sociopolitical, language barriers, cultural, economic risks and conflicts. Many banks put a hope by using ICT. And the challenges of ICT in these banks are by no means identical to the ones in the other institutions. And the challenges faced by banks in exploitation the full potential of ICT don't appear many differences from those of that are faced by the other institutions.
Today ICT is the best choice in all the banks to upgrade their performance and come to be competitive in the global market place [2]. The ICT based economies has simplified the most complex economies of the world and improved the productivity to the level where some banks have made its way out of the entire banking industry deficit and turned into a surplus in recent years [5]. The world economy now has moved from low-value basic industries to a fast paced high-value information based economy [12].

6.1 ICT Concepts and Definitions

Information Technology (IT) is the automation of processes, controls, and information production using computers, telecommunications, software and ancillary equipment such as automated teller machine and debit cards [6]. It is a term that generally covers the harnessing of electronic technology for the information needs of a business at all levels. Communication Technology deals with the Physical devices and software that link various computer hardware components and transfer data from one physical location to another [8]. ICT products in use in the banking industry include Automated Teller Machine, Smart Cards, Telephone Banking, MICR, Electronic Funds

Transfer, Electronic Data Interchange, Electronic Home and Office Banking. It is not easy to find a definition that clearly describes the difference between IT and ICT. In short, the definitions of both IT and ICT talk about hardware, software and networks by which information is retrieved, recorded and displayed. The World Bank, which is working on providing help to the developing countries with funds and proficiency needed to exploiting ICT, defines IT and ICT as following:

Information technology *"Information Technology refers to the creation, storage and processing of data, including hardware (computer networks, servers, storage devices, and desktop computers), system software (operating systems, middleware, programming languages), and software applications"* World Bank (2002, p.3).
Another of World Bank's definitions says that: *"Information Technology is the use of technology, such as computing and software, in support of information processing "World* Bank (2005, p.6).
ICT consists of all technical means used to handle information and aid communication, including computer and network hardware as well as necessary software. In other words, ICT consists of IT as well as telephony, broadcast media, and all types of audio and video processing and transmission. The expression was first used in 1997 in a report by Dennis Stevenson to the UK government and promoted by the new National Curriculum documents for the UK in 2000. *ICT* is often used in the context of "ICT roadmap" to indicate the path that an organization will take with their ICT needs. The term *ICT* is now also used to refer to the merging (convergence) of telephone networks with computer networks through a single cabling or link system. There are large economic incentives (huge cost savings due to elimination of the telephone network) to merge the telephone network with the computer network system.

6.2 Commercial Application of ICT in the Banking Industry

The rapid advancement in ICT has had a profound effect on the banking industry and the wider financial sector over the last 2 decades. ICT is now a tool that facilitates the bank's organizational structures, business strategies and customer services [14]. ICT enables the development of sophisticated products, better market infrastructure, and implementation of reliable techniques for control of risks and helps the banks to reach geographically distant and diversified markets.
Today, the effective use of ICT is assisting bank to be more customer centred in their operations by building a solid foundation for customer relationship management. The system also helps to grow a range of products or services while mitigating fraud levels and improving risk management, reduce transaction and operational cost and to remain competitive. In sum, technology has changed the contours of the three major functions performed by banks: access to liquidity, transformation of assets and monitoring risks [7].

6.3 Forms of Information Technology Innovation (Electronic Delivery Channels)

Technological innovation has been identified to contribute to the distribution channels to banks. The electronic delivery channels are mostly referred to as electronic banking. Electronic banking allows customer to have easy access to their money and also control their financial business anywhere at any time through automated teller machine (ATM), or direct deposit of pay-cheques into checking or savings accounts. Electronic banking also known as electronic funds transfer (EFT), uses computer and electronic technology as a substitutes for cheques and other paper transactions. Many banks and financial institutions use ATM or debit cards and personal identification number (PIN) for this purpose. The various electronic delivery channels are discussed below.

6.3.1 Automated Teller Machine (ATM)

Automated teller machine is an electronic terminal that let customer banks almost anytime. It's a computerized telecommunication device that provides the customer of a bank or financial institutions with access to financial transaction in public space without the need of bank teller. On modern, automated teller machines (ATM) the customers are identified by inserting a plastic ATM cards with magnetic stripes or a plastic smartcard with a chip, that contains a unique card number and some security information such as expiring date, or CVV. Security is provided by customer entering a personal identification number (PIN). To withdraw cash, make deposit and transfer funds between accounts, customers generally insert an ATM card and enter their personal Identification number (PIN).
Some financial institutions, banks and ATM owners charge a fee, particularly to customers who don't have accounts with them. ATM tells the customers the charge fee and its amount on the screen before transactions are completed. In Ghana the Trust Bank is the first bank to installed ATM in the country in 1995. Ghana Commercial Bank started its ATM offering in 2001 in collaborations with Agricultural Development Bank. Not long after, most of the major banks began their ATM networks at competitive position. Due to technological advancement in technology, ATM are able to provide a wide range of services such as cash withdrawal, change pin, account balance enquiry, funds transfer, print mini statement, request for bank statement and cheque book, make a cheque request, and pay utility bills.

6.3.2 Telephone Banking

Telephone Banking is a service provided by banks, which allows its customers to transact banking business over the telephone. Telephone Banking is a very convenient, fast and easy method of gaining access to your bank account at any time from any location, seven days a week, and twenty four hours a day using any touch tone telephone. Most telephone banking uses an automated phone answering system with phone keypad response or voice recognition capability. For security reasons, the customer must first authenticate through a numeric or verbal password or through a security question asked by live customer care personal. Telephone banking has numerous benefits for both banks and their customers [9]. On the part of the customers, it provides increased convenience, expanded access and significant time saving. With respect to the banks, the cost of delivering telephone-based service is cost saving than those of branch based services.
With telephone banking, customers cannot make cash withdrawals and deposit but other transactions like account balance information, electronic bill payment, funds transfers between a customer's accounts. Its saves customers from number of hours spent in the banking hall, making the banks to give good quality of services to their customers and making time to attend to other issues. Telephone banking offers retail

banking service to customers at their office/home as an alternative to going to the banking hall or to the ATM for transactions.

6.3.3 Internet Banking

The idea of internet banking was to give customers access to their personal account online anytime, anywhere. It's the perfect way for customers to do their banking transaction whether from home, the office or when travelling because the service is available 24 hours a day seven days a week. Internet banking gives the customers the opportunity to choose their own banking hours giving them greater control of their finances. It's secure, fast and easy to use. Customers have access to their latest balance, statement, view account details, customise, print, download statement and obtain history of all their account.

Internet banking gives customers the chance to check their balance and transaction details, transfer money to business partners, family and friends, pay utility bills, enquire of daily forex rate, request for cheque book and statement of account from the comfort of their home or offices without having to visit the bank. With internet banking customers need to log in with their log in ID and password in order to access their account. Banks like Barclays Ghana, Standard Charted Bank Ghana, Ecobank Ghana, Prudential Bank and other banks have all launched their internet banking services.

6.3.4 Personal Computer Banking

A personal Computing Banking let customers handles many banks transactions via their personal computers, with the help of proprietary software installed on their computers. Once access is gained, the customer can perform a lot of retail banking services. The increasing awareness of the importance of computer literacy has resulted in increasing the use of personal computers. This certainly supports the growth of personal computer banking which virtually establishes a branch in the customer's home or office and offers 24 hours service, seven days a week.

6.3.4 Branch Networking

Networking of branches is the computerization and inter-connection of geographically scattered stand-alone bank branches, into one unified system in the form of Wide Area Network (WAN) or Enterprise Network (EN), for the creating and sharing of customer information and records. It offers quicker rate of inter-branch transactions as the consequence of distance and time are eliminated. Also, with the several networked branches serving the customer populace as one system, there is simulated division of labour among banks branches with its associated positive impact on productivity among the branches. With this system customers can transact banking business in any branch of their without travelling to their bank where they have their account.

6.4 ICT and Banks Competitiveness

Flexibility is considered to be a major source of competitiveness for banks compared to larger enterprises. The use of ICT could now on the one hand increase the competitiveness of banks as they enable the creation of more flexible links with trading partners because of faster and more reliable communication channels. On the other hand ICTs could help banks to increase their flexibility through a restructuring of the banks which will enable them to adapt quicker to changing conditions. Therefore the competitive advantage of banks could also decline. In general banks rely much more on formal information systems than smaller enterprises. To get the relevant information that is needed for a rational decision is costly especially as in banks usually there are major decision makers. External transaction costs are associated with the initiation, negotiation and enforcement of contracts. Especially the internet helps to screen the banks environment for relevant information and thereby get information about its customers that were previously out of reach [11]. However for the actual delivery of service and the transmission of payments also other infrastructure like transport and a reliable banking system has to be in place. However, for bigger banks that are growth oriented, belong to the formal sector etc. information becomes more important and therefore more advanced ICTs can be helpful for building business linkages. The survey banks in Botswana revealed the biggest information gap in market information pertaining to new customers and the need to expand into other markets. Information is also lacking about external finance and sources of skills and training. This lack of information was found to raise costs and reduce income. "ICTs can reduce time and money costs of business processes and can improve the certainty and quality of those processes." These benefits occur mainly in banks with bigger size (with annual turnover of a few tens of thousands of US$) where the Internet can be used as a marketing tool. However for 90 % of the survey banks lack of finance and skills are the main constraints and they cannot afford to buy a computer or make efficient use of it in the short or even medium term [4].

7 ANALYSIS OF FINDINGS, RESULTS AND DISCUSSIONS

This section contains the analysis and report of the responses ascertained from the questionnaires distributed on the subject of "The impact of information, communication and technology on the performance of Ghanaian Banks". The analysis was done in five (5) parts; the first part contained the demographic characteristics of the respondents who formed the sample population of the project, the other four parts were on four objectives of the study. The study was organized under themes to coincide with the study objectives. The presentation was analyzed with the results obtained from the data gathered from interviews held with one-hundred (100) staff at all levels at Gh.ana Commercial Bank Limited.

7.1 Demographic Characteristics of the Respondents

The study first presents the distribution of the respondents. A total of one-hundred (100) questionnaires were distributed and

administered among the staffs of the bank. The study revealed that the majority of the respondents 55% were males and the rest 45% were females. The study age division results revealed that staff 18% was aged from 18-30 years. In addition, the study aged from 31-40 shows 45%, 41-50 depicts 25% with the remaining 7% and 5% were aged 51-60 years and 61 years above respectively.

The study revealed that the majority of the staff representing 60% of the staff had First Degrees. In addition, the Masters Degree represents 15% as their educational background and HND holders representing 10% of the staff while PhD and other professional certificate represented 15%. The findings revealed that majority of the staff have at least First Degrees, which indicated that the staffs are well educated and knowledgeable in the field of Information, Communication and Technology (ICT).

The study shows the number years the respondents has been working with Ghana Commercial Bank Limited. The study shows that 6% of the staff of the sampled population had worked for the institution for less than a year. In addition, 10% of the staff of the sampled population has worked for the institution between 1-5 years. Furthermore, 30% of the staff had worked with the bank for 5-10 years. Moreover, 40% of the sampled population had worked for the bank between 10-15 years and lastly, 15% of the sampled population are over 15 years and over. The above findings illustrates that majority of the employees are very experienced because they had worked at the Department for a very long time.

7.2 The Meaning of Information and Communication Technology (ICT)

The study respondents were asked to explain the meaning of information and communication technology in their own views. The study findings indicated a total expression of one- hundred 100 respondents of which most of them shared a common view on the meaning of information and communication technology. The study shows that the majority of the respondents which indicated a percentage of 40% shared their views on ICT as an electronic means of capturing, processing, storing and communicating information. Secondly, the respondents shared other views on ICT as a form of tool for information collection storage and management to facilitate communication and decision making process with a maximum percentage of 30%. Moreover, the respondents did express the meaning of information and communication technology as the use of technology such as computer and software in support of information processing with a maximum percentage of 15%. Furthermore, the respondents shared their view on the meaning of ICT as automation of processes, controls and information production using computer telecommunication software with a percentage of 10%. Lastly, the respondents expressed their view on the meaning of ICT as the means by which technology helps in improving the efficiency and effectiveness of services offered to customers, improve business process as well as enhancing managerial decisions making and workgroup collaboration with a percentage of 5%. The study findings revealed that information and communication technology is expressed as the process of using technology in a diverse ways to help in facilitating the activities of the organisation through technological means in order to communicate information across the organisation.

7.3 Expression on the importance of Information and Communication Technology

The study made way for the respondents to share and express their opinion on the importance of information and communication technology (ICT) to the performance of Ghana Commercial Bank Limited (GCB), the study revealed that 93% of the respondents were affirmative whereas 7% stated that they do not know the importance of information and communication technology to the performance of the bank under study. The study shows that information and communication technology was indicated by the respondents as important to the performance of the bank, even though 7% indicated that it does not.

7.4 The Importance of Information and Communication Technology (ICT) to the Performance of Ghana Commercial Bank Limited

The study findings concerning the importance of information and communication technology to the performance of Ghana Commercial Bank Limited revealed that ICT helps to increase productivity with a percentage of 35%. This is followed by simplifying and making work easier at the bank indicated by the respondents as one of the importance of ICT to the performance of the organisation with a percentage of 30%.The next of importance was the enhancement in product quantity, quality and speed of work was indicated by the respondents with a percentage of 20%. Lastly, the least of the importance of information and communication and technology was to facilitate the acquisition of realistic plans in the organisation with a maximum percentage of 15%. The study findings revealed that ICT has been a great importance to the performance of Ghana Commercial Bank Limited because through the use of ICT, productivity of the bank has been improved since its facilitate and make the banking activities very easy and simple to operate.

7.5 Expression from respondents concerning impact of ICT on the financial performance of the Bank

The study findings concerning the impact of information and communication technology on the financial performance of the bank indicated by the respondents shows that more than two-third of the sample population representing 97% indicated that information and communication technology has an impact on the financial performance of the Ghana Commercial Bank Limited, but the remaining 3% indicated that the introduction of information and communication technology does not have any impact on the financial performance of the Ghana Commercial Bank Limited. The study revealed that majority of the respondents expressed that ICT has an impact on the financial performance of the bank and this implies that ICT has been of great benefit to the staffs and the bank as a whole through their job activities at the bank.

7.6 The Impact of Information and Communication Technology (ICT) on the Financial Performance of Ghana Commercial Bank Limited

The study findings concerning the impact of information and communication technology on the financial performance of Ghana Commercial Bank Limited of which most of the respondents shared a common view on the subject. The study revealed that the majority of the respondents indicated an increase in productivity representing 30%. This next of ICT impact on financial performance was innovation dynamics with a maximum percentage of 27%. This is followed by increases in market structure with an average percentage of 24% indicated by the respondents. Furthermore, supply chain was one of the impact of ICT on bank performance were shared by the respondents with a total percentage of 10%. Lastly, the respondents shared opinion on increases in fraudulent activities as one of views the respondents indicated as an impact of ICT on bank performance.

The study findings revealed that the use of information and communication technology (ICT) has had an impact on the financial performance of the Ghana Commercial Bank Limited (GCB) by way of increasing productivity and this has had an effect on labor productivity and total factor productivity of GCB. The study shows that the banking industry of GCB has been enjoying the largest productivity growth effect of ICT by way through which ICT has helped the cahiers of the bank to be more productive in their duties and banking operations which turns to speed and brings efficiency in the job, however, increasing the turn-around time at the banking the hall which increases savings, hence profitability increases. The study shows that the bank used of computers and peripherals simplifies the task of getting customers' data and counting money to effect transaction. This enables a single cashier to serve thousands of customers in a day which would have cost the bank enormous staff strength and large building operating cost.

In addition, the study revealed that the bank used of ICT has had a positive impact on the financial performance of Ghana Commercial Bank Limited through technological innovative products and process such as SMS alert, ATM, mobile banking and electronic banking. The study shows that the bank used of technological change had a massive diffusion of ICT innovation strategies at the bank which helps in easy mobilization and transfer of funds thereby increasing the deposit mobilization hence, improving the financial performance of the GCB. This implies that the bank did frequently introduced products or production processes based on a new technology which did drive the bank into new market in the banking industry. The bank ICT innovation has helped in to reduce the cost of banking which made the process of transaction easier and more convenient. The study shows that through the used of ICT, the bank made its customers to enjoy the fastest information updates about the bank and their daily transactions through SMS and email alerts which enhance customer's loyalty and confidence and also the study shows that the bank encouraged the use of other ICT media such as the internet and POS which enhances cashless banking which proved productive in the sales of the bank hence, improving the financial performance of the bank.

Moreover, the study shows the extent to which Information Technology (IT) had improved supply chain agility as measured by the ability to sense and respond to market changes, and the impact the use of (IT) through supply chain agility has had on the bank performance. The study revealed that the bank used of Information Technology (IT) increases the supply chains ability to sense market changes by improving information quality in terms of adequacy, accuracy, accessibility, and timeliness. The study also shows that IT has increased the supply chains ability to respond to market changes by developing and executing a coordinated plan with the supply chain in terms of cost, quality, and timeliness. This implies that the supply chain agility has improved the bank performance in terms of sales, market share, profitability, speed to market, and customer satisfaction.

Lastly, the study finding shows that the use of Information and Communication Technology (ICT) by the bank paved way for fraudulent activities at the bank. The study revealed information technology made it possible for fraudsters to hack into the system of the bank by way of having access to the information of the bank clients and customers of which these fraudsters managed to freeze and defaulted some of the customer's accounts. This implies that instead of the use of the information and communication technology having a positive impact on the bank financial performance, it slightly had a negative effect on the bank performance through the activities of fraudsters taking advantage of the weakness in the system.

7.7 The contribution of ICT based products to customer satisfaction

The study showed the contribution of information and communication technology based products to customer satisfaction. The study revealed that 45% of the respondents indicated branch network as one of the ICT based product. Secondly, 25% of the respondents indicated the use of Automated Teller Machine (ATM) as part of the ICT based product used by the bank to satisfy its customers. Moreover, 18% of the respondents indicated electronic banking as part of ICT based product offered by the bank to its clients and lastly, 12% of the respondents indicated that SMS Alert as part of the ICT products used by the bank.

The study findings shows that the management and the staff of Ghana Commercial Bank Limited had made it possible for the customers of the bank to experience the quality of service through ICT based products offered to its numerous customers across the country of which the products has been a key strategic component to the bank customer satisfaction.

The study finding shows that the bank branch network has been one of the powerful tools used by the bank through the information based technology. This implies that the customers of bank under study still finds it useful to visit their bank branches regularly every month to transact some banking business such as detailed bank statement requests, loan application, foreign funds transfer, deposits etc. This signified the importance of the customer's satisfaction that they have been experiencing by way of having access to their funds anywhere at the branch and it saved them time and cost visiting their original branches.

The study findings revealed that the bank innovative used of Automated Teller Machine (ATM) has helped in the contribution of reducing the bank customers frequent visits to the banking hall and also with the fact that the study shows that ATM reduced the time involved in the bank transactions, reduced long queues in the banking hall and enhanced access to funds anywhere the ATMs were located.

The study shows that the importance of ICT based products service delivery and its impact on improving customer

satisfaction and retention of customers had improved the bank sales and market share, and also improved the corporate image of the bank. The study revealed that ICT based products has an appreciable positive effects on customer service. This affects the growth of the bank positively because customers can now collect money from any branch of the bank. Also, customers do not need to move about with large sums of money, and customers are being attended to within a short period of time. New technologies have created unparalleled wired economy and that the transfer of money from point 'A' to point 'B' has resulted in turning the actual money into bits and bytes through satellite transponders, fibre optic cables or regular telephone [13].

7.8 Contribution of ICT to the reduction of operational cost of the bank

According to the respondent's opinion concerning Information and Communication Technology (ICT) and contribution to the reduction of operational cost at Ghana Commercial Bank Limited. The study shows that more than half representing 87% indicated that ICT had contributed to the reduction of operational cost of the bank, but the remaining 13% pointed out that the reduction of operational cost at the bank does not necessary involves the bank application of information and communication technology.

7.9 Expression from respondents concerning the effect of ICT on the liquidity of the Bank

According to the respondent's opinion concerning the effect of Information and Communication Technology (ICT) on the liquidity of Ghana Commercial Bank Limited. The study findings indicated that more than half representing 96% shared an opinion that ICT has an effect on the liquidity of Ghana Commercial Bank Limited, but the remaining 4% indicated that Information and Communication Technology (ICT) does not have any effect on the liquidity of the bank.

7.10 Expression on satisfaction of the respondent's concerning the impact of ICT on the performance of Ghana Commercial Bank Limited

The study sought the expressions concerning the satisfaction of the respondents concerning the impact of Information and Communication Technology (ICT) on the performance of Ghana Commercial Bank limited. The study finding shows that the respondents shared a common view that the Information and Communication Technology (ICT) has had positive impact on the performance of the bank with majority of the respondents thus, 82% were fully satisfied with the impact of ICT on the performance of the bank and 18% were dissatisfied.

8. SUMMARY OF THE MAIN FINDINGS

The study findings revealed that the use of information and communication technology (ICT) has had an impact on the financial performance of the Ghana Commercial Bank Limited (GCB) by way of increasing productivity and this has had an effect on labour productivity and total factor productivity of GCB.

The study revealed that the bank used of ICT has had a positive impact on the financial performance of Ghana Commercial Bank Limited through technological innovative products and process such as SMS alert, ATM, mobile banking and electronic banking.

The study shows that the bank used of technological change had a massive diffusion of ICT innovation strategies at the bank which helps in easy mobilization and transfer of funds thereby increasing the deposit mobilization hence, improving the financial performance of the GCB. This implies that the bank did frequently introduced products or production processes based on a new technology which did drive the bank into new market in the banking industry.

The study revealed that information technology made it possible for fraudsters to hack into the system of the bank by way of having access to the information of the bank clients and customers of which these fraudsters managed to freeze and defaulted some of the customer's accounts. This implies that instead of the use of the information and communication technology having a positive impact on the bank financial performance, it slightly had a negative effect on the bank performance through the activities of fraudsters taking advantage of the weakness in the system.

The study finding shows that the bank branch network has been one of the powerful tools used by the bank through the information based technology. This implies that the customers of bank under study still finds it useful to visit their bank branches regularly every month to transact some banking business such as detailed bank statement requests, loan application, foreign funds transfer, deposits etc. This signified the importance of the customer's satisfaction that they have been experiencing by way of having access to their funds anywhere at the branch.

The study findings revealed that the bank innovative used of Automated Teller Machine (ATM) has helped in the contribution of reducing the bank customers frequent visits to the banking hall and also with the fact that the study shows that ATM reduced the time involved in the bank transactions, reduced long queues in the banking hall and enhanced access to funds anywhere the ATMs were located.

The study revealed that ICT based products has an appreciable positive effects on customer service. This affects the growth of the bank positively because customers can now collect money from any branch of the bank. Also, customers do not need to move about with large sums of money, and customers are being attended to within a short period of time. The study shows that the bank ICT based products had hold the promise of a new commercial revolution by offering an inexpensive and direct way for its customers to exchange information and to sell or buy products and services. This revolution made it possible for the bank to create a market place which has turned to set in motion a revolution in the banking sector for the provision of a payment system that is compatible with the demands of the electronic marketplace.

The study finding shows that Information and Communication Technology (ICT) helped to provide supportive role for the human resources (staff) activities at the bank. This implies that the bank operations and activities concerning providing services to its numerous customer are been done through the human resources at the bank, hence, through ICT at the bank had made Ghana Commercial Bank Limited to reduce the number of staffs who are responsible for the daily operations of the bank thus cutting the wage bills of the bank in terms of the cost involved in paying the huge staff in undertaking the operations at the bank.

The study findings also shows that the bank made used of Information and Communication Technology of which this helped the bank to saved time in their operations at the bank. This implies that bank activities were mostly facilitated by the use of ICT machines which made the bank to provide its services to its customers on time made it faster. This means that the budget that was allocated to the running of the operational activities of the bank was been cut off through the bank technological used of ICT at the bank

The study findings shows that the bank made used of Information and Communication Technology (ICT) contributed to the reduction of operational cost of Ghana Commercial Bank Limited by way of the bank using ICT to fasten the decision making process at the bank. This implies that the bank made used of ICT which did support the decision making process at the bank. The operational cost involved in Ghana Commercial Bank Limited is very huge due to its largest coverage across the country; hence, the bank used ICT to help in making decision through video conferencing, CCT cameras and telephones.

The study findings concerning the effects of Information and Communication Technology (ICT) helped to improve the bank monetary reserve. This implied that information technology was used as medium to channel and assisted in mobilizing deposit from the bank numerous customers and also shows that the bank made used of ICT as form of technology which did help the bank turnaround time at the banking hall because of the kind of technology that were used by the banking staff in providing better services to the customers. This encouraged the customers to save much of their income with the bank hence, increasing the bank profitability and its liquidity at large.

The study finding shows that the bank used of Information and Communication Technology (ICT) helped in improving the bank fixed assets at the bank. This means that the bank made used of information technology in its products and services to its customers which made the bank to have access to modern fixed assets in their disposal for providing and assisting the bank in their services.

The study findings shows that the bank technological advancement through the used of Information and Communication Technology (ICT) helped in protecting the bank liquidity either in the form of money or property. The study revealed that the bank used information technology to helped in protecting the banking hall through the use of modern CCTV cameras in detecting fraud. The study also shows that the bank used modern technology in encryption their confidential information on the bank data from allowing fraudsters from hacking into the banking system and defrauding the bank.

9. CONCLUSION

In conclusion, the study observed by drawing its conclusion from the findings of which the results of the research indicate that investment on ICT systems has become a key element in the performance at Ghana Commercial Bank Limited. The study revealed that ICT-Capital has accelerated growth in the bank. Also, ICT facilitates the absorption of high and medium skilled labor. This has a positive effect on the labor output at the bank. The study findings revealed that the bank used of ICT also enabled Ghana Commercial Bank Limited to offer a broad variety of services to customers, coordinate branch activities, meet up with changes in government regulations and policies as well as adjust to market demands and competition. The

study observed by concluding that the bank used of Information and Communication Technology (ICT) has had an impact on the performance of the bank by way of increasing the bank profitability and also reducing the operational cost of the bank. The results suggest that improving the service quality through ICT experience had improved the overall customer perception of the quality of service provided by the bank. The study observed that the customers of the bank appear generally satisfied with the technological aspects of the bank. In other words, the bank customers are now familiar with the basic service provided by ATMs and bank websites through which the customer's satisfaction for these ICT devices has been great of benefit to the bank. They revealed that the bank has taken their ICT to the next level in order to stay one step ahead of the competition.

10. RECOMMENDATIONS

The study recommends that the bank should research and have better feasibility studies on their customers to know the kind of information technology that will be suitable for the bank and its customers to help in the bank financial performance.

The study also recommends that the bank should involve itself more in the use of electronic banking rather than the traditional way of customers always having to visit the banking hall of which it will fasten and will encourage deposit mobilization because of the convenience and easy way of assessing transactions.

The study recommends that the bank's management should be well advised to conduct periodic marketing research studies on their own customers, to include developing a database containing both the demographic and psychographic profiles of its customer base. Such a database would then allow the management to make a more informed decision about the level of technological services needed at any one time in relation to the level of personal staffing to ensure adequate service delivery.

The study recommends that the bank should concentrate their efforts on those areas customers feel are most important (i.e. accuracy, timeliness, and convenience). These are areas in which technology (capital expense) has been well utilized to replace more costly labour resources to accomplish essentially those services that the customer is accustomed.

11. FURTHER STUDIES

There are a number of research opportunities to explore in the future based on this study. The following recommendations are therefore made for future research.

- Future research should look at Information and Communication Technology (ICT) and its contribution to the development of the country at large.
- Future research should also take into consideration a larger sample size upon which better conclusion can be drawn from them.
- Future research should aim to improve impact of ICT, by examining whether ICT usage or adoption varies in importance across the banking sector in the country.

12. REFERENCES

[1] Castells, F (2001). The Adoption of Direct Banking Services, *Journal of Marketing Management*, *23*, 91-111.

[2] Chan, K (2000). *Effects of Information Technology on Customer Services in the Banking Industry in Ghana*. GWEP Publications: Lagos.

[3] Davis, F.D (1993). Perceived usefulness, perceived ease of use, and user acceptance of information technology, *MIS Quarterly*, vol. 13, no. 3, pp. 318-39.

[4] Duncombe, R. & Heeks, R. (2001). *Information and Communication Technologies and Small Enterprise in Africa, Lessons from Botswana*. IDPM University of Manchester.

[5] Jayamaha, O (2008). Assessment of Impacts of Information Technology on Selected Service Industries in South Western Nigeria, MSc. Thesis in technology Management, Obafemi Awolowo University, Nigeria.

[6] Khalifa , G (2000). Deploying internet Banking and e-Commerce: Case Study of a Private Sector Bank in India, Information Technology for Development, Vol. 11, No. 4,pp. 363-379.

[7] Khanna, S (2003). The Use of Information Technology to Transform the Banking Sector in Developing Nations, Information Technology for Development, Vol.11, No. 4, pp. 305-312.

[8] Laudon, D.P. & Laudon, J.P. (1991). Business Information System: A Problem Solving Approach, New York, HBJ, College Publishers.

[9] Leow, H. B. (1999). New Distribution Channels in banking Services. *Banker's Journal Malaysia*, 110, 48-56.

[10] Mehta Kamlesh T. & Shah Vivek (1998). "Workforce, Information Technology and global unemployment", Industrial Management & Data Systems, Vol 98, no. 5, pp. 228.

[11] Müller-Falcke, D. () *The Use of Telecommunication and Information Technologies in Small Businesses - Evidence from Indian Small Scale Industry*. Forthcoming.

[12] Odedra, M. & Kluizer, S (1988). "Bibliography for Information Technology in Developing Counties," Information Technology for Development. pp. 297-356.

[13] Ovia, J. (1997). *New Technologies and Performance Enhancement*, Paper presented at the 13th Annual Bank Directors' Seminar Held at Abuja between 17th and 19th of June, 1997.

[14] Yin, R. (1994). *Case study research: Design and methods* (2nd ed). Beverly Hills, CA: Sage Publishing

Water Supply Optimization Strategy Using Qualitative Analysis Approaches

Guiyuan Yuan
Shandong University of
Science and Technology,
Qingdao 266590, China

Shanshan Li
Shandong University of
Science and Technology,
Qingdao 266590, China

Cong Liu*
Shandong University of
Science and Technology,
Qingdao 266590, China

Abstract: This paper aims to list the possible risk of water resources with fully considering of multiple factors, and analyse the factors influence on the shortage of water resources, and finally put forward optimization strategy and reasonable suggestions through a systematic qualitative analysis. More specifically, (1) establishing four models to measure water supply capacity: From the natural environment, social water, economic factors, and the new technology factors to create the model, (2) taking Beijing as a survey centre, to get the weight of each risk factor to draw the top 5 main risk factors, (3) building the equation by multiple linear regression, through the curve fitting analysis to predict water status in Beijing city for the past15 years, (4) making an intervention plan and modify model parameters on the basis of economy and environment, and (5) Based on the intervention plan, time variable is modified to predict whether Beijing will be affected by water shortage in the future.

1. DESCRIPTION

Water resource on earth generally refers to the overall content of the hydrosphere[1,2]. Including the human control and direct irrigation, power generation, water supply, shipping, aquaculture and other purposes of surface water and groundwater, as well as rivers, lakes, wells, springs, tidal, harbor and aquaculture waters, etc. In a narrow sense, it refers to the amount of fresh water that can be recovered and renewed every year[3,4]. Water resources are important natural resources for the development of the national economy[5]. In many parts of the world, the demand for water has exceeded the level of the water resources, and there are many areas on the verge of the use of water resources is not balanced[6,7,8]. Therefore, we need to establish a set of effective water resource model to solve the world water problem. Throughout the modeling process, we fully consider the feasibility of our model for effectiveness, efficiency, and cost.

The problems to be solved inludes:

1) Develop a model that provides a measure of the ability of a region to provide clean water to meet the needs of its population. You may need to consider the dynamic nature of the factors that affect both supply and demand in your modeling process.

2) Using the UN water scarcity map pick one country or region where water is either heavily or moderately overloaded. Explain why and how water is scarce in that region. Make sure to explain both the social and environmental drivers by addressing physical and/or economic scarcity.

2. SYMBOL CONVENTIONS
2.1 Symbol Conventions
ZY: Total water consumption;

GY: industrial water consumption;

NY: agricultural water consumption;

SH: third industrial and domestic water consumption;

ZL: total water resources;

JS: precipitation;

DX: underground water;

DB: surface water content of;

ZS: reclaimed water;

λ: characteristic root;

E: cumulative contribution rate;

U: risk factors of water shortage;

V: hierarchical domain;

T: membership degree;

w^*: entropy weight;

B: comprehensive evaluation risk value;

S: correlation matrix;

R: fuzzy matrix.

3. ANALYSIS OF THE PROBLEM

For question 1, a four model was established to measure the water supply capacity, and a model was created to solve the task from four aspects: natural environment, social water, economic factors, and new technology.

For question 2, according to the selected regions of the economic environment variable is introduced to solve the model; Beijing, China as a water reference City, Beijing city's natural environment and climate, economic and population factors, water supply system, technical reserves into the model,

through social and environmental driving factors to explain reasons of water shortage in Beijing City.

4. MODELS
4.1 Water Resources Model
Beijing city as a research center, according to the economic environment of Beijing City, the introduction of the model to solve the environmental variables.

4.1.1 Analysis of the present situation
Will Beijing's natural environment, topography and climate precipitation, economic and population factors, water supply system, technical reserves, water pollution, water price, water purification ability is introduced into the model through social and environmental driving factors to explain the cause of water shortage in Beijing city. Then statistical results is the following table:

Table 1. Beijing water resources statistical results

project	2005	2006	2007	2008	2009	2010	2011	2012
Ground water	15.7	14.7	14.8	16.5	18.5	18.5	19.2	21.4
Surface water	7.8	5.3	6.1	8.2	7.6	6.0	7.6	12.8
Reclaimed water	2.5	2.0	2.1	2.0	2.6	3.6	5.0	2.1
Amount of water	26	22	23	26.7	28.2	28.1	31.8	36.3

4.1.2 Model Testing
Design water supply for y, the factors affecting the total water resources amount is the amount of surface water resources, total amount of groundwater resources and renewable water were recorded for x_1, x_2, x_3, establish three element linear regression model $= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$, among them$\varepsilon$is random error obeys normal distribution n $(0, \delta), \beta_0, \beta_1, \beta_2, \beta_3$ regression coefficients.

Using a 2005-2012 chronology data and on the type, then

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon (\alpha = 1, 2, 3 \ldots, 8)$$

Let b_0, b_1, b_2, b_3 respectively, the parameters of $\beta_0, \beta_1, \beta_2, \beta_3$ of the estimated value, then get the regression equation $\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3$

$$\hat{y}_\alpha = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 (\alpha = 1, 2, \ldots, 8)$$

$y_\alpha (\alpha = 1, 2 \ldots, 8)$ as sample value; y goes alpha as the regression equation given value becomes the regression value,

called the $y_\alpha - \hat{y}_\alpha (\alpha = 1, 2 \ldots, 8)$ as residual, it can be classified as the deviation of the sample value and the return value.

According to the least square method, the least square sum of residuals is achieved

$$Q = \sum_{\alpha=1} (y_\alpha - \hat{y}_\alpha)^2$$

According to the limit principle of calculus b_0, b_1, b_2, b_3 must be satisfied for the minimum

$$\frac{\delta Q}{\delta b_i} = 0 (i = 0, 1, 2, 3)$$

The normal equations in $\left(X^T X \right) \hat{B} = X^T Y$

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{13} \\ 1 & x_{21} & x_{22} & x_{23} \\ 1 & x_{31} & x_{32} & x_{33} \\ 1 & x_{41} & x_{42} & x_{43} \\ 1 & x_{51} & x_{52} & x_{53} \\ 1 & x_{61} & x_{62} & x_{63} \\ 1 & x_{71} & x_{72} & x_{73} \\ 1 & x_{81} & x_{82} & x_{83} \end{pmatrix} \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \end{pmatrix} \quad \hat{B} = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

Estimated value of solution:

$$\hat{B} = (X^T X)^{-1} (X^T Y) = \begin{pmatrix} -4.732 \\ 2.138 \\ 0.498 \\ 0.274 \end{pmatrix}$$

This is calculated by $\hat{B}_i$ regression equation through the regression equation as follows:

$$\hat{y} = -4.653 + 2.091 x_1 + 0.398 x_2 + 0.286 x_3$$

hold $(x_{1\alpha}, x_{2\alpha}, x_{3\alpha})$ and on the type of solution of each regression value $(\hat{y}_1, \hat{y}_2 \ldots, \hat{y}_8)$

Calculate   ,

$$\bar{y} = \frac{1}{8}\sum_{\alpha=1}^{8} y_\alpha \rightarrow U = \sum_{\alpha=1}^{8}(\hat{y}_\alpha - \bar{y})^2 = 181$$

Degree of freedom $f_u = p = 3$

$$Q = \sum_{\alpha=1}^{8}(\hat{y}_\alpha - y_\alpha)^2 = 28 \quad \text{Degree of freedom}$$

$$f_\alpha = n - p - 1 = 8 - 3 - 1 = 4$$

Total deviation square $S_{yy} = U + Q$

Verifying the significance of regression equation is equivalent to test hypothesis

$$H_0 : \beta_1 = 0, \beta_2 = 0, \beta_3 = 0.$$

For statistical measurement

$$F = \frac{U/f_u}{Q/f_Q} = \frac{U/P}{Q/n-P-1} = \frac{181/3}{30/4} = 8.67$$

Use $\alpha = 0.05$ search table with a critical value of $F_{0.05}(3,4) = 6.59$,easy to see the

8.67>6.59, and therefore refused to H_0.

Namely, the effect of regression equation was significant.

4.2 model of influence factors of water consumption

From the chart we can see that before 1994 due to policy, farming, the industrial structure of the problem data fluctuations, and in 1994 after can be regarded as a continuous curve, so we choose after 2003 data to carry on the fitting analysis, relational expression is obtained.
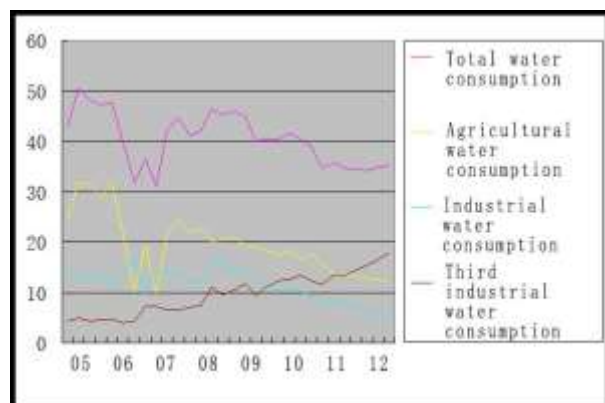

Figure. 1 the total water use agricultural water consumption, industrial water consumption, and the third industry and other water use fitting tables

Considering the continuous change of time, the curve of agricultural water consumption was obtained by fitting, as shown in Figure 2.
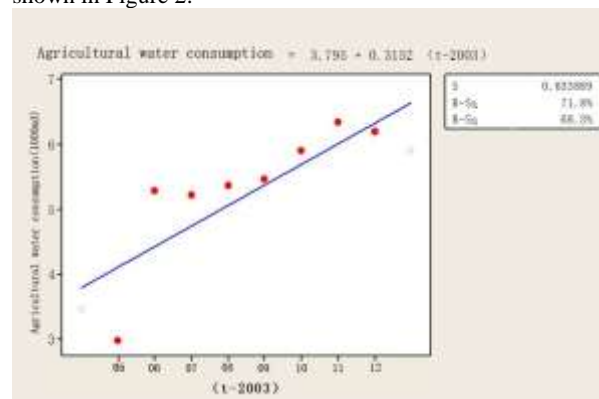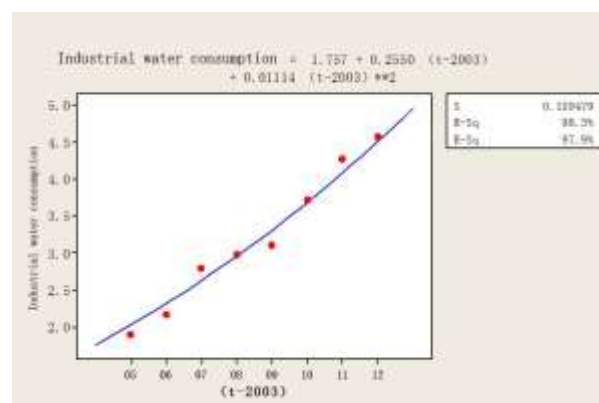

Figure. 2 fitting image of agricultural water

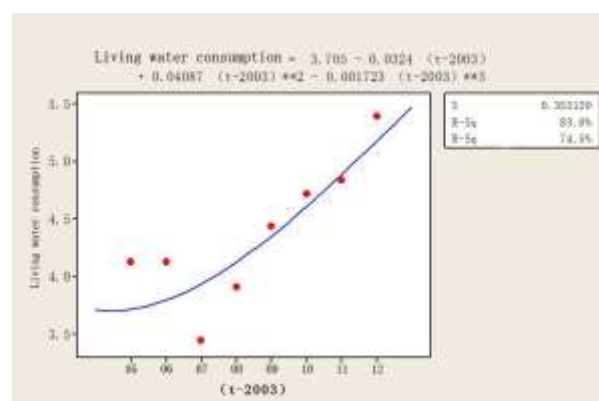Figure. 3 fitting image of industrial water
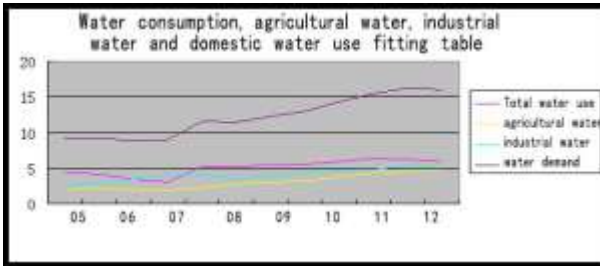

Figure. 4 the water of life

Figure. 5 Total water consumption, agricultural water consumption, industrial water consumption and other water use fitting table third

4.3 model of risk of factors of water shortage

4.3.1 Analysis on risk factors of water shortage
There are many factors that affect the water resources of Beijing City, the paper divides the risk factors from two aspects: water supply and water use.

Table 2. statistical analysis of risk factors of water resources shortage in Beijing city from 2005 to 2012

Risk factors	2005	2006	2007	2008	2009	2010	2011	2012
DX	15.7	14.7	14.8	16.5	18.5	18.5	19.2	21.4
DB	7.8	5.3	6.1	8.2	7.6	6.0	7.6	12.8
ZS	2.5	2.0	2.1	2.0	2.6	3.6	5.0	2.1
GY	9.2	7.5	8.4	7.7	6.8	6.2	5.8	5.2
NY	17.4	15.5	13.8	13.5	13.2	12.8	12.4	12.0
SH	12.0	10.8	13.0	12.8	13.4	13.7	13.9	14.7

To quantitatively analyze the influence degree of each risk factor of the status of water resources, the judgment matrix in AHP pairwise comparison of each risk factor, to determine the weights of various risk factors, which determine their influence on the total amount of water resources.

The establishment of the matrix A is as follows：

$$
\begin{pmatrix}
& C1 & C2 & C3 & C4 & C5 & C6 & C7 & R & W \\
C1 & 1 & 3.0776 & 1.3884 & 8.1893 & 2.8837 & 4.7657 & 2.9678 & 24.2725 & 0.271337 \\
C2 & 0.3247 & 1 & 0.4511 & 2.6609 & 0.9370 & 1.5485 & 0.9643 & 7.8868 & 0.088164 \\
C3 & 0.7203 & 2.2167 & 1 & 5.8589 & 2.0771 & 3.4326 & 2.1376 & 17.4826 & 0.195434 \\
C4 & 0.1221 & 0.3758 & 0.1695 & 1 & 0.3521 & 0.5819 & 0.3624 & 18.2615 & 0.204141 \\
C5 & 0.2098 & 1.0672 & 0.4815 & 2.8398 & 1 & 1.6526 & 1.0291 & 8.2801 & 0.092561 \\
C6 & 0.2098 & 0.6458 & 0.2913 & 1.7184 & 0.6051 & 1 & 0.6227 & 5.0932 & 0.056935 \\
C7 & 0.3370 & 1.0370 & 0.4678 & 2.7594 & 0.9717 & 1.6058 & 1 & 8.1787 & 0.091427
\end{pmatrix}
$$

among : $A = \left(a_{ij} \right),\ S = d_i - d_j, d_i = \sum_j c_{ij}$

$$R_i = \sum_{j=1}^{7} a_{ij} \qquad W_i = R_i \Big/ \sum_{i=1}^{7} R_i$$

W_i :The weight of the risk factor

$$W_1 = 0.271337 \quad W_2 = 0.88164 \quad W_3 = 0.195434$$
$$W_4 = 0.204141 \quad W_5 = 0.092961 \quad W_6 = 0.056935$$
$$W_7 = 0.091527$$

Through the calculation of 5 main risk factors, respectively, the weight of the relatively large surface water C4, C2 reclaimed water, C5 agricultural water, C6, industrial water, C7 third industry and domestic water.

4.3.2 Principal component regression
According to the ontology with fewer samples, variable dimension, multiple linear problem is serious by principal component regression model can in reducing the dimension of the variables at the same time eliminate variables highly correlated due to multicollinearity problems, so as to improve the accuracy of the regression model parameter estimation. At the same time, the establishment of the model also from the side to test the effectiveness of the above extraction of five risk factors.

A water shortage rate is Y, $Y = (U - V)/U$. U for the annual total water, V for the annual total water resources. In order to X1, X2, X3, X4, X5 as explanatory variables, Y as the explanatory variables for multiple linear regression. Regression model can be written as

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5$$

water resource scarcity risk factors were obtained from the principal component analysis:X1,X2,X3,X4,X5, The regression equation obtained by SPSS is:

$$Y = -0.024X_1 + 0.1X_2 - 0.124X_3 - 0.15X_4 - 0.09X_5 + 0.271$$

Goodness of fit test and F test of the regression equation: the test of goodness of fit, the model had an adjusted $R_2 = 0.898$, variance analysis of the significant level of less than 0.05. The test results show that the regression equation fitted well with the original data.

X_1 is known for industrial water comprehensive factors, X_2 for life and water environment comprehensive factors, X_5 for agricultural water integrated factors, by regression equations available, according to their contribution to the rate of water shortage of the rate of its importance to sort for:

Industrial water > living and environmental water > agricultural water.

It can be known that the regression model can well explain the relationship between water shortage rate and risk factors, and the significance of the index is consistent with the actual situation, which can be applied.

4.4 predicted and forecast model

4.4.1 Predicted total water consumption

Task three solution: water condition forecast for the next 15 years Environmental drivers: select Beijing City, according to the economic environment variables and time variables of Beijing city to solve the model. On the basis of population increase, scientific and technological progress and environmental change, the water status of Beijing city in the next 15 years is predicted.

Agricultural water consumption:

$$NY = 0.3152(t-2003)+3.795$$

Industrial water consumption:

$$GY =$$
$$0.01114(t-2003)2+0.2550(t-2003)+1.757$$

Third industry and other water use:

$$SH = -0.001723(t-2003)3+$$
$$0.04807(t-2003)2+0.0324(t-2003)+3.705$$

The total water use: $ZY = NY + GY + SH$

Predict the next 15 years, the total annual water consumption results is the following table:

Table 3. the total annual water consumption

Year	NY	GY	SH	Total
2016	7.10	17.00	12.33	23.53
2017	7.12	17.20	12.41	23.71
2018	7.14	17.52	12.25	23.87
2019	7.17	17.86	12.33	24.00
2020	7.25	17.69	12.45	25.23
2021	7.25	17.81	12.56	25.31
2022	7.56	17.73	12.67	25.28
2023	7.12	17.92	12.70	25.54
2024	7.67	17.93	12.71	25.57
2025	7.78	17.92	12.78	25.56
2026	7.81	17.87	12.85	25.63
2027	7.92	17.89	12.88	25.69
2028	7.93	17.92	12.92	25.77
2029	7.89	17.98	12.98	25.85
2030	7.90	18.12	13.23	16.23

4.4.2 Forecast total water resources

Based on our analysis on data from 2005 to 2012, surface water and groundwater in a certain range fluctuations, reclaimed water quantity rise in the basic trend, in view of this,

we have a bold prediction of four years of water resources in the future. According to the principle of Statistics found that the basic surface water and underground water fluctuations in a certain range, while the sewage treatment capacity because of the support of national policy it has been in a rising state, we take years of data for comparative analysis, speculated that the next four years surface water and underground water to maintain in the eight years ago, the average level of sewage treatment capacity for nearly three years on average, so the water resources amount in the next four years is speculated as follows table:

Table 4. water resources amount in the next four years

Year	Surface water quantity	Underground water	Sewage treatment capacity	Total water resources
2016	7.70	17.00	12.33	23.53
2017	7.70	17.20	12.41	23.71
2018	7.70	17.52	12.25	23.87
2019	7.70	17.86	12.33	24.00
2020	7.70	17.69	12.45	25.23
2021	7.71	17.81	12.56	25.31
2022	7.72	17.73	12.67	25.28
2023	7.72	17.92	12.70	25.54
2024	7.83	17.93	12.71	25.57
2025	7.78	17.92	12.78	25.56
2026	7.81	17.87	12.85	25.63
2027	7.92	17.89	12.88	25.69
2028	7.93	17.92	12.92	25.77
2029	7.89	17.98	12.98	25.85
2030	7.90	18.12	13.23	16.23

Description: the total amount of water resources can be solved by linear regression equation:

$$\hat{y} = -4.732 + 2.138x1 + 0.498x2 + 0.274x3.$$

5. WATER SHORTAGE DRIVEN FACTOR INTERVENTION PROGRAM

5.1 Holistic analysis

Intervention program from the natural environment, topography and climate, economic and population factors, water supply system, technical reserves, water pollution, water price, water purification ability of each model and the model parameters are taken into account, as far as possible according to the different situation of different places to draw the best interventions.

5.2 Model introduction

Since The index system is a reasonable abstraction and Simulation for the sustainable utilization of water resources, but the contribution of each index to the system is different, which can be used to a set of normalized weight W_1, W_2,... ,

W_n to describe. Traditional weight distribution commonly used Delphi expert consultation method and analytic hierarchy process (AHP). Among them, the Delphi method is not only a higher requirement for consulting experts, but also a large number of rounds of consulting work. AHP uses nine scale score comparison to judge the result, makes the person feel the operation difficulty, and the computation complexity, but also needs to carry on the consistency check. Under the premise of guaranteeing the scientific nature, in order to calculate the simple and convenient, this paper uses the improved three scale analytic hierarchy process, its computation steps like formula (1) ~ (3).

Subjective comparison matrix：$C = \left[c_{ij} \right]_{n \times n}$,

Where n is the number of indicators, the element C_{ij} is expressed in terms of S_k, the relative importance of C_i relative to C_j,

Establishment of sensory judgment matrix:

$$S = \left[s_{ij} \right]_{n \times n}, (2)$$

Among them, $S = d_i - d_j, d_i = \sum_j c_{ij}$

Calculation of objective judgment matrix:

$$R = \left[r_{ij} \right]_{n \times n}, (3)$$

Among them,

$$r_{ij} = p^{(S_{ij}/S_m)}$$

$$s_m = \max_{i,j} s_{ij} = \max_i(d_i) - \min_j(d_j)$$

P is defined as the scale extension of the user, such as p=3 or 7, this article takes 3. We can get the weight vector R, W2, by normalizing the arbitrary column of matrix [W1,... , Wn]T.

The criterion layer 4 indexes, through consultation and calculation, it is concluded that their weights. The present situation of water resources for 0.2765 and water resources utilization efficiency 0.1917, the pressure of the sustainable utilization of water resources for 0.1329 and ability of sustainable utilization of water resources is 0.3988. Then the index system in the next level of the indicators were calculated, the last layer of indicators in the index system of the weight of the factors (Table 5).

Table 5. w*ater resources sustainable use specific evaluation index weight*

target	Criterion layer	weight
General level of sustainable utilization of water resources	Current situation of water resources *(S)*	0.2765
	Water consumption *(E)*	0.1917
	Water resources sustainable utilization pressure *(P)*	0.1329
	Sustainable utilization of water resources(C)	0.3988

Events related to network news, forums, blog, microblogging and other Internet information resources, information capacity, high redundancy and susceptible to noise interference. To collect comprehensive information on these and filtering, on the one hand a huge workload and high labor costs, On the other hand due to the passage of time, some of the information on the Internet will disappear, to be covered by the new information, Therefore, quantitative study of this article only collect and use information representative corpus as samples. Although, in the foregoing chapter we have made under the existing study based on the results of the theory and use of the public emergency network news also from a public opinion which reflects the evolution of the development and changes of the network news coverage. However, Insufficient sample of the present study was to analyze the qualitative part brings some limitations.

Since public opinion guidance and management is a very practical work, specific to the implementation level. We need to deal with a large number of complex issues and deal with complicated things, due to the lack of implementation experience, guidance and management research in the public emergency network public opinion aspects, This article does not find significant innovations, suggestions and measures proposed are mainly to further enrich the existing research-based.

6. MODEL EVALUATION
6.1 Strengths
(1)interfere with the plan in the principal component analysis model based on, eliminating the precipitation, GDP, industrial water and other indicators related effect, reduce the workload of index selection, more objectively reflect the display relationship between risk factors and water shortage.

(2) in this paper, a comprehensive evaluation model of water resources based on entropy weight is established, which can avoid the problem that the pairwise contrast structure judgment matrix is easy to happen. The theory of entropy in the information theory was introduced from the data itself reflects the information disordering effect value to calculate the weight coefficients of the weight determination of the certain theory basis to reduce the subjectivity of evaluation and subjective factors influence and evaluation of the results.

(3) the establishment of a gray prediction model, to predict the risk of water shortage in Beijing 2011, 2012 years, and effectively quantify the risk of water shortages in the next two years in Beijing.

(4) all models of this paper, the whole process of clear thinking, simple calculation, the objective of the evaluation results.

6.2 Weaknesses
(1) the inevitable part of the principal component analysis model is to be lost.

(2) the grey model is used to predict, in view of the smoothness of the sequence is not good, the selection of the background value is more difficult.

(3) due to the environment and human impact of water resources is greater, so the results of the prediction error is inevitable.

(4) In the future, we prefer to integrate the optimization with some traditional data mining approaches [8], [9], [10] to find more effective solutions.

7. REFERENCES

[1] Gordon F L Florida west coast desalter negotiations-progress report #3 1998(03) Scott A S The work exchanger for SWRO energy recovery 1999(04).

[2] Gordon F L.Developer selected for 25MGD(94625m3/d) Florida west coast seawater desalting plant[J]. The international Desalination &Water Reuse, 1999, 9(1): 11~16 .

[3] Xiaojun Wang; Hui Zhao;Zhi Geng,Current groundwater exploitation situation and protection countermeasure in China, China Water Resources; 2010 13.

[4] Siqing Lin; Weirun Zhang, Cost of seawater desalination. J MEMBRANE SCIENCE AND TECHNOLOGY, Vol. 22 No.4: 1007-8924,2001 .

[5] L. Awerbuch, Integration of desalination and po wer: the challenge in the Middle East. Paper presented at Middle East Energy. Dubai (8–10 October 2001) .

[6] Ting Lu, Qi Gao, Xudong Wang, Cong Liu, "Modified Page Rank Model in Investigation of Criminal Gang", International Journal of Science and Engineering Applications, 4(3): 100-104, 2015.

[7] Baohua Liu, Xudong Wang, Qi Gao, Cong Liu, "A K-Means based Model towards Ebola Virus Prorogation Prediction", International Journal of Science and Engineering Applications, 4(3): 95-99, 2015.

[8] Libao Zhang, Faming Lu, An Liu, Pingping Guo and Cong Liu*, "Application of K-Means Clustering Algorithm for Classification of NBA Guards", International Journal of Science and Engineering Applications, 5(1), pp. 1-6, 2016.

An Easy-Using Evaluation System Model in Terrorism Based on Individual Risk Index and Classification

Jing Tang
College of Information Science and Engineering
Shandong University of Science and Technology,
Qingdao, Shandong, China

493117148@qq.com

Ning Zhang
College of Information Science and Engineering
Shandong University of Science and Technology,
Qingdao, Shandong, China

zning@sdust.edu.cn

Abstract: In this paper, we first establish the evaluation model of the individual risk index based on the social network analysis. The attributes of the individual through the behavior of the performance is divided into social attributes (including psychological attributes) and network attributes, and then to the social attributes of refinement, and that with some social attributes of people more dangerous. In which, we analyze the psychological factors, and the comprehensive evaluation of the factors, the network attributes are linked to the second tasks in the model. Then, we establish the model of double layer classification based on fuzzy clustering model and weighted Bias classification model. So as to make a preliminary classification of a large number of data, and finally get the specific behavior of the classification table "The Definition Sample of Cut Set of Fuzzy Matrix ". Then the weighted Naive Bayesian classifier is used to classify the individual behavior sets, so that each behavior is obtained from the correlation with terrorism. Then, we can use the weight and then return to the analysis of the individual can be derived from each person's expectations, the expectations of everyone, and applied to the first task of the risk index evaluation.

Keywords: Combat Terrorism; Individual Risk Index; Evaluation Model; Fuzzy Clustering; Weighted Bias Classification; Double Classification Model;

1. INTRODUCTION

Figure 1. Terrorism is becoming the focus of attention (Photo from Internet)

Since 9.11, terrorism has become one of the main factors that disrupts the international order, threatens national security and social stability. [1] With the development of network, modern terrorism has presented a series of new features and it has put a serious challenge to the national antiterrorist activities.

Recently, a group of the extreme religious jihad who calling itself ISIS (of Islamic State of Iraq and Short al-Sham) is increasingly rampant and anti-terrorism situation is imminent. Under the current situation, the terrorist activities have become more wanton rampant rather than Low-key. Today, in the Internet boom, they are distorted by instant messaging, social networks and other forms of propaganda and make more people join them. [2] [3] But a lot of people, who have psychological or mental problems, listen to their propaganda and then engage in terrorist activities. How can we find them, and make timely intervention? This is a proactive approach. Our paper will focus on this phenomenon to predict their behavior through the network and data flow so that we can identify the people who real need "repentance".

2. THE DESCRIPTION OF THE PROBLEM

In our paper, we will analyze the following contents:

1. For first task, we will establish an evaluation model based on social network analysis of individual risk index.

   By analyzing the social attributes (including psychological attributes) and network properties, we will determine whether a person is a terrorist and assess his (or her) acceptance of terrorism. Based on the outcome, we will give his (or her) proposed measures accordingly.

2. For the second task, we will establish a data classification model based on Naive Bayes classification. Based on the original big data provided by experts, we carried out a preliminary screening classification and obtain the status of each person's behavioral characteristics. At the same time, we can get the situation of terrorism in the region.

   We believe that the information obtained by the task two will be helpful for the analysis of the individual risk index in the task one. And the specific role of this part of the mission will be specifically explained in the model of task one.

3. In the end, we will propose a reasonable suggestion to President Obama in response to the rampant ISIS in the near future. And at the end of the article, we will evaluate and summarize our model.

3. THE HYPOTHESIS OF PROBLEM

In the section, we use some symbols for constructing the model as follows.

We assume that:

1. The data that experts have provided is very useful for us. And they provide the Internet keyword about terrorist organization or terrorist activities.

2. We believe that the individual of the given data will not make physical and psychological changes within the time limit (72 hours).

4. THE DESCRIPTION OF SYMBOL

Table 1. The Description and Explanation of Symbol in this Paper

Symbol	Description and Explanation
$C_d(u)$	Point degree centrality
$r(u,v)$	It represents the node u, whether there is a link with v
$C_c(u)$	Approaching Centrality
$C_B(k)$	Mediation Centrality
$C'_B(k)$	relatively Centrality
σ_{ij}	Number of the shortest path between the node B and S
$\sigma_{ij}(k)$	Number between any two points i and j with k shortest path through the junction
a_{ij}	Relationship between nodes i and j
xi	Feature vector Centrality
λ	The maximum eigenvalue
μ	Scaling factor
C_{eff}	Graph Theory (network), the average efficiency
$E(G)$	Network Efficiency
d_{ij}	The shortest distance between nodes i and j
I	Individual Risk Index

5. MODELING

5.1 An Evaluation Model in Individual Risk Index Base on the Social Network Analysis

In recent years, with the dynamic social network analysis, we can find complex covert network, and through a series of algorithms to estimates people who have been in contact with terrorism individuals risk index, in order to determine a key figure what we need. And then take corresponding measures to solve it.

In terms of the key figures digging, Dr. Nasrullah Mammon proposed concept of Network Efficiency, Location of Role Centers and Dependence centrality based on Traditional Center index. It provided calculating individual risk index with a new set of indicators. Dr. D. M. Akbar Hussein proposed a new method to calculate individual risk index, these are of high academic value in the key figure in the fight against terrorist networks.

5.1.1 The constitution of Individual Risk Index Assessment

Individual risk index is divided into two parts, social attributes and network attributes. By calculating the two parts of the property value, we can get an individual preliminary risk index of who have been in contact with terrorism.

1. Social attributes, social attributes are the measurement indicators based on the basic situation of the individual with the person who has been exposed to terrorism. Generally considered, individuals with some social attributes are more threatening than others, so it could potentially be a key figure. According to the analysis of the relevant terrorist information display [4] [7] [10] [11] [14], people who have had contact with terrorism have a greater similarity in the following 14 attributes. Among them, the psychological factor accounted for a large proportion of the comprehensive measurement.

Therefore, we mainly select these 14 social attributes, the specific description of the table as follow:

Table 2. The selected social attributes (Total 14)

Attribute name	Explain
Name	Name appeared in the international or national terrorist files are more threatening
Age	In general, young people are more enthusiastic and aggressive, more than the elderly threat
Sex	Men are often more threatening than women.
Marriage	Married people tend to be stable, so the threat is relatively small.
Nationality	People who come from the countries of terrorism are more threatening.
Religion	People who have faith and believe in religion are more threatening than those who are not.
Education	Under normal circumstances, the higher education level, the greater the threat.
Occupation	People engaged in blasting, chemical, physical, transportation, computer and finance and other industries have professional skills, able to provide technical support for terrorist organizations, funds and other support, and therefore may be more threatening

Attribute name	Explain
Crime history	It can be considered that the people who have the history of crime are more threatening.
Psychological index	People who have a certain degree of psychological problems（For example, depression, or dying）are easier access to terrorism, the greater its threat
Comprehensive measurement	According to the data collected by the intelligence, According to the comprehensive index of its behavior, from this comprehensive indicator of their daily behavior, then give the impact factor
Participate in the number of terrorist activities	People involved in many terrorist activities the more experienced, therefore more threatening
Social features	If a person's social appearance is bad, compared the good person, can be considered as a threat to the greater
Recruiting ability	If someone has the ability to recruit terrorists, it is more threatening.

2．Network attributes. Network attribute is the measurement index determined by a series of social network analysis method. Social attributes are based on the personal characteristics of the individual with the contact with terrorism, there are some subjective and one-sided. [5] [9] [12] Here, we analyze the network properties based on social network analysis, to be more objective, comprehensive evaluation of the risk of exposure to the individual risk index. In social network analysis, the center is one of the most important indicators. Therefore, we determine the network properties based on central degree. The main network attributes include: Center of point, Proximity Center, Center of media, Eigenvector centrality and network efficiency.

1) Point degree center. The center of the point is the typical representative of the center; it refers to the degree of concentration as determined by each node in the network diagram. Index of individual activity, it is the most direct and simple way to measure individual centrality. This concept comes from the concept of "Star" in the social metrology. In terrorist activities, terrorists are often more important to have a relatively high degree of centrality. If a person has been exposed to terrorism has a direct connection with many other members of the terrorist activities, then it can be considered as an important platform for the members to obtain and exchange information in the network.

The most common definition of the center of the point is:

$$C_d(u) = \sum r(u,v)$$

Among them, r (U, V) is a two variable, which is used to indicate whether there is a link between u and v.

(2) Close to the center. Close to the center of the network is based on the distance between a node and other nodes. The shorter the total distance, the higher the degree of the network. Close to the center of the information to measure how long it can spread from one node to the other nodes in the network,

shows the close degree between the actors and the other actors. In terrorist activities, the individuals who are close to the central level are in the central position of the contact with terrorism, they can get in touch with other members of the terrorist activities through a minimum of intermediate position, so that more independent of the other terrorists.

For the most common definition of proximity to the center:

$$C_c(u) = \frac{1}{\sum_{v \in V} d(u,v)}$$

Among them, $\sum_{v \in V} d(u,v)$ representing the distance of any point $u \in V$ in graph theory $G = (V,E)$.It should be noted that, unlike the point of the center of the degree, the proximity of the center is a measure of the overall network.

(3) The center of the media. Medium centrality is the frequency that a node appears on the other nodes to contact the path, and measure the degree of the connection between the nodes as the other nodes. Terrorist activities, a node that has a low degree but it contacts with a terrorist usually has a high degree of mediation, such a person plays the role of a broker or a goalkeeper to control the communication and exchange of information. [13] [15] [16] [17] But in the real environment, communication is often not strictly in accordance with the network path, in view of this situation, some scholars put forward two kinds of methods to measure the Intermediate Center. The first one is based on the calculation of all the possible paths between a pair of nodes, and the second is based on the random path calculation.

At present, the most common definition of the center of the center is:

$$C_B(k) = \sum_{i \neq j \neq k \in V} \frac{\sigma_{ij}(k)}{\sigma_{ij}}$$

Among them，σ_{ij} is the number of shortest paths between B and S, $\sigma_{ij}(k)$ is the shortest path through the number of nodes K between i and j.

However, in the real world, a pair of nodes often has many paths, which makes the calculation of the center of the media becomes very complicated. Similarly, the computation of $C_B(k)$ is also influenced by the network size. By dividing the K by dividing the number of nodes outside the $(n-1)(n-2)$ can be standardized, the relative center degree is obtained. Freeman gives the definition of relative centrality:

$$C'_B(k) = \frac{C_B(k)}{(n^2 - 3n + 2)/2}$$

(4) Feature vector center degree. The concept of feature vector centrality is firstly proposed by Bonefish. Subsequently, the PageRank algorithm of Google improved it. The nodes defined by the center of the feature vector are the nodes which are in the central position in the central position. Let express an Adjacency matrix of $n \times n$. If the node i and node j by the line directly connected, then $a_{ij} = 1$; Conversely，$a_{ij} = 0$.Because A is a symmetric matrix, so its eigenvalues are real numbers, the feature vectors are orthogonal and can be used to be of the right. Then the feature vector of the center node xi is defined as the maximum eigenvector of the A in the i standard. The λ

for maximum eigenvalue, the corresponding eigenvector x, then

$A\lambda = \lambda x$, then $x = \frac{1}{\lambda} Ax$, and as follows:

$$xi = \mu \sum_{j=1}^{n} a_{ij} x_j$$

Proportional coefficient $\mu = \frac{1}{\lambda}$, This xi is proportional to the center of the eigenvector of the other nodes connected to it.

(5) Network efficiency. Complex networks can be understood from the point of view of efficiency, For example, it is believed that a network with a small world character is very efficient in information transmission. Network efficiency $E(G)$ is a measure of the nodes in a network to exchange information efficiency indicators. In order to define the efficiency of M, we first calculate the shortest distance $d_{i_0,j}$ between nodes J and I. Assume that each node sends a message through their contact. The communication efficiency of the node i and j is inversely proportional to its shortest distance.

When there is no path between the nodes 'i' and 'j', $d_{i_0,j} = +\infty$, and the network efficiency is 0.Suppose N is the number of nodes in the network size or graph theory, The average efficiency of graph theory (Network) G can be defined as:

$$C_{eff} = E(G) = \frac{1}{N(N-1)} \sum_{i \neq j \in G} \frac{1}{d_{ij}} (C_{ef} E \in [0,1])$$

From the above formula can be obtained in the $[0,1]$ interval value.

5.1.2 Individual Risk Index Calculation in who have been in contact with terrorist

The traditional social network analysis method often evaluates the key figures only through a series of network indicators, and do not consider the individual with the exposure of terrorism has its own characteristics.

This is a good way to ensure the objectivity of the evaluation results, but in the practical application, it will cause the one-sided and inaccurate. We learn from key figures evaluation method proposed by D. M. Akbar Hussein. On this basis, we introduce the degree coefficient of event participation and the individual risk index evaluation system to measure and calculate the risk index, and then evaluate the key figures. [19]

1. Property classification and evaluation methods. Different methods for classifying and assigning attributes are the key to build individual risk index evaluation system. First of all, the classification method of the attributes is explained by the religious attributes (See table below).

In comparison with those who have no faith, religion is more threatening.

Table 3. Religious attributes

Attribute name	Possibility	Fraction
Religion	A	4
	B	3
	C	2
	Faithless	1

In order to avoid controversy, we use A, B, C to replace different religions and specific religion can be selected

according to the specific circumstances in the actual analysis. In the above table, we divide our religious attributes into 4 levels, that is, A, B, C, and no religion. Based on previous experience of terrorist incidents, it can be found that the terrorists often carry out terrorist activities in the name of religion.

In comparison with those who have no faith, religion is more threatening. At the same time, in the case of specific terrorist incidents, such as religious conflicts caused by terrorist incidents, terrorists tend to have a common religious belief. In this case, people who believe in a particular religion are more likely to be terrorists than those who believe in other religions. Therefore, for a terrorist event, it can be considered that the people who have his belief of a religion (such as A) are more threatening and more likely to become a key figure. Therefore, for religious A, we give the highest score - 4 points. A or B except for C, we give the lower score-3 or 2, respectively. The people who have no faith are considered to be the least threatening, so it is given the lowest score of 1 point. [21] [22]

It should be noted that individuals engaged in terrorist activities are often concealed. Due to the impact of the national intelligence agencies, their information may be difficult to obtain. If the specific information of a terrorist is not obtained, then the attribute is given a minimum of 1 point. In another example, the table below grade social landscape attributes as follows:

Table 4. Social features

Attribute name	Possibility	Fraction
Social features	(suspected) Terrorism	5
	Violence / Racism / Drug / Crime	4
	Solitary / Independent	3
	General	2
	Good	1

Psychological indicators are shown in the following table:

Table 5. Psychological Index property

Attribute name	Possibility	Fraction
Psychological Index	Violent mental illness/ Serious violence obsessive-compulsive disorder / Mental illness delusions of persecution	4
	With excitement tendency of psychological disease/With mild, moderate excited state of mind	3
	Have depression type mental disease / Have other mild mental disease	2
	Mental normal / Mental disorders	1

The comprehensive measurement attributes are shown in the following table:

Table 6. Comprehensive measurement attribute

Attribute name	Possibility	Fraction
Comprehensive measurement	A lot of contact with the nature of the content related to terrorism	5
	Having contact with the contents of a terrorist nature	4
	Who do not have contact with the terrorist nature of the content / Who have contact minimal nature of the content.	1

The rest of the property will not go; their design principle is the same.

Of course, these selected properties are not absolute. Depending on terrorist activities and their structure, we can adjust the content of the above indicators and make it more accurate estimates.

2. Individual Risk Index calculation steps.

Individual risk index calculations are divided into the following steps:

1) The social attributes and network properties that we select are subdivided into several stages and we give a different score for each level. On this basis, we will build individual risk index evaluation system.

2) According to the information collected from the members of the terrorist activities, we will use the individual risk index evaluation system to score;

3) We can calculate the risk index of each node of the network by using the attribute values of each member to sum and multiplying its corresponding event participation degree coefficient. By comparison, we can determine the position and importance of each node in the network.

Expressed by the formula:

$$I = P \square \sum_{i=1}^{n} (S_i + N_i)$$

Finally, the key figures in the network can be obtained by comparing their individual risk index. It can be considered that the individual risk index value is the key person in the network. It is worth noting that they are not necessarily at the center of the network.

5.2 A Double Classification Data Model Based on Fuzzy Clustering Model and the Bayesian Classification Model with Weighted

From the point of view in behavior, the individual long-term network behavior has a certain stability. And may correspond to social reality people's social behavior. Therefore, analysis of network behavior is important. However, depending on the focus of research and application, the classification method is different. So, faced with a particular problem, we need to classify from a particular angle.

In the face of a large amount of data, we preprocess the data first, using small sample to get the distance matrix, and then the fuzzy clustering method for a large number of data in the network behavior of the initial classification. Then use the weighted Bias classification method to set up the key words in the sample, and then segment the data which has already been classified, and filter out the content we need. [23] [24]

5.2.1 Data pre-processing
Take out the data from the 100 IP addresses, from a pair of multiple perspectives, the corresponding behavior of each IP address to get out, a table as shown below. That each action corresponds to a value. The value should be determined according to the actual situation, for determining the classification of indicators.

As shown in the table below, the X_1、X_2、X_3….X_n score table for different user behavior observation points:

Table 7. User score table for different behavior observation points

Behavior / User	X_1	X_2	X_3	…	X_n
A	a_1	a_2	a_3	…	a_n
B	b_1	b_2	b_3	…	b_n
C	c_1	c_2	c_3	…	c_n
D	d_1	d_2	d_3	…	d_n

First, the average value of the sample of each behavior is calculated:

$$\overline{x}_k = \frac{1}{n} \sum_{i=1}^{n} x_{ki}$$

And standard deviation:

$$S_k = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_{ki} - \overline{x}_k)^2}$$

Secondly, the standard value of each data XXX, and the standardization of data compression to YYY closed interval:

$$x_{ki}^{''} = \frac{x_{ki}^{'} - x_{k\min}^{'}}{x_{k\max}^{'} - x_{k\min}^{'}}$$

（$x_{k\max}^{'}$ and $x_{k\min}^{'}$ said the maximum value and the minimum value respectively）

5.2.2 Fuzzy clustering behavior pattern classification
Now, x_i and x_j are two of the data that have been processed, and the function d（x_i，x_j） satisfies the following conditions:

1. Non negative：

2. To all x_i and x_j, d（x_i，x_j）=0 $\Longleftrightarrow$ x_i=x_j

3. Symmetry: for all x_i and x_j, d（x_i，x_j）= d（x_j，x_i）

4. Triangle inequality: for all x_i，x_j，x_k, d（x_i，x_j）+ d（x_j，x_k）≥d（x_i，x_k）

Then, in the p-dimensional Euclidean space, the Euclidean distance between any two elements I and j is defined as the square root of the difference of the vector:

$$d(x_i, x_j) = [\sum_{s=1}^{p} (x_{is} - x_{js})^2]^{1/2}$$

In order to avoid the large difference between the results of a one-dimensional distance, it is necessary to carry out a linear transformation, the standard of all variables, each variable will be subtracted from the mean, then divided by the square root. After that, each variable has zero mean and unit variance, with the same weight to participate in the distance calculation.

In order to obtain the grouping of the vector of all network user behavior, the distance of all the vectors is usually considered, and the distance of the 100 samples is also expressed as the distance of the sample:

$$\begin{pmatrix} 0 & & & & \\ d(2,1) & 0 & & & \\ d(3,1) & d(3,2) & 0 & & \\ \cdots & \cdots & & 0 & \\ d(100,1) & d(100,2) & d(100,3) & \cdots & 0 \end{pmatrix}$$

Among them, d（x_i, x_j）represents the distance between the behavior of i and the behavior of j.

After that, the cut sets of fuzzy matrices are defined, and the classification of the large numbers of data can be classified:

Table 8. The Definition Sample of Cut Set of Fuzzy Matrix

Category	Access to Information	Communication	Leisure and Entertainment	E-commerce	…
Body behavior	News, information acquisition, work related information, search engines, electronic journals, online search, browse the web, read information, online	Online chat, email, SMS, BBS forum, community, discussion group, free personal home page, Internet phone, alumni, RENREN, Wechat, Tieba… …	Online games, multimedia entertainment, video viewing, live online, online reading ……	Online education, software upload / download services, online hospital, online banking, online design ……	…
	reading ……				
Keywords	a_1、$a_2 ... a_n$	b_1、$b_2 ... b_n$	c_1、$c_2 ... c_n$	d_1、$d_2 ... d_n$	…

5.2.3 Weighted Bayesian behavior pattern analysis

In this context, we preprocess the raw data and classify it based on fuzzy clustering. By the above classification, we can know what behavior is the user's risk behavior through the data acquired by the large data. On this basis, we will again through the Bias classifier to the specific behavior of the key words for a classification. This classification is mainly to have good user behavior and the abnormal behavior of the two classification. In order to effectively apply the data after the second classification, we will use the method based on the weighted Bias classification to analyze the behavior patterns that have been collected. The expression of the discriminant function is as follows:

$$e(x) = \arg \max_{1 \le j \le m} p(C_j) \prod_{k=1}^{d} p(x_i \mid C_j)^{v_k}$$

Based on the reparability judgment, we can obtain the final weights. We use JF(k) to represent the function, which means that the weight of the right of the classification of each feature.

We use $D = (z_1, z_2, \cdots, z_n)$ to represent the specific behavior of the keyword set (the following abbreviation for the sample word set). $A_1, A_2, \cdots, A_d$ represents a set of continuous variables. $C = \{C_1, C_2, \cdots, C_m\}$ is used to represent the set of samples, and j is expressed by $(z_1^{(j)}, z_2^{(j)}, \cdots, z_{n_j}^{(j)})$. And $j = 1, 2, 3, \cdots, m$. For any attribute A_k, it is compared with the variable of the class object concentration. The expression of the C_j component of A is calculated as follows :

$$\mu_{jk} = \frac{1}{n_j} \sum_{j=1}^{n_j} z_{jk}^{(j)}$$

The total mean value is:

Table 9. Weighted Bias classification weight based on fuzzy classification of specific behavior

	a_1	a_2	$\cdots$	a_n	b_1	b_2	$\cdots$	b_n	c_1	c_2	$\cdots$	c_n	d_1	d_2	$\cdots$	d_n	$\cdots$
Access to Information	w_1	w_2	$\cdots$	w_n	0	0	$\cdots$	0	0	0	$\cdots$	0	0	0	$\cdots$	0	$\cdots$
Communication	0	0	$\cdots$	0	w_{n+1}	w_{n+2}	$\cdots$	w_{2n}	0	0	$\cdots$	0	0	0	$\cdots$	0	$\cdots$
Leisure and Entertainment	0	0	$\cdots$	0	0	0	$\cdots$	0	w_{2n+1}	w_{2n+2}	$\cdots$	w_{3n}	0	0	$\cdots$	0	$\cdots$
E-commerce	0	0	$\cdots$	0	0	0	$\cdots$	0	0	0	$\cdots$	0	w_{3n+1}	w_{3n+2}	$\cdots$	w_{4n}	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$

$$\mu_k = \frac{1}{n}\sum_{j=1}^{n} z_{jk}$$

Class distance difference:

$$SB(k) = \sum_{j=1}^{m} n_j (\mu_{jk} - \mu_k)^2$$

Class within the difference:

$$SW(k) = \sum_{j=1}^{m}\sum_{i=1}^{n_j} (z_{ik}^{(j)} - \mu_{ik})^2$$

The value of JF(k) can be obtained:

$$JF(k) = \frac{SB(k)}{SW(k)}$$

In this function, SB (k) is used to describe the dissimilarity between the class t class and SW (k) is used to describe the dissimilarity of different samples within the same class. According to the definition of the criterion function, we can infer that the value of JF (k) is becoming more and larger and belongs to the same class. At the same time, the degree of similarity between different samples is lower. With the increase of the value of JF(k), the classification of its types is gradually enhanced.

For this function, when the value of JF is above the average, in order to improve the function of it, we should increase the weight to more than 1. When the value of JF is less than the average, the weight should be reduced to less than 1 in order to weaken the function of tithe method of establishing weights is given below:

Assuming the weight of Fisher is v_k :

$$v_k = \frac{JF(k)\Box d}{\sum_{k=1}^{d} JF(k)}$$

So the definition of weighted Bias classifier based on Fisher criterion function is as follows:

$$P(C_j \mid y) = \frac{\alpha_j \prod_{k=1}^{d} \frac{1}{\sqrt{2\pi}\sigma_{jk}} \exp\left\{-\frac{(y_k - u_{jk})^2}{2\sigma_{jk}^2}\right\}^{vk}}{\sum_{j=1}^{m} \alpha_j \prod_{k=1}^{d} \frac{1}{\sqrt{2\pi}\sigma_{jk}} \exp\left\{-\frac{(y_k - u_{jk})^2}{2\sigma_{jk}^2}\right\}^{vk}}$$

By using the above model, we have made the two processing of the data that have been processed before the specific behavior of the key words. So we can get the results, and be able to judge the situation of the regional terrorism related communication activities, as well as the impact of the key social or network factors.

In the table, the value $w_1, w_2, \cdots, w_n, \cdots, w_{4n}$ presents the weights given by each keyword after the classification has been classified by weighted Bias. Based on the weight of keywords, we can know what the specific content is similar to this kind of. In using this model, we will prepare a part of the similar sample data to be simulated and get the weights.

Therefore, the weights and the results from the previous training simulation are compared with the results of the previous training simulation, which can find out the individual and regional terrorist activities. Title and Authors

The title (Helvetica 18-point bold), authors' names (Helvetica 12-point) and affiliations (Helvetica 10-point) run across the full width of the page – one column wide. We also recommend e-mail address (Helvetica 12-point). See the top of this page for three addresses. If only one address is needed, center all address text. For two addresses, use two centered tabs, and so on. For three authors, you may have to improvise.

6. MODEL SUMMARY AND IMPROVEMENT

With the improvement of living standards, network security has become a hot issue of social concern, with the help of the network, the activities of terrorist organizations have been greatly expanded, and the research of network user behavior has become an important research topic. In the first question, we established the evaluation model of the individual risk index based on the social network analysis, and described the evaluation of the key figures in the terrorist network. In this second question, we first preprocess the data, and then we establish a fuzzy matrix model and a data classification model based on weighted Naive Bayesian classification to achieve a fast data classification, and we can get the information of each weights, laying the foundation for the next step in the analysis of personal information. This is of great significance to the management of network security. The advantages and disadvantages of the two models in this paper are as follows:

6.1 The advantage & disadvantage of Evaluation Model in Individual Risk Index

Advantages:

1. The personal attributes are classified into social attributes and network attributes, which are described, and the corresponding evaluation index system is established. We can take advantage of

the rank difference of two attributes; analyze what kind of activity that a person has to do, in order to make the concrete action to prepare.

2. In this model, a Meta network (human relationship network) and multi - Network (task network, skill network) are described.

Disadvantages:

1. This paper is based on the static network model, and it is from the perspective of the individual, it can be further extended by the key individuals to find a key group, from the perspective of dynamic network to construct a network model of terrorist groups.

2. Because this model relates to social and network two attributes, so in this based ranking, it is possible to weaken the influence of a certain attribute, and thus cause the error.

6.2 The advantage & disadvantage of Data Classification Model

Advantages:

1. The weighted Naive Bayesian algorithm can estimate the required parameters based on a small amount of training data. Based on these, we can classify a large number of data.

2. The weighted Naive Bayesian algorithm is the simplest algorithm in the Bias classification algorithm and it can be obtained by the data after the classification of the data on the basis of each data classification.

3. It has great flexibility for different types of data and different requirements of data classification. It also can establish a relatively stable classification model and not rigidly adhere to several data classification. And according to the actual needs of information classification, the model can be improved.

4. Bias network has a strong ability to deal with uncertainty problem. Bias network uses conditional probability to express the relationship between the various elements of information. It can learn and reason under the conditions of limited, incomplete and uncertain information. So we can effectively integrate information with the key words of our hypothesis.

Disadvantages:

1. In this model, it can be used to classify the data in the algorithm. So it is not possible to achieve the actual application. There will always be some errors.

2. The use of this model is to know a prior probability. It means that the model only can be obtained by a large number of data learning. This requires that a large amount of data is required to train the model to improve the accuracy of the model.

3. Classification decision error rate. The main reason for the induction of these errors is due to the problems in the training process. The main reasons

may be that the model design is unreasonable; the model training data is not sufficient or not typical and so on.

6.3 The advantages and disadvantages of fuzzy matrix classification model

Advantages:

1. Many concepts in real life are very difficult to distinguish between "with" and "no". Therefore, fuzzy clustering is used to describe the relationship between the behavior and the effect of the initial classification.

Disadvantages:

1. The selection of the fuzzy matrix has some errors, too large and too small can lead to classification error.

7. REFERENCES

[1] Popp R, Poindexter J. Countering terrorism through information and privacy protection technologies [J]. Security & Privacy, IEEE, 2006, 4(6): 18-27.

[2] Liu X. P. Modeling method for terrorist social network based on social network analysis [D]. Changsha: National University of Defense Technology, 2010.

[3] Shang R., Sun D. Y. A key figure in the evaluation based on the individual threat index [J]. Journal of Wuhan Public Security Cadre's College, 2011, 25(4): 83-87.

[4] Shang R. Individual Threat Index Based Key Member Detecting Research in Covert Network [D]. National University of Defense Technology, 2011.

[5] Long J. A Study on Tag-based Folksonomy on the Internet [D]. Beijing: Peking University, 2007.

[6] Zhang H. A Study on Terrorist Organization from the Perspective of Social Network Analysis [D]. Changsha: National University of Defense Technology, 2010.

[7] Wang Q. W., Deng X. Z. A Study on Mental Models of Web Users' Information Behavior in Browsing [J]. Document, Information and Knowledge, 2010 (5): 93-96.

[8] Rong G. J. Mathematical Models of Relationship between Psychology and Behaviors in Clothing Wearing and Purchasing [J]. Journal of Beijing Institute of Clothing Technology, 1989, 1: 006.

[9] Liu T., Bu ZH. X., Wang G. H. Based on psychological factors and behavioral patterns of consumers brand choice model [J]. Enterprise economy, 2015, 7: 009.

[10] GUO Z. Y. On the social Psychological model of the Psychological structure of morality [J]. Journal of Shaanxi Normal University (Philosophy and Social Sciences Edition), 2000, 29(2): 151-157.

[11] GAO W. B., Chen Z. Y. A Study on Psychopathology and Psychotherapy of Internet Addiction [J]. Advances in Psychological Science, 2006, 14(4): 596-603.

[12] Xue L. Study of Mental Pattern and Utilization of Social Support of Drug Addicts. [J]. China Journal of Health Psychology, 2007, 15(9): 833-835.

[13] Wan H. Y. Research on Link-based Classification in Social Networks [D]. Beijing Jiaotong University, 2012.

[14] DING H. J., CHEN D. J. ISIS's Cyber Terrorism Activity Impacts on China's Counter-terrorism and the

Countermeasures [J]. China Public Security. Academy Edition, 2015 (2): 95-97.

[15] Conway M. Terrorist 'Use' of the Internet and Fighting Back [J]. Information and Security, 2006, 19: 9.

[16] Kuang W. B. On characteristics of netizens' behavior in special times [J]. Journal of International press, 2003, 5: 004.

[17] Zhou R. T., Zheng B. Folksonomy New Information Classification System in Internet Age [J]. New Technology of Library and Information Service, 2006 (3): 72-75.

[18] Zhang N. Analysis of Folksonomy System Users' Behavioral Characteristics [D]. Taiyuan: Shanxi University, 2013.

[19] Li Y. G., Li X. CH. Area information of Measure Model and its application based on variable weights [J]. Journal of Intelligence, 2006, 25(2): 107-109.

[20] Li D. D. Research on the Methods of Network Traffic Classification Based on Behavior [D]. Harbin University of Science and Technology, 2015.

[21] Ren W. J, Chen J. Y. Research of web-based user behavior analysis [J]. 2012.

[22] Ma G., Du Y. G., An B., etc. Risk Evaluation of Complex Information System Base on Threat Propagation Sampling [J]. Journal of Computer Research and Development, 2015, 52(7): 1642-1659.

[23] Song N., Fu J. L., Bao Q., etc. Cyber terrorism spreading and optimal intervention policies based on a scale-free network [J]. Systems Engineering—Theory & Practice, 2015, 35(3): 630-640.

[24] Liu CH. Y., Hu X. F., Si G. Y., etc. Public Opinion Propagation Model Based on Small World Networks [J]. Journal of System Simulation, 2007, 18(12): 3608-3610.

[25] Li Q., He H., Fang B. X., etc. Awareness of the Network Group Anomalous Behaviors Based on Network Trust [J]. CHINESE JOURNAL OF COMPUTERS, 2014, 37(1): 1-14.

[26] Zou L. J. Research on Clustering and Classification Algorithm of Streaming Data [D]. Yangzhou: Yangzhou University, 2008.

[27] Sui ZH. Research and Implication of Group Network Behaviors Model [D]. Sichuan: Sichuan University, 2006.

[28] Li ZH. H., He J. L., Wu P. F. The Time Period Characteristic of Information Communication Model and its Management Strategies of Paroxysmal Public Crisis [J]. LIBRARY AND INFORMATION SERVICE, 2007, 51(10): 0-0.

[29] Dong F. Q. Study on Network Users' Behaviors Analysis and its Application [D]. Xi'an: Xidian University, 2005.

[30] Ji R. M. RESEARCH AND DEVELOPMENT OF WEBPAGE TEXT KEYWORDS FILTERING [D]. Inner Mongolia University, 2012.

[31] Yao ZH. W. Dissemination of modern terrorism steering [J]. Journal of Yunnan Social Sciences, 2012, 2: 026.

[32] Dai Y. RESEARCH AND DESIGN OF USER BEHAVIOR ANALYSIS AND CLASSIFICATION IN UNIVERSAL NETWORK [D]. Beijing University of Posts and Telecommunication, 2013.

Evaluation and Influence on the Development Strategy of the Maritime Silk Road

Ning Zhang
College of Information Science and Engineering
Shandong University of Science and Technology,
Qingdao, Shandong, China

zning@sdust.edu.cn

Jing Tang
College of Information Science and Engineering
Shandong University of Science and Technology,
Qingdao, Shandong, China

493117148@qq.com

Abstract: This paper studies the economic and trade of Maritime Silk Road from the perspective of trade facilitation. Through the establishment of multiple linear regression model and trade gravity model, the paper makes a quantitative evaluation on the impact of the Silk Road on the sea. The value of trade facilitation is constantly improving. From this we give the suggestion that in the domestic resources to improve the range of 5 points of trade facilitation, if the limited resources, it should be based on the relevant calculation, determine which trade facilitation of the contribution of the sub index, priority to the use of resources in a greater impact on the trade facilitation of points. Specifically, governments should increase the investment in the infrastructure of trade facilitation, not only the construction of hardware facilities, but also the investment of software facilities.

Keywords: Maritime Silk Road; Strategic Policy; Multiple Linear Regression Model; Trade Gravity Model; Evaluation and Impact; SPSS Analysis

1. INTRODUCTION

In October 2013, during his visit to the countries of Southeast Asia. Chairman xi speech, said Southeast Asia since ancient times is the important hub "Marine silk road". So far, China's strategic plan of "One Belt And One Road" is ripe, the Silk Road economic belt and the maritime silk road in the 21st century. [1] [2] [3]

The maritime Silk Road in the 21st century should be a global network based on trade, also is the international trade network. Tariff has deep cuts in the countries all over the world, how to use trade facilitation to promote trade both sides of the economic development, it is theoretical issues of common concern to all countries in the world. [9] [10] As a result, the new sea route of the silk road is a dynamic rather than static, not only covers China and Southeast Asia, South Asia and west Asia and East Africa, should also be extended to Australia, and North America, and through this international trade network reflect the economic and trade relations between China and the countries of the world trade network.

Construction and development of the Marine silk road, should look at the goods trade as the main means, make good use of the Chinese government to set up the China-Asian maritime cooperation fund, based on the sea, developing good Marine cooperation partnership, deepening and extension of the China-Asian free trade area, promote regional economic and security economic and security.

2. THE DESCRIPTION OF THE PROBLEM

In our paper, we will analyze the following contents:

1. In this paper, we will refer to the relevant information. And Select the most important factor in the maritime Silk Road - economic trade, as the focus of this paper and analyze the relevant indicators. Based on the multiple linear regression model and the trade gravity model, we study the development strategy of the Silk Road on the sea.

3. THE HYPOTHESIS OF PROBLEM

We assume that:

1. All of the countries have typical representative, and they have linked with Maritime Silk Road.

2. As mentioned in this article, the distance between countries, by default, is replaced by the distance between the capitals of the two countries.

4. THE DESCRIPTION OF SYMBOL

Table 1. The Description and Explanation of Symbol in this Paper

Symbol	Description and Explanation
CC_j	Comprehensive index of trade facilitation.
$POLICY_{ij}$	Bilateral trade environment, if the member state is the Maritime Silk Road along the country, the value is 1, otherwise it is 0.
TAR_j	The average tariff of Country j, it's the average value of the tariff rate for all trade goods.
BOR_{ij}	BOR_{ij} is a dummy variable that Partner j and Chinese bordering the territory when the value is 1, otherwise it is 0.
Y_i, Y_{max}	Y_i is the original data, Y_{max} is the maximum value of the index.
Z_i	$\frac{Y_i}{Y_{max}} = Z_i$, Z_i is standardized data after treatment.
$Trade_j$	Trade volume between Country i and j.

Symbol	Description and Explanation
GDP_i	Gross Domestic Product of Country i.
GDP_j	Gross Domestic Product of Country j.
$PGDP_i$	Per capita Gross Domestic Product of Country i.
$PGDP_j$	Per capita Gross Domestic Product of Country j.
JL_{ij}	Distance between Country i and j.

5. MODELING

5.1 Modeling of Trade Evaluation Model in Maritime Silk Road

In recent years, with the dynamic social network analysis, we can find complex covert network, and through a series of algorithms to estimates people who have been in contact with terrorism individuals risk index, in order to determine a key figure what we need. And then take corresponding measures to solve it. [4] [5]

5.1.1 Theory Building

Summarized above measure of trade facilitation in the literature review has the following categories: 1. Port Efficiency (KA) 2. Customs environment 3. Regulatory and institutional environment 4. Marketing Factors 5. Infrastructure.

In this paper, based on the above five indexes to study China and trade between the countries along the maritime silk route to measure the level of trade facilitation, and then through the contrast can be concluded that the maritime silk road trade development strategy is put forward for China. Us to the five indicators, digital processing, multiple linear regression analysis by SPSS can get all kinds of factor value, the greater the number, shows that the index of the greater the degree of influence on trade facilitation.

1. Port Efficiency (KA)

Port efficiency refers to the efficiency of the operations of ports, airports and other infrastructure, the index of coefficient value is higher, said the airport, the port cargo throughput and cargo loading and unloading, handling, storage and the higher the efficiency. Specific indicators include Quality of port infrastructure (P1) and Quality of air transport infrastructure (P2)

2. Customs environment (TE)

The customs environment refers to the cost of the cost and time needed for customs clearance, and under the customs law and regulations, the size of the rent-seeking space, specific include:

(1) Irregular payments and bribes (C1). The indicator to measure whether in order to obtain the relevant import and export license to the customs personnel for rent-seeking and the resultant costs or is the enterprise of planned spending. 1 for irregular payments, often values 7 said had never happened in the abnormal phenomena of payment.

(2) Ethics and corruption (C2). People in the field of the indicators is related to a country customs corruption degree of evaluation, the numerical value range of fluctuation represent different meanings. 1 minute, serious corruption, 7 points, said there is no corruption, very honest.

(3) Burden of customs procedures (C3). The customs declaration formalities trival degree is not the same. 1 minute

said customs formalities is extremely complicated and 7 points said entry is simple and efficient.

(4) Degree of customer orientation (C4). The indicator to measure a country's customs overall service capacity, the higher the score represents the stronger service ability, the highest is divided into seven points.

3. Regulatory and institutional environment (LE)

A country's laws and regulations and the system more transparent, the smaller the rent-seeking space, import and export business, the available in a more transparent society under the rule of law for international trade.

(1) Transparency of government policymaking (R1). The higher the score the higher the transparency of the on behalf of the country's law. The highest is divided into seven points.

(2) Efficiency of legal framework in settling disputes (R2). The higher the score on behalf of the country's laws and regulations are strictly enforced. 1 minute means the country did not perform the relevant laws and regulations.

4. Extent of marketing (DZ)

The coming of the era of big data, making the number of the marketing factors are increases. Companies use different marketing factors, can be more convenient to keep in close contact with customers, suppliers, customs, reduce the risk caused by information asymmetry.

5. Infrastructure (TC)

Infrastructure of transportation problem is a measure of export enterprises for the country's logistics department of transport and logistics ability indicators. Specific indicators are as follows:

(1)Quality of roads (A1). The higher the score, the smaller the transportation time and cost.

(2)Quality of railroad infrastructure (A2). The higher the score, the smaller the transportation time and cost.

Table 2. Trade facilitation measure

Level indicators	Secondary indicators	Three indicators
Port Efficiency(KA)	Quality of port infrastructure(P1)	1-7
	Quality of air transport infrastructure(P2)	1-7
Customs environment(TE)	Irregular payments and bribes(C1)	1-7
	Ethics and corruption(C2)	1-7
	Burden of customs procedures(C3)	1-7
	Degree of customer orientation(C4)	1-7
Regulatory and institutional environment(LE)	Transparency of government policymaking(R1)	1-7
	Efficiency of legal framework in settling disputes(R2)	1-7

Level indicators	Secondary indicators	Three indicators
Extent of marketing (DZ)	Extent of marketing(DZ)	1-7
Infrastructure(TC)	Quality of roads(A1)	1-7
	Quality of railroad infrastructure(A2)	1-7

Note：Data from the 2015 "Global Competitiveness Report" http://www3.weforum.org/docs/WEF_GlobalCompetitiveness Report_2014-15.pdf

5.1.2 Required Data for Standardization

Because the measure index of trade facilitation is taken in the case of time range is not unified, the data are not comparable. As a result, the need for data processing, standardization of data obtained, to use, specific USES the following practices, remember to take a secondary index, on behalf of the original raw data, on behalf of the index can achieve maximum value, and is standardized processing of data, finally will be all kinds of secondary standardization index level resulting from the weighted average index. [6] [7] [8]

Port Efficiency (KA) = (Quality of port infrastructure (P1) +Quality of air transport infrastructure (P2))/2

Customs environment(TE)=(Irregular payments and bribes(C1)+Ethics and corruption(C2)+Burden of customs procedures(C3)+Degree of customer orientation(C4))/4

Regulatory and institutional environment (LE) = (Transparency of government policymaking (R1) +Efficiency of legal framework in settling disputes (R2))/2

Infrastructure (TC) = (Quality of roads A1+ Quality of railroad infrastructure A2)/2

The comprehensive indicators of trade facilitation (CC)=(KA+TE+LE+DZ+TC)/5

Data calculated by later learned that within the ASEAN Trade Facilitation level uneven levels, such as Singapore trade facilitation within the overall score is much higher than other ASEAN countries.

5.1.3 Model Establishment

We introduce the standard model of international trade - the gravity model. And we set the gravity model is a natural log-linear form, the basic model is set as follows:

$$\ln Trade_{ij} = c_0 + c_1 \ln GDP_i + c_2 \ln GDP_j + c_3 \ln PGDP_i + c_4 \ln PGDP_j + c_5 \ln JL_{ij} + u_{ij}$$

In this paper, we study between China and countries along the maritime Silk Road trade situation, selecting the section data of 2014, according to our research purpose, the modified gravity model:

$$\ln Trade_{ij} = c_0 + c_1 \ln GDP_j + c_2 \ln PGDP_j + c_3 \ln JL_{ij} + \alpha \ln CC_j +$$
$$c_4 \ln TAR_j + c_5 POLICY_{ij} + c_6 BOR_{ij} + u_{ij}$$

Compared with type(1) , can be found as follows: First, we examine the trade between China and other countries, and therefore China's data is constant, there is no effect on the results of the regression model, the coefficient variable is 0, can be removed from the model; Secondly, the trade facilitation comprehensive index into the model, indicating the bilateral trade environment, if Member States are members of the China-ASEAN Free Trade Area, which takes a variable value, and 0 otherwise, and therefore, the variable is discrete variables; Finally, the former represents the average tariff j country,

namely the average of all trade in goods tariff rates. Dummy variables, when its value is 1 when the show j and the Chinese territory bordering the partner countries, if the value is 0, indicating no border with the Chinese territory. Related to the specific meaning of the variables as shown in Table 1: The Description and Explanation of Symbol in this Paper in IV. The Description of Symbol.

5.2 Application and Conclusion of Trade Evaluation Model in Maritime Silk Road

5.2.1 Comprehensive Application and Improvement in Multivariate linear regression model and Trade Gravity Model

1. The Basic Regression Model

Using the least squares method, using SPSS 19.0 software basic model, the regression results in Table 3:

Table 3. Variable regression model results

Independent variable	Regression Coefficients	T-Statistics
C-	-	0.925
$\ln GDP$	0.943	0.006
$\ln PGDP$	0.091	0.611
$\ln JL$	0.073	0.729

Note: According to the results of the regression SPSS 19.0 software finishing.

Available from the results, it is 0.887; F=15.696, P=0.003 < 0.01. Thus available, the explanatory power of the model is acceptable and relatively stable. GDP explanatory variables and regression coefficients of large absolute PGDP before, indicating the influence of the explanatory variables is large, but also through the t value judgment, these variables are significant.

2. Gravitational regression model modification

The trade facilitation index CC is introduced into the regression model, can be obtained in Table 4:

Table 4. Correction of regression results

Independent variable	Regression Coefficients	T-Statistics
C		0.041
$\ln GDP$	0.415	0.854
$\ln PGDP$	-0.114	-0.563
$\ln JL$	-0.120	-0.501
$\ln CC$	0.446	2.186
$\ln TAR$	-0.053	0.331
POLICY	0.288	-0.600
BOR	0.011	0.060

Available from the results, it is 0.982; F=16.026, P=0.006 < 0.01,After the adoption of contrast Tables 3 and 4, the introduction of trade facilitation indicators, whether or value of F are improving, the explanatory power of the model described improved than before, And after analysis of explanatory variables, we found GDP is still the key variables that affect trade, but trade facilitation CC coefficient is 0.446.Description: trade facilitation indicators for each one percentage point to enhance the volume of trade will double, you can observe, before the tariff coefficient is negative, the reduction of tariffs

to promote trade, but the effect is far less than trade facilitation has brought results.

5.2.2 The impact of trade facilitation in the twenty-first Century Maritime Silk Road

In order to analyze the impact of different trade facilitation measures on the trade volume, in order to make a more targeted policy recommendations, we have to amend the gravity model to join 5 trade facilitation indicators.

$$\ln Trade_{ij} = c_0 + c_1 \ln GDP_j + c_2 \ln PGDP_j + c_3 \ln JL_{ij} + \alpha_1 \ln KA_j +$$
$$c_4 \ln TAR_j + c_5 POLICY_{ij} + c_6 BOR_{ij} + u_{ij}$$

$$\ln Trade_{ij} = c_0 + c_1 \ln GDP_j + c_2 \ln PGDP_j + c_3 \ln JL_{ij} + \alpha_2 \ln TE_j +$$
$$c_4 \ln TAR_j + c_5 POLICY_{ij} + c_6 BOR_{ij} + u_{ij}$$

$$\ln Trade_{ij} = c_0 + c_1 \ln GDP_j + c_2 \ln PGDP_j + c_3 \ln JL_{ij} + \alpha_3 \ln LE_j +$$
$$c_4 \ln TAR_j + c_5 POLICY_{ij} + c_6 BOR_{ij} + u_{ij}$$

$$\ln Trade_{ij} = c_0 + c_1 \ln GDP_j + c_2 \ln PGDP_j + c_3 \ln JL_{ij} + \alpha_4 \ln DZ_j +$$
$$c_4 \ln TAR_j + c_5 POLICY_{ij} + c_6 BOR_{ij} + u_{ij}$$

$$\ln Trade_{ij} = c_0 + c_1 \ln GDP_j + c_2 \ln PGDP_j + c_3 \ln JL_{ij} + \alpha_5 \ln TC_j +$$
$$c_4 \ln TAR_j + c_5 POLICY_{ij} + c_6 BOR_{ij} + u_{ij}$$

After correlation calculation, as shown in Table 5.

Table 5. Trade growth forecast after the implementation of trade facilitation (%)

Economy Group	KA	TE	LE	DZ	TC
Brunei Darussalam	28.85	5.58	8.71	16.48	9.47
Malaysia	10.12	44.89	21.04	28.87	14.01
Philippines	2.38	97.87	45.96	50.12	33.95
Singapore	11.96	5.78	8.94	18.02	9.87
Thailand	4.56	62.77	36.48	42.74	52.31
India	9.66	35.67	18.74	32.45	10.03
Pakistan	8.73	31.23	15.78	29.84	9.92
Myanmar	3.93	65.83	38.57	45.04	54.23
Cambodia	3.54	149.32	54.54	76.53	40.97
Japan	15.76	4.97	8.84	19.02	9.69

The results show that the degree of trade facilitation is the highest and the space is not large. Philippines, Vietnam, Kampuchea will give priority to the customs and e-commerce applications can improve the trade volume.

From this we give the suggestion that in the domestic resources to improve the range of 5 points of trade facilitation, if the limited resources, it should be based on the relevant calculation, determine which trade facilitation of the contribution of the sub index, priority to the use of resources in a greater impact on the trade facilitation of points.

Specifically, governments should increase the investment in the infrastructure of trade facilitation, not only the construction of hardware facilities, but also the investment of software facilities. Second, the development of e-commerce, the relevant countries, especially ASEAN countries, in communication networks and big data processing is a serious short board, which greatly restricts the trade volume. Therefore, China and other countries should vigorously develop the Internet in

accordance with their own national conditions, the implementation of government network, paperless office. Finally, we must improve the customs policies and regulations, strengthen the transparency of information and the overall service capacity of the customs.

Trade facilitation is the first proposed and studied by the developed countries in twenty-first Century, the maritime Silk Road strategy should be studied, the use of relevant knowledge, promote the "along the way" of economic and trade development, and enhance the confidence of the relevant countries.

6. MODEL SUMMARY AND IMPROVEMENT

With the improvement of living standards, network security has become a hot issue of social concern, with the help of the network, the activities of terrorist organizations have been greatly expanded.

6.1 The advantage & disadvantage of Trade Gravity Model

Advantages:

1. Trade gravity model on the determinants of the bilateral trade flows has strong explanatory power and achieved great success in many applications. Most of the research shows that, whether from the trade as a whole, or only from the industry level, trade partner economies of scale (GDP), space distance, population and system arrangement are significant effect factors, especially the role of the economy. The role of the institutional arrangements for differences, mainly depends on the total member economies of scale.

2. With the development of the institutional economics, people are paying more attention to the formal system and informal system's impact on international trade flows. Many studies confirm that legal system of trade partners, contract implementation, security, property safety and other factors to a large extent affects the traders expectations about the safety of transaction, so the system is superior to bilateral trade flows has a prominent effect. In addition, the system of quality similar countries easier to build up trust, thus is advantageous to the bilateral trade.

Disadvantages:

Despite the gravity model of trade between the countries, and even trade creation has strong explanatory power, but for some trade phenomenon, gravity model is ragged.

1. Industry trade between the countries is zero, CES preference doesn't tally with the real data, the trade of income elasticity deviation unit elasticity and between the industry is not the same and not reasonable explanation.

2. Gravity model estimation method is easy to underestimate trade trade between countries, overestimate small volume of trade between nations. If included in the gravity model is specialized in primary products production and export of small, can bring interference to model, estimation and prediction results are inaccurate.

The reason is that small country's main export products more onefold, and mostly labor intensive products or natural products, gravity model is often include national income, the distance is no longer the main factors affecting these products trade. Gravity model, therefore, are often unable to trade goods single effective analysis of trade between nations.

3. When using the gravity model analysis to export problems also exist obvious malfunction.

6.2 The advantage & disadvantage of Data Classification Model

Advantages:

1. Regression analysis based on the analysis of multi-factor model, a more simple and convenient;

2. Using the regression model, as long as the same model and data, through the standard statistical method can calculate the only result;

3. Regression analysis can accurately measure the factors fitting degree of regression and correlation degree between high and low, improve the effect of prediction equations; Due to actual a variable in the regression analysis method is rarely influenced by individual factors, only pay attention to the mode of suitable range, so the multivariate regression analysis method is suitable for practical economic problems, comprehensive influence by various factors. Disadvantages:

4. In this model, it can be used to classify the data in the algorithm. So it is not possible to achieve the actual application. There will always be some errors.

Disadvantages:

Sometimes in regression analysis, choose what factor and the factor what expression is a kind of speculation, which affects the utilization factor of diversity and unpredictable, some factors regression analysis in some cases was limited.

7. REFERENCES

[1] Liu Jia. "The new Silk Road Economic Belt" strategy considerations [D]. Beijing Foreign Studies University, 2015.

[2] Hou Fulai. "The Belt and Road" under the national strategy China EU economic and trade relationship between [D]. China Foreign Affairs University, 2015.

[3] Fan Xiufeng, Cheng Wenxian. Study on trade facilitation of the new Silk Road on the sea [J]. International Business (Journal of University of International Business and Economics), 2015,05:144-152.

[4] Liang Feiran, new China's new child. Silk Road Economic Belt total factor productivity from the perspective of financial measure of [J]. Industrial Technology Economy, 2015,10:59-68.

[5] Ma Bo. "Along the way" and the "global maritime fulcrum" of strategic docking research [J]. International Outlook, 2015,06:33-50+152-153.

[6] Zhang Yong. The strategic significance of the construction of the Silk Road on the sea in twenty-first Century [J]. new economy, 2014,31:47-53.

[7] Li Xiangyang. The diversity and cooperation mechanism of the Silk Road on the sea [J]. world economy and politics, 2014,11:4-17+155.

[8] Wang Aihu. The value and significance of the development of the Silk Road from the sea to the value and significance of the new maritime Silk Road [J]. Journal of South China University of Technology (SOCIAL SCIENCE EDITION), 2015,01:1-14.

[9] Zhou Maorong, Tan Xiujie. The trade potential and its influencing factors of the Silk Road on the sea in twenty-first Century, [J]. International trade issues, 2015,02:3-12.

[10] Zhao Yiqi, Wang Zhaoxin. The maritime Silk Road Economic Zone Tourism Service Industry in Hainan Province, an empirical study on the influence of regional economic [J]. Journal of Commercial College of Guizhou, 2015,02:36-40.

[11] Chen Wanling, He Chuantian, the parties to the maritime Silk Road and its economic and trade position [J]. Reform, 2014,03:74-83.

[12] Zhang Yong. A brief discussion on the national development strategy of the Silk Road on the sea in twenty-first Century [J]. Journal of Ocean University of China (SOCIAL SCIENCE EDITION), 2014,05:13-18.

[13] Quan Yi, Wang Jie, Liu Wanting. The strategic vision and construction strategy of the Silk Road on the sea in twenty-first Century [J]. International trade, 2014,08:4-15.

[14] Shan Na. Corporate environmental strategy decision [D]. study and performance evaluation of Ocean University of China, 2012.

[15] Peng Rongsheng. The connotation, mechanism and evaluation of regional economic coordinated development [D]. Henan University, 2007.

[16] Yao Li. Economic growth effect of regional economic integration [D]. Northeast Normal University, 2015.

[17] Feng Weijiang. International political and economic analysis of the Silk Road Economic Zone [J]. 2014,06:73-98+157-158.

[18] Zhang Yongfeng, Du Zhongchao. Analysis of the competitiveness of the 10 cities along the Silk Road in Northwest China [J]. Arid Zone Resources and Environment, 2009,10:194-200.

[19] Chen Xiuzhen - CDI. Quantitative economic integration between Hong Kong and the mainland open evaluation index of [J]. Review of the Degree of Economic Integration between Hong Kong and Mainland China, 2005,04:62-68.

[20] Bi Wei. Theory and practice of circular economy and the comprehensive evaluation system of [D]. Tianjin University, 2006.

Statistical Review of Various Object Detection Techniques

Bindu N.S
PES College of Engineering, Mysore
Karnataka, India

H.S.Sheshadri
PES College of Engineering, Mysore
Karnataka, India

Abstract: The main and important objective of tracking an object is to segment a ROI from a given video scene. Once the object is tracked it is continued to track its motion occlusion and most importantly its position. Detection and classifications of objects are the previous steps to track an object in a given video. Object detection process is done to check the existence or presence of objects in a given video and also to exactly locate the existed object. After detection, the objects that are detected are classified into varieties of categories namely humans, pedestrian, vehicles and also other objects which are under motion. Object tracking process is done by monitoring the objects temporal as well as spatial changes throughout a given video sequence which includes the presence, size and position etc. Object tracking has been used in numerous applications like video surveillance, artificial intelligence, traffic monitoring and also in video animations. This work presents a short-term survey of various object detection algorithms which is available in literature survey together with analysis and also a comparative learning of diverse detection techniques used.

Keywords: Object detection, Background subtraction, frame differencing, optical flow, voting based detection

1. INTRODUCTION

In reality videos are a sequences of frames which are displayed in a fast frequency mode so that a human eyes be able to percept the continuousness of the content. It is more obvious that each and every techniques that make use of image processing be able to apply to an individual frames. Further, the contents or the information of the two consecutive frames will be generally closely related [1].The appropriate identification of these ROI are typically going to be the first phase in many of the applications which is based on computer vision and these applications comprises of event detection, motion detection and video surveillance. A common detection algorithm might be appropriate, but then it is tremendously difficult to appropriately handle the unknown objects and also objects with substantial variations in terms of color, texture and shape. Therefore, a lot of practical systems based on computer vision assumes of fixed camera environs, so that it makes an object detection procedure a straightforward [2]. An image is usually divided into a two complimentary groups of pixels, where first group consists of pixels which are correspond to a foreground objects and the second group consists of background pixels. This results are represented as mask or a binary image. It is pretty difficult to identify a complete standard with reverence to what must be recognized as foreground, also what must be marked as a background since this description is somewhat an application specific. Usually, foreground objects will be a moving objects alike people, cars, pedestrians and boats and also everything other is background [3]. Most of the times, shadow will be classified as a foreground object giving an improper and bad output.

2. DETECTION METHODS

Since the moving objects remain as a prime source of information, maximum methods focuses on those objects which has to be detected. Detailed explanation of various approaches are listed and explained below.

A. Frame differencing

The existence of objects which are under motion are generally determined by computing the dissimilarities between the two consecutive frames. The calculation is very easy and also it's simple to implement too. For a diversity of dynamic environs, it has a very strong adaptability, however it is usually difficult to get complete summary of a moving object, which are in turn responsible to seem the unfilled phenomenon, resulting in an inaccurate outputs of objects which are moving [4]. The below example in terms of Fig.1 illustrates how the frame differencing works


(a) Sample scene
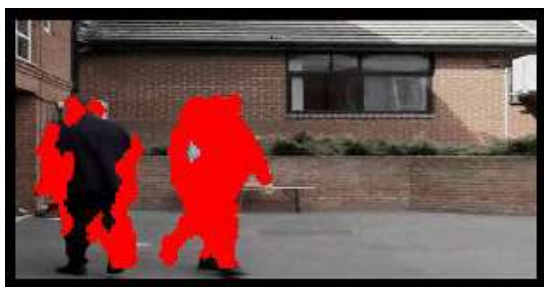
(b) Failure of Frame differencing
Fig.1. Example of Image differencing

B. Optical Flow

Optical flow technique [5] is used to compute the image field i.e., optical flow playing field, and also it does the clustering processing accordingly with the optical flow scattering characteristics of an image. This method might get complete information about the movement and also it can detect the objects which are moving from the so called background. Conversely, an enormous amount of calculation, sensitivity towards noise, and also poor noise performance, makes it as not appropriate for the real-time challenging occasions. Fig.2. illustrate the examples of how the optical flow approach results.
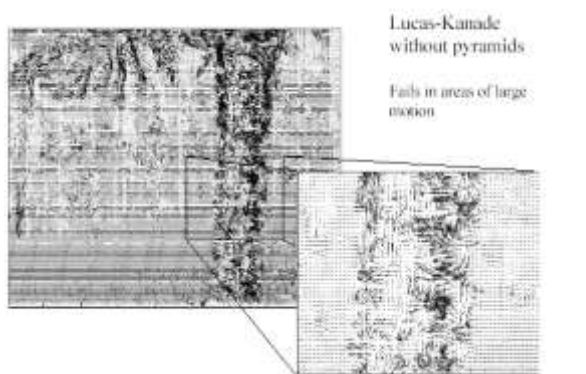
(a) Lucas-kanade-without
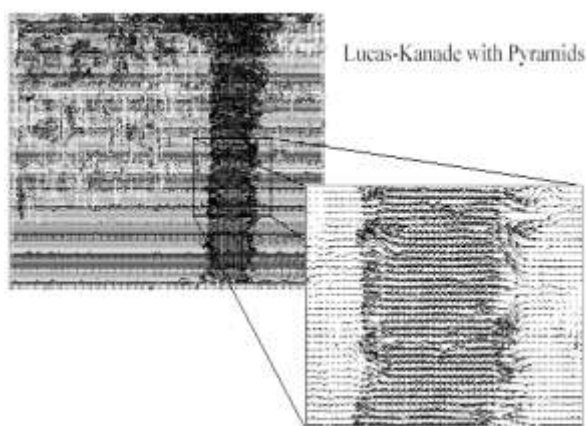
(a) Lucas-kanade-with
Fig.2. Example of how optical flow results

C. Background subtraction

Background modelling is the first step used for so called background subtraction. It has been the central importance for a background subtraction techniques. Background Modeling should be sensitive enough in order to recognize the moving objects [6]. Background Modeling yields the reference model. Later this model will be used in subtracting the background in which individual video sequence will be compared in contrast to the so called reference model which in turn aims at determining the possible Variations. These variations amongst the current frame and the reference frame with respect to pixels signifies the existence of an objects that are moving [6]. At present, mean and median filters [7] are more extensively used in order to realize the background modeling. However, this provides the maximum complete information about the object if background is well-known. Fig.3. gives a general background subtraction approach and Fig.4. Gives the results.
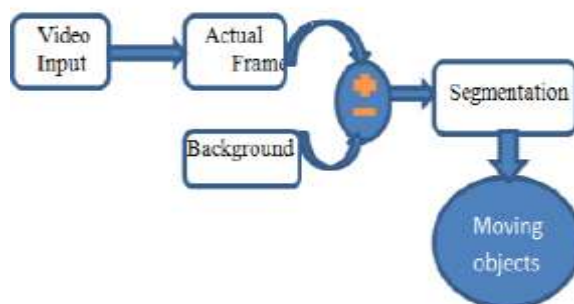

Fig.3. General Background subtraction approach

Fig.4. Background Subtraction

In [3], as described there are mainly two background subtraction approaches namely

1. Recursive algorithm

Recursive techniques [3, 8] does not keep up a buffer in estimating a background, Instead it recursively keeps updating a single model for each of the input frame. This results in input frames having an effect towards the present background model. On comparing to non-recursive-techniques, the recursive techniques will require a less storage. This technique also includes numerous methods like approximate median, GOM and adaptive background.

2. *Non-Recursive Algorithm*

Non-recursive-techniques in [8, 3] makes use of sliding-window method to estimate the background. It also stores the buffer of a preceding video frames which in turn evaluates the so called background image which is based on temporal variations in each of the pixel that are buffered. Techniques which are based on non-recursive are found to be highly adaptive since they it does not depend completely on the past beyond those individual frames that are stored in that buffer. Conversely, the storage necessity or requirement will be consistently significant if in case a huge buffer is desired to cope up with the slow-moving-traffic.

D. *Motion Estimation based on voting*

In order to estimate a proper motion of the moving objects, this voting based algorithm is developed and used. A powerful and also speedy voting based algorithm is developed for an indoor surveillance to evaluate the motion of the camera without tracking the features and also knowing clear knowledge of the camera motion. Due to the mobile camera, estimation of motion exploits the set of voting decision as of the group of these motion vectors which are found by edge feature or background in order to precisely estimate and also compensate in place of shifting motion. The VBME technique will decrease the time complexity and also avoid the incorrect tracking of feature due to edge characteristics and only it is employed [9]. Temporal sampling based on content and also spatial coding based on priority are the two algorithms which involve the voting based method.

1. Temporal Sampling Based on content

The calculation of movement vector is done by considering the two static-images. According to this sampling algorithm, information which is more important will be preserved prudently and also the redundant frames which are similar will be removed effectively [10].

2. Spatial Coding Based on priority

Spatial domain contains two parts i.e. important and also unimportant region. Algorithm based on spatial coding is used to encrypt those regions having important information in frame with greater visual quality. The regions without any important information will have a very low visual quality [10]. It also includes two key parts namely

(a) Set the priority to a region-based resulting in an edge detection of moving objects. The pixels in these contour maps is categorized into 3 distinct types namely moving edge, no edge and finally a static edge and also their respective priority will be high, low and medium.

(b) Generating image code which is based on the priority- In this SPIHT algorithm is adopted and used. This

algorithms initially partitions each and every blocks by same priority. On a whole, the entire encoding process would upgrade the so called quality on those regions that are occupied at the moving objects edges but it degrades the other regions quality. Therefore, when the accessible bandwidth is under limit, the quality of those high-priority info can be maintained [10].

TABLE 1

(a) COMPARATIVE STUDY BASED ON ACCURACY AND COMPUTATION TIME

Methods		Accuracy	Computation Time
Background Subtraction	GOM	Moderate	Moderate
	Approximate Median	Low-to-moderate	Moderate
Optical Flow		Moderate	High
Frame Differencing		High	Low-to-Moderate
Voting Based		High	High-to-Moderate

TABLE 2

(B) COMPARATIVE STUDY AND THE COMMENTS ON DIFFERENT DETECTION TECHNIQUES

Methods		Comments
Background Subtraction	GOM	✓ On the positive side, it requires low memory ✓ On the negative side, it cannot cope up with the multimodal background
	Approximate Median	✓ On the positive side, it do not require any frame sub sampling ✓ On the negative side, its computation time requires a small buffer with an recent values i.e. pixel values
Optical Flow		✓ On the positive side, it gives the complete information about the movement ✓ On the negative side, it requires a large volume of calculations
Frame Differencing		✓ On the positive side, it is an easier method ✓ On the positive side, it requires background devoid of moving objects
Voting Based		✓ On the positive side, it performs well in static background

CONCLUSION

This paper gives a brief survey of various methods that has been used to detect the objects. Each available techniques are explained with a relevant examples for each of the techniques. The various methods are discussed in the above sections which includes frame differencing, background subtraction, optical flow and voting based motion detection. It is summarized that the background subtraction technique found to be a simple and easiest method which provides complete info of an objects compared to other methods.

ACKNOWLEDGMENT

REFERENCES

[1]. Abhishek Kumar Chauhan, Prashant Krishan, "Moving Object Tracking Using Gaussian Mixture Model And Optical Flow", International Journal of Advanced Research in Computer Science and Software Engineering, April 2013.

[2]. Cheng-Laing Lai; Kai-Wei Lin, "Automatic path modeling by image processing techniques," Machine Learning and Cybernetics (ICMLC), 2010 International Conference on, vol.5, no., pp.2589, 2594, 11-14 July 2010.

[3]. Sen-Ching S. Cheung and Chandrika Kamath, "Robust techniques for background subtraction in urban traffic video".

[4]. Rupali S.Rakibe, Bharati D.Patil, "Background Subtraction Algorithm Based Human Motion Detection", International Journal of Scientific and Research Publications, May 2013

[5]. Abhishek Kumar Chauhan, Prashant Krishan, "Moving Object Tracking Using Gaussian Mixture Model And Optical Flow", International Journal of Advanced Research in Computer Science and Software Engineering, April 2013

[6]. J.Joshan Athanesious, P.Suresh, "Systematic Survey on Object Tracking Methods in Video", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) October 2012, 242-247.

[7]. M.Sankari, C. Meena, "Estimation of Dynamic Background and Object Detection in Noisy Visual Surveillance", International Journal of Advanced Computer Science and Applications, 2011, 77-83

[8]. K.Srinivasan, K.Porkumaran, G.Sainarayanan, "Improved Background Subtraction Techniques For Security In Video Applications"

[9]. Sarita P. Shinde, Prof. A. R. Askhedkar, "Motion Estimation using Mobile Camera From Video", International Journal of Latest Trends in Engineering and Technology (IJLTET Vol. 3 Issue 3 January 2014.

[10]. Feng-Li Lian, Yi-Chun Lin, Chien-Ting Kuo, and Jong-Hann Jean" Voting-Based Motion Estimation for Real-Time Video Transmission in Networked Mobile Camera Systems" *IEEE transactions on industrial informatics*, vol. 9, no. February 2013

Appraisal of Queuing Performance in Chester Mega Petroleum Station Enugu

Onyechi, Pius Chukwukelue.[1], Ogunoh, Victor Arinze [2] and Ezeliora, Daniel Chukwuemeka[3]

Department of Industrial and Production Engineering, Faculty of Engineering NnamdiAzikiwe University Awka

Abstract.

The operation manager of Mega Petroleum Station was finding it difficult in managing their queuing system and was not able to determine the best number of servers that can serve arriving customers at various demand periods which affects their queue performance. This study was conducted at Chester Mega Petroleum Station in Enugu, Nigeria with the aim of addressing the identified problem. The result from the Queue Evaluation Environment showed that 8 servers gave the best system utilization values of 73.2% which is expected to reduce the respective customers waiting times (Ws) by 84.7% for the case study establishments. The result further showed that with 8 servers for the service systems, there will be no more need of server increase at the respective average arrival rates of customers/minutes and also the expected probabilities of system idleness for the case studies were negligible at 8 server utilization. The Queue Evaluation Environment was later adopted in developing a Decision Support System for the referenced service facilities. The model was recommended for PMS refill only.

Key Words: Queuing, waiting line, arrival rate, service rate, waiting time, system utilization, Chester mega station, probability and decision support system

1. Introduction

A queue is a waiting line for service that forms either due to inability to meet up with arriving demands as a result of insufficient service capacity or due to stochastic nature of customer arrival and demand. A queue could deteriorate to congestion if effective queue management decisions are

not implemented. Queuing theory is the mathematical study of waiting lines [1]. The theory permits the derivation and calculation of several performance measures which includes the average waiting time in the queue or the system, the expected number waiting or receiving service, the probability of encountering the system empty, having an available server or having to wait a certain time to be served and most importantly the system utilization [2]. As a result of its applications in industries, technology, telecommunications networks, information technology and management sciences, it has been an interesting research area for many researchers active in the field.

In recent times, queuing theory and the diverse areas of its applications has grown tremendously. Takagi (1991) considered queuing phenomena with regard to its applications and performance evaluation in computer and communication systems [3]. Obamiro applied Queuing Model in Determining the Optimum number of Service Facility needed in Nigerian Hospitals. He however achieved this by determining some queuing parameters which enabled him to improve the performance of the system [4].

Chinwuko and Nwosu adopted the single line multi-server queuing existing model to analyze the queuing system of First Bank Nigeria PLC. In their work, they suggested the need to increase the number of servers in order to serve customers better in the case study organization [5].

Ohaneme et al, proposed the single line multi-server queuing system which they simulated using c-programming to be adopted at NNPC Mega petroleum station in Awka, Anambra State in order to avoid congestion and delay of customers [1].

Presently, studies on performance evaluation of queuing have not been conducted at Chester Mega Petroleum Station Enugu. However, this work goes further in evaluating the performance of the queuing systems, creating a Queue Evaluation Environment that gives expected queue performance and developing a Decision Support System that recommends the best number of servers to use at various demand periods.Basic structure of queuing model can be separated into

input and output queuing system, which include queue that must obey a queuing rule and service mechanics [6].

1.1. Input and Output Process: Input process is known as the arrival process. Customers are known as arrivals which are generated one time by an input source randomly from finite or infinite population. These customers enter the queuing system and join a queue to be served. The required service is then performed for the customer by the service mechanism, after which the customer leaves the queuing system [6]. The provision of services using certain rule and discharge of customers is referred to as output process. Another fact worth mentioning here is that the key word in queuing models is "average". It takes the average of the random numbers of customers arriving [7].

1.2. Queuing System Characteristics: The queuing system consists primarily of the waiting line(s) and the available number of servers. Factors to consider with waiting lines include the line length, number of lines, and queue discipline.Queuing phenomenon comprises of the following basic characteristics: Arrival characteristics; Queue or the physical line itself; Number of servers or service channels; Queue discipline; Service mechanism; Capacity of the system; and Departure [8].

1.3. Waiting Line or Queue: A waiting line or queue occurs when customers wait before being served because the service facility is temporarily engaged. A queue is characterized by the maximum permissible number of customers that it can contain. Queues are called infinite or finite. An infinite queue is one in which for all practical purposes, an unlimited number of customers can be held there while a finite queue refers to the limited-size customer pool that will use a service system and, at times, form a line [9].
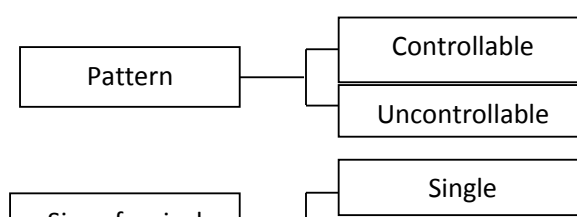
Figure 1: Arrival Characteristics in Queue
Source: Adopted from Davis et al., [10]

1.4. Types of Queuing Systems: There are four major types of queuing system. Lapin broadly categorized queuing system structures into the following [11].

- **Single-server, Single-phase system**:

    This is a situation in which single queue of customers are to be served by a single service facility (server) one after the other. An example is bottles or cans of minerals or beer to be cocked in a production process.

- **Single-server, Multiple-phases System**:

    In this situation, there's still a single queue but customers receive more than one kind of service before departing the queuing system. For example, in the university, students first arrive at the registration desk, get the registration done and then wait in a queue for their forms to be signed, after signing; they join another queue for submission.

- **Multiple-servers, Single-phase System**:

  This is a queuing system characterized by a situation whereby there is a more than one service facility (servers) providing identical service but drawn on a single waiting line. An example is a petroleum service station.

- **Multiple servers, Multiple-phases System:**

  This type of system has numerous queues and a complex network of multiple phases of services involved. This type of service is typically seen in a hospital setting, multi-specialty outpatient clinics, patient first form the queue for registration, and then he/she is triage for assessment, then for diagnostics, review, treatment, intervention or prescription and finally exits from the system or triage to different provider [7].

1.5. Performance Measures of a Queuing System: Hillier and Lieberman put forth the following performance parameters in a queuing system [6]:

- System Utilization (P): System Utilization is the most important measure of a queuing system. It is the ratio of system capacity used to available capacity. It measures the average time the system is busy. System utilization of zero means that there is nobody in the system. On the other hand, a system utilization of one or more signifies that there is infinite number of people on the waiting line. This means that the available servers cannot cope with the arriving demand. Thus something has to be done on the service facility. Egolum, further stated that the best value of system utilization should be greater than 0 but less than 0.8 i.e. 80% [12].

- Mean Number in the system (L_s): Mean number in the system is the average number of system users (entities) in the system; it includes those in the queue and those being served by the server(s).

- Mean Number in Queue (L_q): Mean number in the queue is the average or expected number of system users in the queue (waiting line), waiting for their turn to be served.

- The average waiting time for an arrival not immediately served (W_a)

- Mean Time in System (W_s): Mean time in the system is the expected value or average waiting time an entity will spend in the queuing system. It includes the average time waiting for service to begin and the average service time.

- Mean Time in Queue (W_q): Mean time in the queue is the expected value or average time an entity will spend in the queue, waiting for service to begin.
- probability of zero customers in the system (P_0)
- Probability of waiting (P_w): This is the probability that an arrival will have to wait for its service to begin.

The aim of this study is to address the queuing problem at Chester Mega Petroleum Station by developing a Decision Support System that recommends the best number of servers needed to be engaged at various demand periods.

The structure of the studied system is shown in figure 2. The structures can be approximated as a single-line multi-server queuing systems. . At the Chester Mega Petroleum Station, there are six dispensers i.e. fuel metering pumps (S_1 to S_6) in the system. Each of the fuel dispensers has two nozzles. This means that at full capacity of operation the service facility should be considered a twelve-server system.

2. Methodology.

There are two major techniques of research methods; they are qualitative research method and quantitative research method. The research method used in this work was the quantitative research approach. The single line multi-server queuing model was adopted for developing the results of the queue performance. This model was adopted because it showed a good representation of the model structure of both case studies of queuing systems.

2.1. Data Analysis

The data generated was first organized and descriptive statistics was used to compute the total average arrival rates and total average combined service rates for the year. The service rates per server of both facilities were established and the single line multi server queuing model was coded in Microsoft Excel using 2 – 12 servers (i.e. when M = 2 – 12 servers) in creating the Queue Evaluation Environment that generates the expected queue performance results at the respective average arrival rates of customers in the referenced service facilities. The results generated were validated using MATLAB (version 7.10.0.499: R2010a). The essence was to see if both results

corresponded with each other. The Queue Evaluation Environment was later adopted in developing the decision support system using the application of Microsoft Excel.

2.2. Models Applied for the Queuing Analysis

Based on the assumptions of the single line multi-server queuing model, the expressions for the performance measures which are derived from the analysis of the birth-and-death models, (Blanc, 2011 [13]; Sztrik, 2011 [14]; and Nain, 2004 [15]) are;

    i.   The average utilization of the system:

        When m = 6 is

$$P = \frac{\bar{\lambda}}{\bar{\mu}_c} \tag{1}$$

        When m = 2 − 12 is

$$P = \frac{\bar{\lambda}}{M(\bar{\mu})} \tag{2}$$

    ii.   The probability that there are no customers in the system is

$$P_0 = \left[\sum_{n=0}^{M-1} \frac{\left(\frac{\bar{\lambda}}{\mu}\right)^n}{n!} + \frac{\left(\frac{\bar{\lambda}}{\mu}\right)^M}{M!\left(1 - \frac{\bar{\lambda}}{M\mu}\right)} \right]^{-1} \tag{3}$$

    iii.   The average number of customers waiting for service.

$$L_q = \frac{\bar{\lambda}\mu\left(\frac{\bar{\lambda}}{\mu}\right)^M}{(M-1)!\left(M\mu-\bar{\lambda}\right)^2} P_0 \tag{4}$$

    iv.   The average number of customers in the system.

$$L_s = L_q + (\bar{\lambda}/\bar{\mu}) \tag{5}$$

    v.   The average time a customer spends in line waiting for service

$$W_q = \frac{L_Q}{\bar{\lambda}} \tag{6}$$

    vi.   The average time a customer spends in the system.

$$W_s = \frac{L_s}{\bar{\lambda}} \tag{7}$$

    vii.   The average waiting time of a customer on arrival not immediately served.

$$W_a = \frac{1}{M\mu-\bar{\lambda}} \tag{8}$$

    viii.   Probability that an arriving customer must wait

$$P_w = \frac{W_Q}{W_a} \tag{9}$$

It is seen that these performance measures depend on two basic queue parameters, namely; $\bar{\lambda}$ and $\bar{\mu}$. Given $\bar{\lambda}$ and $\bar{\mu}$, the values computed for these measures gives an indication of how well the referenced service facilities handle the volume of arriving customers.
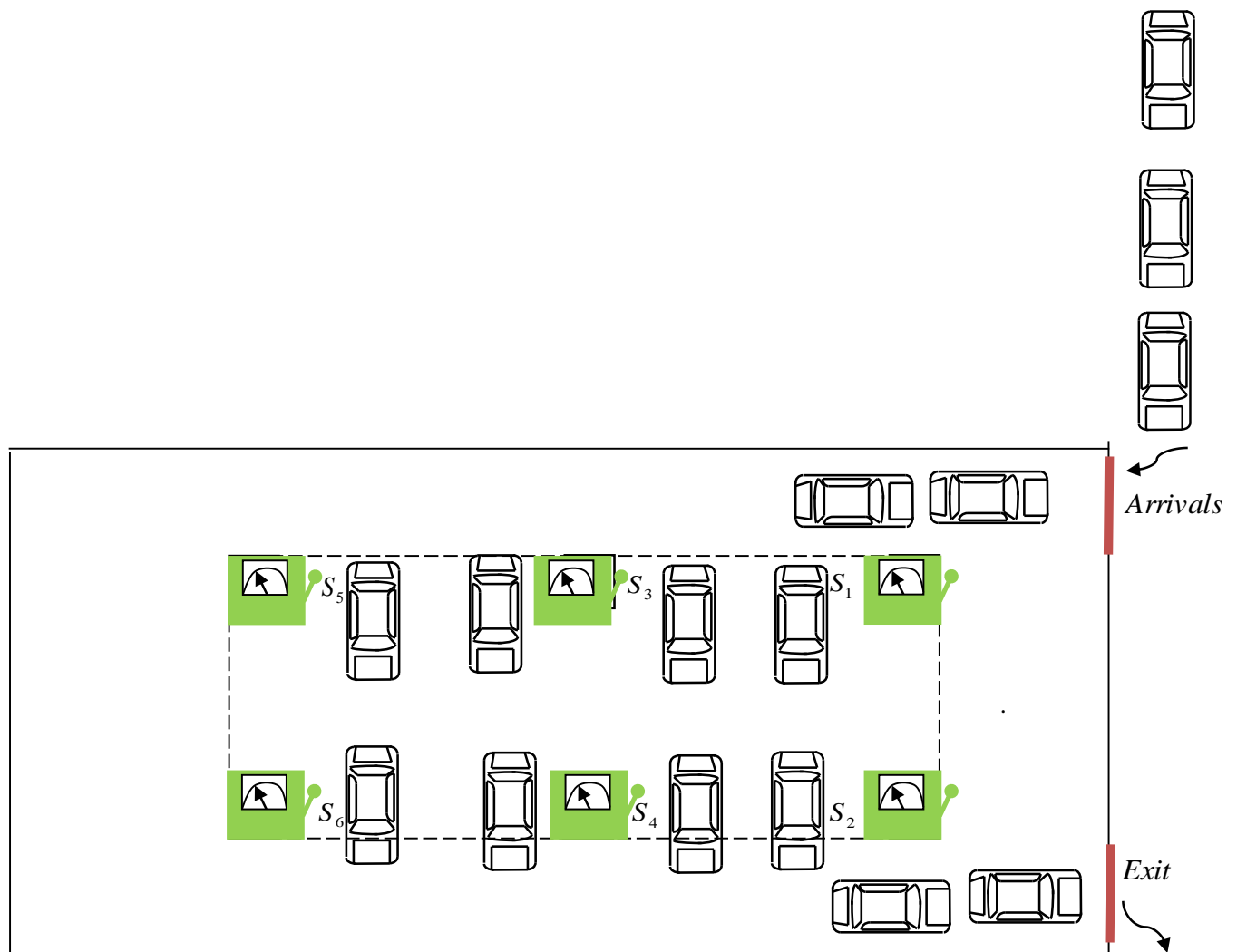

Figure 2: Structure of the PMS Dispensary pump system of the studied Chester mega petroleum station Enugu.

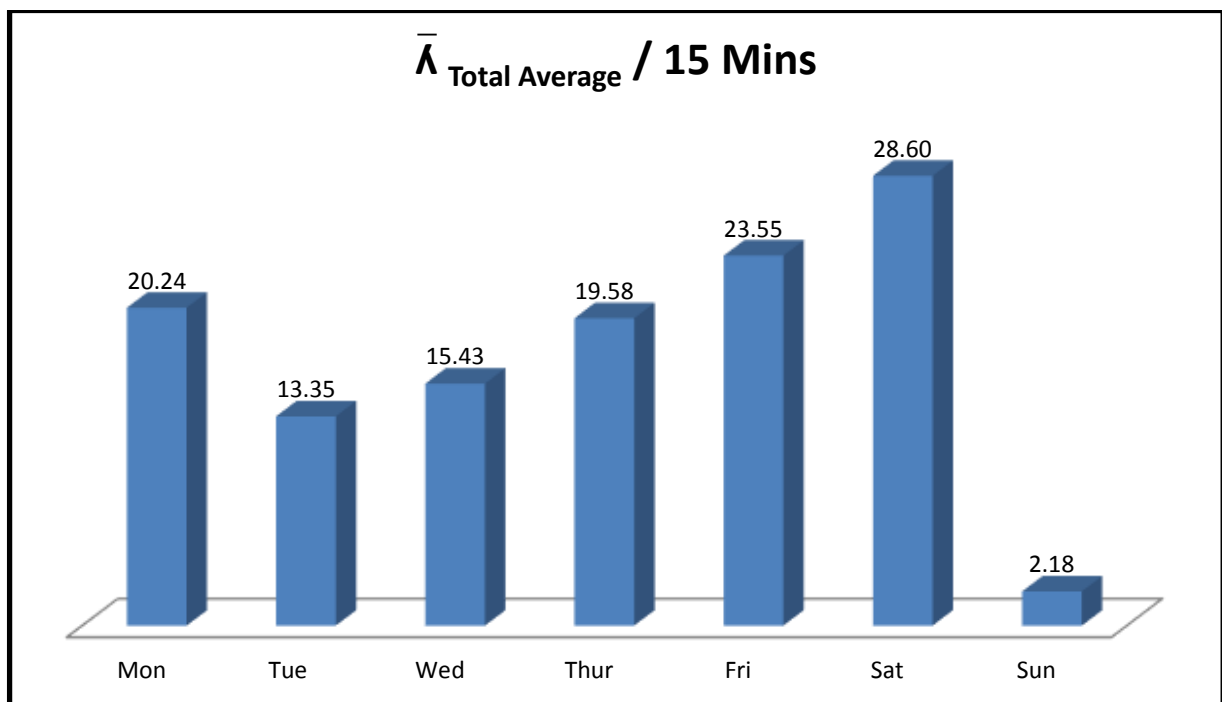## 3. Data Analysis Result of Chester Mega Petroleum Station in Enugu

Figure 3: Total Daily Average Arrival Rate / 15 Minutes

In figure 3, the bar chart shows the total daily average arrival rate of customers per 15 minutes for the year from Monday to Sunday at Chester Mega Petroleum station Enugu. From the chart, it is observed that Saturday was with the highest arrivals which shows that the mega station is being patronage more by customers on Saturdays than the rest of the days being the fact that Saturday is a work free day for civil servants and most public servants so customers buy large quantity of PMS to last them for the week day activities as well as weekend travels. While Sunday was with the lowest arrivals which shows less patronage of customers being the fact that Sunday is a worship day for Christians and the Mega Station don't always open for service on that day and also customers most especially civil and public servants must have bought large quantity of PMS on Saturday and Friday to last them for the week. The week days (i.e. Mondays to Fridays) were mostly patronage more by commercial transporters

Table 1: Weekly Mean Arrival Rate of Customers, Weekly Mean Combined Service Rate of Customers and Mean Number of Servers Engaged at Chester Mega Petroleum Station Enugu for the year

		Weekly Average Arrival Rate Per Mins	Weekly Average Combined Service Rate Per Mins	Weekly Average Number of Servers Being Used (M)
DEC	1st Week	0.878	0.9095	7
	Last Week	0.3625	0.4071	3
JAN	1st Week	0.519	0.5911	5
	Last Week	1.2863	1.3042	5
FEB	1st Week	1.2744	1.297	8
	Last Week	1.2542	1.2887	6
MAR	1st Week	1.1798	1.206	6
	Last Week	1.225	1.2488	6
APR	1st Week	0.9137	0.9333	7
	Last Week	1.1982	1.2214	7
MAY	1st Week	1.2369	1.2577	8
	Last Week	1.256	1.2821	7
JUN	1st Week	1.2238	1.2554	6
	Last Week	1.2488	1.2762	6
JUL	1st Week	1.2601	1.3	8
	Last Week	1.1	1.1226	7
AUG	1st Week	1.3458	1.3756	6
	Last Week	1.3137	1.3345	6
SEP	1st Week	1.3506	1.3792	7
	Last Week	1.2792	1.3089	7
OCT	1st Week	1.3137	1.3411	6
	Last Week	1.3054	1.3286	6
NOV	1st Week	1.3976	1.4155	7
	Last Week	1.3762	1.394	7
	Total	28.0989	28.7785	154
	Average	1.1708	1.1991	6

Total average arrival rate for the year $= \frac{\sum \bar{\lambda}_{(Weekly)}}{24} = 1.1708$ cars/minutes

Total average number of servers being used for the year $= \frac{\sum \bar{M}_{(Weekly)}}{24} = 6$

Total average combined service rate for the year $= \frac{\sum \bar{\mu}_{c(Weekly)}}{24} = 1.1991$ cars/minutes.

It is assumed that each server contributes an average service rate of $\frac{\bar{\mu}_{c(Year)}}{\bar{M}}$ cars/minutes. Where $\bar{M} = 6$, and $\bar{\mu}_{c(Year)} = 1.1991$ cars/minutes. This implies that each server contributes an average service rate of 0.1999 cars/minutes in the service facility

4. Presentation of Queue Performance Evaluation Results

The results of the performance measures of the queuing system are presented below.

Table 2: Results of the performance evaluation of the queuing system with parameters $\bar{\lambda} = 1.1708$ cars/minutes and $\bar{\mu}_c = 1.1991$ cars/minutes when ($M = 6$

Average Arival Rate ⋏	1.1708		Average Combined Service Rate μc	1.1991
System Utilization	p	0.9762		
Probability system is empty	P0	4E-04		
Probability Arrival must wait	Pw	0.9348		
Average no in line	Lq	38.3		
Average no in System	Ls	44.125		
Average Time in Line	Wq	32.686		
Average Time in System	Ws	37.688		
Average Waiting Time	Wa	34.965		

Formula bar (I7): =(((((C1/G1)^0)/1)+(((C1/G1)^1)/1)+(((C1/G1)^2)/2)+(((C1/G1)^3)/6)+(((C1/G1)^4)/24)+(((C1/G1)^5)/120)+(((C1/G1)^6)/(720*(1-(C1/(6*G1)))))))^-1

Single Line Multi Server Queuing Simulation 1 [Compatibility Mode]

			2	3	4	5	6	7	8	9	10	11	12
Average Arrival Rate λ	1.1708		Service Rate μ	0.1999									
			(Per Server)										
Number of Servers (Max 12)		M	2	3	4	5	6	7	8	9	10	11	12
System Utilization		p	2.9285	1.9523	1.464	1.1714	0.9762	0.8367	0.7321	0.6508	0.5857	0.5324	0.4881
Probability system is empty		P0	-0.491	-0.09	-0.021	-0.004	0.0004	0.00197	0.00253	0.00274	0.00282	0.0028	0.0029
Probability Arrival must wait		Pw	4.366	3.153	2.1948	1.466	0.9348	0.5663	0.3245	0.1752	0.0889	0.0424	0.019
Average no in line		Lq	-6.63	-6.463	-6.922	-10.02	38.3	2.9015	0.8867	0.3265	0.1257	0.0483	0.0181
Average no in System		Ls	-0.773	-0.606	-1.066	-4.163	44.125	8.7584	6.7436	6.1834	5.9827	5.9052	5.875
Average Time in Line		Wq	-5.663	-5.52	-5.913	-8.558	32.686	2.4782	0.7574	0.2789	0.1074	0.0412	0.0154
Average Time in System		Ws	-0.66	-0.518	-0.91	-3.556	37.688	7.4807	5.7599	5.2814	5.1099	5.0437	5.0179
Average Waiting Time		Wa	-1.297	-1.751	-2.694	-5.838	34.965	4.3764	2.3343	1.5916	1.2074	0.9727	0.8143

Best Server Utilization

Figure 4: Queue Evaluation Environment displaying the results of the queue performance when $\bar{\lambda}$ is fixed, $\bar{\mu}$= 0.1999 cars/minutes (per server) and $M = 2 – 12$ servers

From figure 4, the charts of the queue output results were developed using the application of Microsoft Excel and trend line was used to test for the best goodness fit in developing the relationship that exists best between the queue output results.

5. Development of the Decision Support System for the Case Study

From the Queue Evaluation Environment created, the service rates (per server) of each of the referenced facilities were fixed and arrival rates were simulated using $2 - 12$ servers to see the expected queue performance and to determine the best number of servers that gives the best system utilization value at various arrival rates of customers. The summary result outputs were plotted on a chart using the application of Microsoft Excel and trend line was used to test for the best goodness fit between the dependent variable i.e. Number of Servers (M) and the independent variable i.e. Average Arrival Rates/Minutes (ʎ). See summary result output and charts below.

Table 3: Summary result output of simulated arrival rates of customers/minutes (Chester Mega Petroleum Station Enugu)

$\bar{\lambda}/Mins$	Best Server Utilization
0.3	2
0.4	3
0.5	4
0.6	4
0.7	5
0.8	5
0.9	6
1.0	7
1.1	7
1.2	8
1.3	9
1.4	9
1.5	10
1.6	10

M vs ʎ/Mins

$y = 8.1369x^5 - 47.477x^4 + 101.99x^3 - 100.12x^2 + 50.647x - 6.5694$

$R^2 = 0.9911$
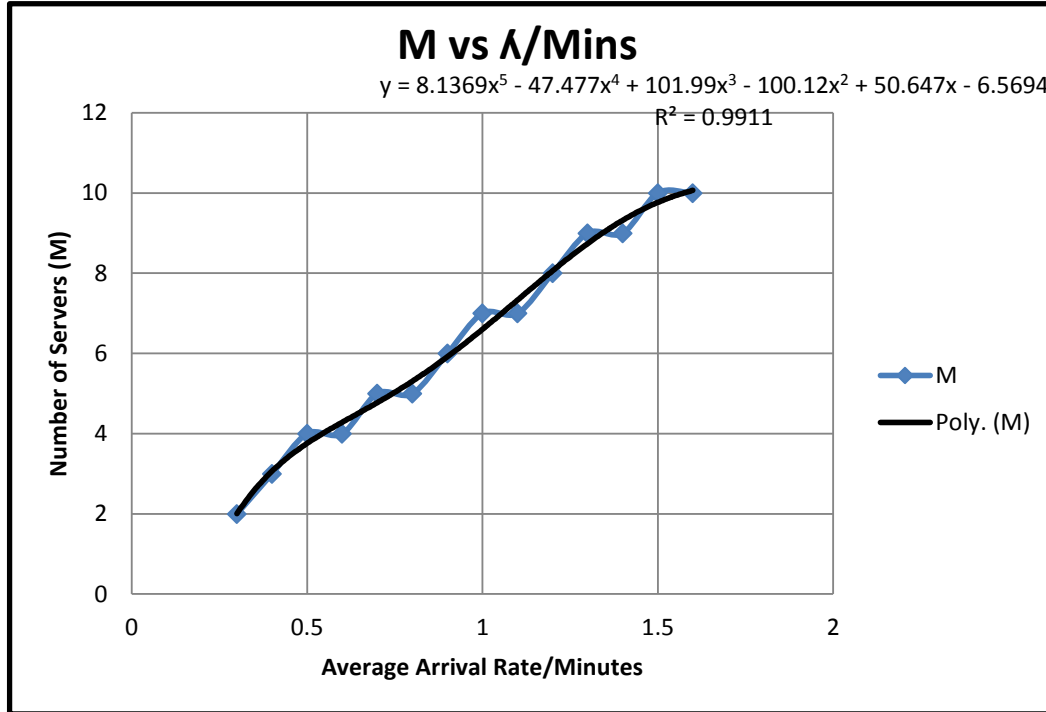
Figure 5: Scatter Plot of Number of Servers (M) vs. Average Arrival Rate/Minutes

From figure 5, the scatter plot shows the number of servers plotted against average arrival rate/minutes in Chester Mega Petroleum Station Enugu. From the chart, it is observed that the number of servers is also expected to be stepping up as average arrival rate increases and the best fit between the two variables i.e. Number of Servers (M) and Average Arrival Rate (ʎ) is a nonlinear polynomial function in fifth order as depicted in the chart.

6. Discussion of Results

From the analysis, table 2, shows the results for the performance measures of the queuing system as seen at the Chester mega petroleum station Enugu. From the results, it was also discovered that with an average number of 6 servers with average combined service rate $(\bar{\mu}_c)$ of 1.1991 cars/minutes and average customer arrival rate $(\overline{\lambda})$ of 1.1708 gave a system utilization (P) of 0.9762 which gives a percentage system utilization of 97.62%, while the probability of the system being empty and the probability of waiting gave 0.0004 and 0.9348 respectively, this means that when service commences, the system is never idle and a customer must wait before receiving service with a 93.48% probability. However, the average number of customers in line and the

average number of customers in system including any being served gave 38.3 and 44.125 respectively. Furthermore, the average waiting time of customers in line, the average waiting time of customers in the system including service and the average waiting time of a customer on arrival not immediately served gave 32.686, 37.688 and 34.965 minutes respectively.

The results of table 2, for the case study showed that the system was heavily utilized at an average of 6 servers because system utilization was almost 100%. This resulted to the longer waiting time of customers experienced at both service facilities. However in respect of this, the service rate per server were determined for the case study and a Queue Evaluation Environment was created using 2 – 12 servers to see the expected queue performance and to determine the best number of servers that gives a good trade-off between system utilization and waiting time at the collected average arrival rates of customers in the referenced service facilities.

The results from the Queue Evaluation Environment showed that 8 servers gave the best system utilization values of 0.7321 which is expected to reduce the respective customers waiting times (Ws) by 84.72% for the case study establishment. This is based on the statement of Egolum, which says that system utilization should be greater than 0 but less than 0.8 [12]. From the charts of system utilization versus waiting time plotted for the case study, it is observed that there's no significant decrease in waiting time anymore from system utilization value of 0.8, which shows that waiting time has reached its optimum at the respective best server utilization values of 0.7321 for the referenced service facility. This shows that there will be no need of making use of more than 8 servers at the respective average arrival rates of customers in the referenced service facility. Also the expected probability of system idleness is negligible at 8 server utilization because at that point, probability of system idleness has also reached its optimum and it no longer has any effect on the service systems.

From the study, it was also revealed that system utilization drops as number of server's increases; the probability of system being empty increased to optimum as number of server's increases; the probability of an arrival waiting reduces as number of server's increases; the average number in line and average number in system drops to optimum as number of server's increased; the average time in line, average time in system and average waiting time drops to optimum as number of server's increased.

Finally, from figure 5, the model for the decision support system was developed using trend line analysis. For Chester Mega Petroleum Station Enugu is given by:

$$M = 8.136\lambda^5 - 47.47\lambda^4 + 101.9\lambda^3 - 100.1\lambda^2 + 50.64\lambda - 6.569 \quad (4.34)$$

7. Conclusion

The evaluation of queuing system in an establishment is very essential for the betterment of the establishment. Most establishments are not aware on the significance of evaluating their queue performance. The implication of this is that operations managers are not able to determine the best number of servers to engage for service at various demand periods which affects their queue performance. As it concerns the case study establishment, the evaluation of their queuing system showed that there service system was over utilized which resulted to customers spending longer time than necessary before receiving service. However, the need of creating a Queue Evaluation Environment to find out the number of servers that gives the best server utilization at the collected average arrival rates became very essential. From the Queue Evaluation Environment, using 8 servers at the collected average arrival rates of customers in the referenced service facilities gave a good trade-off between system utilization and waiting time which is expected to reduce the waiting time of customers in the system while server idleness is neglected. In conclusion, the Queue Evaluation Environment created and the decision support system developed for the case study establishment will go a very long way in addressing their queuing problems.

References

[1]. Ohaneme C.O, Ohaneme L.C, Eneh I.L and Nwosu A.W (2011) "Performance Evaluation of Queuing in an Established Petroleum Dispensary System Using Simulation Technique" International Journal of Computer Networks and Wireless Communications (IJCNWC), Vol.2, No.2, April 2012

[2]. Medhi J. (2003) Stochastic Models in Queuing Theory, Amsterdam: Academic Press, 2nd pp 470 - 482

[3]. Takagi.H. (1991).Queuing analysis: A foundation of performance evaluation. European journal of operations research. Volume 1: Pp. 10-19Vacation and priority systems, Part 1. Elsevier Science Publisher, B. V.,Amsterdam.

[4]. Obamiro, J.K.,(2003) "Application of Queuing Model in Determining the Optimum Number of Service Facilities in Nigerian Hospitals", M. Sc. Project Submitted to Department Business Administration, University of Ilorin

[5]. Chinwuko E.C and Nwosu M.C (2014), Analysis of a Queuing System in an Organization (A Case Study of First Bank PLC). 2014 National Conference on Engineering for Sustainable Development, Nnamdi Azikiwe University Awka, Anambra State Nigeria: Paper F29, pp 256-267 Manufacturing and Industrial Infrastructure Issues.

[6]. Hiller S.F and Lieberman J.G. (2005) Introduction to Operations Research, Boston: McGraw Hill, Eight Edition.

[7]. Singh V., (2007) "Use of Queuing Models in Health Care", Department of Health Policy and Management, University of Arkansas for Medicals Sciences.

[8]. Adedayo A.O, Ojo, O and Obamiro J.K. (2006), Operations Research in Decision Analysis and Production Management, Lagos: Pumark Nigeria Limited.

[9]. Hiller S.F and Lieberman J.G. (2001) Introduction to Operations Research, Boston: McGraw Hill, Seventh Edition.

[10].Davis M.M, Aquilano, J.N. and Chase, B.R. (2003) Fundamentals of Operations Management, Boston: McGraw-Hill Irwin.

[11].Lapin L.L. (1981), Quantitative Methods for Business Decision, Harcourt: B. Jouanaurch, 2nd Edition. pp 182 - 201

[12].Egolum C.C (2001), Quantitative Technique for Management Decisions. Nnamdi Azikiwe University Awka, Anambra State. Pp. 91-131.

[13].Blanc, J.P.C. (2011). Queuing Models: Analytical and Numerical Methods (Course 35M2C8), Department of Econometrics and Operations Research Tilburg University, pp. 30-57.

[14].Sztrik, J. (2011). Basic Queuing Theory, University of Debrecen, Faculty of Informatics, pp 17-57.

[15].Nain P. (2004). Basic Elements of Queuing Theory Application to the Modelling of Computer Systems (Lecture notes), Pp. 16-31

Bioactive Compounds Produced by Strain of *Penicillium* sp.

M. Shaaban
Nat. Comp. Dept., Division
of Pharm. Indust., National
Research Centre, Dokki-
Cairo, Egypt

G. Elhady Sohsah
Chem. Dept., Fac. of Sci.,
Mansoura University
Mansoura, Egypt

M. Magdy El-Metwally
Bot. and Microbiol. Dept.,
Fac. of Sci., Damanhour
University, Damanhour,
Egypt

M. G. Elfedawy
Chem. Dept., Fac. of Sci.,
Mansoura University
Mansoura, Egypt

M. Abdel-Mogib
Chem. Dept., Fac. of Sci.,
Mansoura University
Mansoura, Egypt

Abstract: During our search for bioactive compounds from fungi, the terrestrial *Penicillium*sp.KH Link 1809 isolate KHMM was fermented on large scale using solid rice medium. After harvesting, working up and purification of the afforded extract using different chromatographic techniques, the bioactive metabolites viridicatol (**1**) and kojic acid (**2**) were isolated. The chemical structures of **1** and **2** were confirmed by extensive 1D and 2D NMR and mass measurements, and by comparison with literature data. The antimicrobial activity of the strain extract was studied using a panel of pathogenic microorganisms.

Keywords: *Penicillium* sp., Bioactive metabolites, Biological activity, Taxonomy.

1. INTRODUCTION

Fungi represent strongly rich producers of potent bioactive secondary metabolites [1], acting as antibiotics [2], antimycotics [3], antiviral [4], and anticancer agents [5], or are pharmacologically active in other ways [6]. The genus of *Penicillium* is one of the most prolific sources of bioactive drugs and large range of compounds including polyketides, alkaloids, terpenoids, and peptides, including a large huge number of cytotoxic compounds [7,8]. Based on the taxonomical characterization, the genus *Penicillium* is representing one of the most complex of the fungal world, with 225 species approximately and a continued discovery of new species [9]. Traditionally, species in the genus *Penicillium*, which are fundamentally saprophytic and ubiquitous, have been regarded as a fruitful investigational ground for the finding of novel bioactive compounds, leading to the discovery of blockbuster drugs, such as penicillin [10] and the anticholesterolemic agent compactin [11], miscellaneous antitumor products [12], and mycotoxins contaminating food [13]. Most of these fundamental studies were carried out on strainsfrom soil and food commodities [14].However, the search for further new bioactive compounds from fungi isolated from untouched habitats is still strongly recommended and needed to overcome the huge gap between the currently inadequate drugs and needed new ones to treat the recently discovered diseases [15,16].

In our continual program for searching of bioactive compounds from the genus of *Penicillium,* the *Penicillium*sp. KHMM isolated from Egyptian habitats [17], has been applied to biological and chemical assays, announcing potent antimicrobial activity against different pathogenic microorganism, mainly Gram positive and Gram negative bacteria, and yeast (Table 1). Chemically, the strain extract showed several middle-polar zones during TLC visualized by UV light and spraying with anisaldehyde/sulphuric acid. A large scale fermentation of the strain on rice-solid medium followed by working up and purification using a series of different chromatographic techniques afforded the bioactive compounds viridicatol (**1**) and kojic acid (**2**). The chemical structures of the isolated compounds were confirmed by extensive 1D and 2D NMR and ESI HR mass measurements, and by comparison with literature data. The antimicrobial activity of the strain was studied using a set of microorganisms. The isolation and taxonomical characterization of the fungal strain is reported as well.

2. EXPERIMENTAL

2.1 General

NMR spectra (^{1}H NMR, ^{13}C NMR, DEPT, COSY, HMQC and HMBC) were measured on Bruker Avance DRX 500 and DRX 600 MHz

spectrometers [Q] using standard pulse sequences and referenced to residual solvent signals. HR-EI-MS was determined using GCT Premier Spectrometer. The ultraviolet and visible (UV–Vis) spectra were measured on Spectro UV–Vis Double Beam PC8 scanning auto Cell UVD-3200, LABOMED, INC [Q].Column chromatography was carried out on silica gel 60 (0.040–0.063 mm, Merck [Q]) and Sephadex LH-20 as the stationary phases. Preparative TLC (0.5 mm thick) and analytical TLC were performed with pre-coated Merck silica gel 60 PF$_{254+366}$. R_f values and Visualisation of chromatograms was carried out under UV light (254 and 366 nm) and further by spraying withanisaldehyde/sulphuric acid followed by heating.

2.1. Isolation and Taxonomy of the producing strain

The fungus *Penicillium* sp. KHMM was isolated from a soil sample collected from Tag elezz agricultural research station, Dakahyia region, Egypt, at a depth of 5 cm under sterile conditions. The fungus was isolated by conventional dilution plate technique [18]. One gram of the sample was suspended in 10 mL of sterilized water and shaked for 2 h. The sediments were then left to settle down for few minutes. The supernatant was then applied to successive dilution up to 10^{-6}. An aliquot (1 mL) of the afforded suspension was spread over Czapek-Dox agar medium (gL^{-1}: 30 sucrose, 3 NaNO$_3$, 1 K$_2$HPO$_4$, 0.5 KCl, 0.5 MgSO$_4$, 0.01 FeSO$_4$, 20 agar-agar, and distilled water (1 L) at pH 7.3]), and incubated at 28°C for 14 days, the growing colonies isolated in slants of the same medium and stored in refrigerator at 4°C until use.

On Czapeck-Dox agar medium, Colonies of the fungal isolate showed vegetative mycelium abundant, colorless or pale- or brightly-colored, Colonies low to moderately deep, plane to very faintly sulcate; margins low, narrow to wide (1–5 mm), sporulation highly dense, conidia dull to greyish green exudate clear to almost a hazy yellow, sometimes absent; soluble pigment mostly yellowish orange.

In micromorphology, conidiophores typically borne from green mycelia embedded in medium, stipes mostly hyaline but sometimes green, mostly biverticillate, infrequently terverticillate, monoverticillate side branches sometimes present; Stipes very short to very long, typically rough, a minor proportion smooth to finely rough. Branches two when present, etulae 3–5 per stipe, Phialides ampulliform, 6–9 per etula, Conidia smooth, globose to subglobose.According to theses cultural and morphological features, and according to Raper and Thom (1949) [19], the terrestrial fungal strain KHMM is belonging to *Penicillium* genus. An authentic isolate of the strain is deposited at collection of Dr M. M. El Metwally, Botany and Microbiology Department, Faculty of Science, Damanhour University, Egypt.

Fermentation, working up and isolation

The *Penicillium* sp. KHMM was inoculated from well grown agar plates with dark green sporulating colonies into 0.1L sterilized glass bottles each containing modified rice medium composition: 8 g commercial rice; 10 mL distilled water. The bottles were incubated for 15 days at 30°C.

After harvesting, 50 mL of 1:1 DCM/MeOH was added to each bottle, followed by aggressive shaking for two hrs, and the afforded organic extract was decanted, filtered, and then concentrated *in vacuo* till dryness, affording (9.5 g) as brown crude extract. The crude extract was then suspended in water and applied to successive extraction in separating funnel (0.5 L), starting with pet. ether, dichloromethane (DCM), ethyl acetate and n-butanol, consequently. The corresponding fractionated extracts were concentrated to dryness affording KH1 (0.3 g), KH2 (0.4 g), KH3 (3.2 g) and KH4 (2.4 g), respectively, as reddish brown fractions. Fractions KH1 and KH2 showed undesired components according to TLC visualization and excluded.

The ethyl acetate fraction KH3 (3.2 g) was subjected to purification on silica gel column (3 × 60 cm), eluted with DCM-MeOH gradient (0.5 L DCM/3%MeOH [97:3], 0.5 L DCM-MeOH [95:5], 0.5 L DCM-MeOH [90:10], 0.5 L DCM-MeOH [80:20] , 0.5 L DCM-MeOH [1:1], 0.2 L MeOH. Three fractions were afforded according to TLC monitoring: FI (3.2 g), FII (1.2 g), FIII (0.3 g). Fraction FI is mostly containing fatsand phthalate and discarded. Purification of fraction FII using Sephadex LH-20 (MeOH), followed by silica gel column eluted by DCM-MeOH and finally with Sephadex LH-20 (MeOH) afforded a colorless solid of viridicatol (**1**, 5 mg). As the same for the ethyl acetate fraction, purification of the n-butanol fraction KH4 (2.4 g) using silica gel column followed by Sephadex LH-20 afforded a colourless solid of kojic acid (2, 8 mg).

Viridicatol (1):

Colorless solid, UV absorbing showing brown coloration on spraying with anisaldehyde/sulphuric acid: R_f = 0.45 (DCM/7% MeOH).^{1}H NMR (CD$_3$OD, 500 MHz) and ^{13}C NMR (CD$_3$OD, 125 MHz) see Table 1.
(+)-ESI MS: *m/z* (%) 276 ([M+Na]$^+$, 100), 529 ([2M+Na]$^+$, 70); (-)-ESI MS: *m/z* (%) 252 ([M-H]$^-$, 100).

Kojic acid (2):

Colorless solid, UV absorbing showing no color staining on spraying with anisaldehyde/sulphuric acid. R_f = 0.22 (DCM/10% MeOH). ^{1}HNMR (DMSO-*d6*, 500 MHz):δ 8.99 (br s, 1H, OH), 5.69 (br s, 1H, OH), 7.97 (d, 1H, *J*=7.97, H-2), 6.34 (s, 1H, H-5), 4.28 (s, 2H, H$_2$-7). ^{13}CNMR (DMSO-*d6*, 125 MHz): δ 174.2 (C$_q$-4), 168.2 (C$_q$-6), 153.0 (C$_q$-3), 138.9 (CH-2), 110.0 (CH-5), 59.7 (CH$_2$-7).

Antimicrobial Assay Using Agar Diffusion Test

Antimicrobial activity testing of the crude extract of the fungal isolate KH was carried out against a set of microorganisms using the agar diffusion technique. Paper-disk diffusion assay [20] with

some modifications has been followed to measure the antimicrobial activity. 20 mL of medium seeded with test organism were poured into 9 cm sterile Petri dishes. After solidification, the paper disks were placed on inoculated agar plates and allowed to diffuse the loaded substances into refrigerator at 4 °C for 2 h. The plates were incubated for 24 h at 35 °C. Both bacteria and yeasts were grown on nutrient agar medium: 3g/L beef extract, 10g/L peptone, and 20g/L agar. The pH was adjusted to 7.2. Fungal strain was grown on potato dextrose agar medium (g/L): Potato extract, 4; Dextrose, 20; Agar No. 1, 15 (pH 6). The samples were dissolved in DCM/10% MeOH. Aliquots of 50 μL (= 50 μg) were soaked on filter paper discs (9 mm) and dried at room temperature under sterilized conditions. The paper discs were placed on inoculated agar plats and incubated for 24 h at 37°C for bacteria and 48 h (30°C) for the fungi. After incubation, the diameters of inhibition zones were measured with a wide panel of test microorganisms comprising Gram positive bacteria (*Bacillus subtilis* ATCC6633 and *Staphylococcus aureus* ATCC6538-P), Gram negative bacteria (*Pseudomonas areuginosa* ATCC 27853), yeast (*Candida albicans* ATCC 10231, and the fungus *Aspergillus niger* NRRL A-326.

3. RESULTS AND DISCUSSIONS

3.1. Fermentation and Structure Elucidation

The *Penicillium* sp. KH Link 1809 was cultured on rice solid medium. Biologically, the afforded extract of the strain exhibited high activity against Gram positive, Gram negative bacteria, and yeast, namely: *Bacillus subtilis* ATCC6633, *Staphylococcus aureus* ATCC6538-P), *Pseudomonas areuginosa* ATCC 27853, and *Candida albicans* ATCC 10231 (Table 1). In the chemical screening monitored by TLC, the fungal extract exhibited two major UV absorbing bands with a narrow polarity range, one of them exhibited a brown staining with anisaldehyde/sulphuric acid. Some other unpolar bands were detected as violet-blue on spraying with the same reagent, which might be attributed to some fats, phthalates or sterols. Separation of the strain produced middle polar bioactive metabolites using a series of chromatographic techniques (see experimental section) afforded Viridicatol (**1**) and kojic acid (**2**).

Viridicatol (1)

As Colorless solid, compound **1** was obtained showing UV absorbance at 254 nm during TLC, which stained as brown on spraying with anisaldehyde/sulphuric acid. The molecular weight of **1** was deduced as 253 Dalton according to ESI MS modes, such that two *pseudo* ion peaks were exhibited at *m/z* 276 [M+Na] and 529 [2M+Na] were exhibited, and one ion peak was visible at *m/z* 252 [M-H] in the ESI negative mode. HRESI MS of 1 revealed the corresponding molecular formula as $C_{15}H_{11}NO_3$, bearing eleven double bond equivalents.

Based on the ^{1}H, ^{13}C and HMQC NMR spectra (CD$_3$OD) (Table 2), the numbers of hydrogen and carbon atoms were in agreement with the molecular formula. Based on the ^{1}H NMR spectrum, H,H COSY experiment, and coupling constants, eight proton signals (δ8.15, 7.36 [2H], 7.34, 7.26, 6.88, 6.83, 6.82),were observed in the aromatic region, indicating the presence for *ortho-* and *meta-*disubstituted benzene systems. In the ^{13}C NMR spectrum and HMQC experiment, one carbonyl carbon signal (δ159.1) and 14 sp^2carbon signals, including eight methine signals (δ129.1, 124.9, 122.4, 121.6, 120.7, 116.5, 115.0, 114.5), were observed. According to the above data, followed by study the structure using HMBC experiment (Fig.1), the structure of the title alkaloid was assigned as viridicatol (**1**) [21,22]. Viridicatol was isolated previously along with viridicatin from a strain of *P. viridicatin* [23]. Functionalized 4-arylquinolin-2(1*H*)-ones constitute generally, a valuable class of biologically active molecules, including several fungal metabolites such as viridicatin [24] and 3-*O*-methylviridicatin [25]. Viridicatol was reported to show cytotoxicity toward KB, KBv200, A549, hepG2, MCF7, K562, SMMC7721, and SGC7901 tumor cell lines with IC$_{50}$ values of 25.0, 16.5, 60.0, 85.0, 45.0, 25.0, 80.2, and 80.0 g/mL, respectively.

Kojic acid (2)

As polar Colorless solid, exhibiting UV absorbance at 254 nm during TLC, compound 2 was obtained, which showed no color staining on spraying with anisaldehyde/sulphuric acid. Based on EI MS, the molecular weight of **2** was deduced as 142 Daltons, and the corresponding molecular formula was established by EI HRMS as $C_6H_6O_4$, containing four double bond equivalents. The ^{1}H NMR spectrum (DMSO-d_6) of **2** displayed two broad signals at δ 8.99 and δ5.69, which could be attributed to exchangeable protons of phenolic and aliphatic hydroxyl groups, respectively. Two further proton singlets each of 1H were visible at δ7.97 and 6.34, in addition to anoxy-methylene signal as singlet at δ4.28. Based on the ^{13}C and APT NMR spectra, three quaternary sp^2 carbon signals were displayed at δ174.2, 168.3 and 153.0, representing a lactone carbonyl (δ174.2), oxygenated sp^2 carbon (168.3) in β-position with respect to a lactone carbonyl, and phenolic one (δ153.0). Two sp^2methine carbons were exhibited at δ138.9, and 110.0, along with an oxy-methylene carbon (δ59.7) as well. Based on the revealed chromatographic features, spectroscopic data, and search in the corresponding data bases (AntiBase [1], Dictionary of Natural Products (DNP) [26], and Scifinder [27], compound **2** was confirmed as kojic acid. Biologically, kojic acid is showing high toxicity and high antibiotic activity against Gram-Positive and Gram negative bacteria, and antifungal properties [28]. Kojic acid is a well-known as tyrosinase inhibitor, and hence it has been used as whitening or anti-hyperpigment agent because of its ability to suppress dermal-melanin production [29].

3.2. Biological activity studies

Antimicrobial activity testing of the stain extract of the terrestrial fungus *Penicillium* sp. KH Link 1809

was carried out against five microorganisms using the agar diffusion technique. The extract showed high antimicrobial activity (50 μg per disk) against Gram-positive bacteria (*Bacillus subtilis* ATCC6633 [27 mm], *Staphylococcus aureus* ATCC6538-P [26 mm]), Gram-negative bacteria (*Pseudomonas aeruginosa* ATCC 27853 [31 mm]), and *Candida albicans* ATCC 10231 [25 mm]. Nevertheless, the extract showed no activity against *Aspergillus niger* NRRL A-326. (Table 1)

Acknowledgement
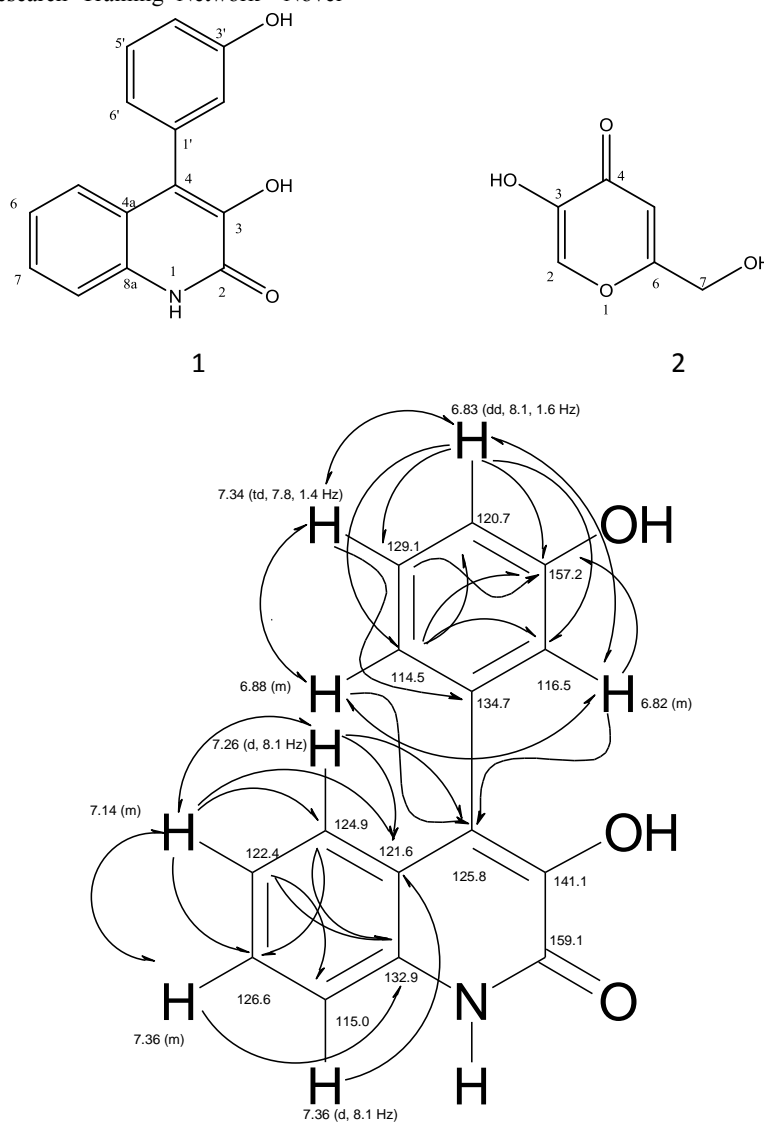
Fig.1: H,H COSY (↔) and HMBC (→) correlations of Viridicatol (**1**)

Table 1: Antimicrobial activities of *Penicillium sp.*KHMM Link 1809 extract

Extract	Diameter of zone inhibition (mm)			
	P. aeruginosa	*St. aureus*	*B. subtilis*	*C. albicans*
KH	31	26	27	25

Table 2: ^{1}H (CD$_3$OD, 500 MHz) and ^{13}C (CD$_3$OD, 125 MHz) NMR data of Viridicatol (**1**)

Nr.	δ_C	δ_H
2	159.1	
3	141.1	
4	125.8	
4a	121.6	8.15 (dd, 8.0, 1.5)
5	124.9	7.26 (d, 8.1)
6	122.4	7.14 (m)
7	126.6	7.36 (m)
8	115.0	7.36 (d, 8.1)
8a	132.9	
1'	134.7	
2'	116.5	6.82 (m)
3'	157.2	
4'	120.7	6.83 (dd, 8.1, 1.6)
5'	129.1	7.34 (td, 7.8, 1.4
6'	114.5	6.88 (m)

4. REFERENCES

[1] Laatsch, H. (2014) *AntiBase*, A Data Base for Rapid Structural Determination of Microbial Natural Products, and annual updates, Wiley-VCH, Weinheim Germany.

[2] Butler, M. S. (2004) J. Nat. Prod. 67, 2141–2153.

[3] Brady, S. F., Clardy, J. (2000) J. Nat. Prod. *63*, 1447–1448.

[4] Singh, S. B., Zink, D. L., Guan, Z., Collado, J., Pelaez, F., Felock, P.J., Hazuda, D. J. (2003) Helv. Chim. Acta 86, 3380–3385.

[5] Zhang, H. W., Song, Y. C., Tan,R. X. (2006) Nat. Prod. Rep.*23*, 753–771.

[6] Song,Y. C., Li,H., Ye,Y. H., Shan,C. Y., Yang,Y. M., Tan,R. X. (2004) FEMS Microbiol. Lett. 241, 67–72.

[7] Bugni, T.S., Ireland, C.M. (2004) Nat. Prod. Rep. 21, 143–163

[8] Grovel, O., Kerzaon, I., Vansteelandt, M., Petit,K. E., Biard, J. F., Pouchus, Y. F. (2008) Planta Med 74 - SL99.

[9] Jens C. Frisvad. (2014) Taxonomy, chemodiversity, and chemoconsistency of *Aspergillus*, *Penicillium,* and *Talaromyces* species. Front Microbiol. 5, 773.

[10] Demain A.L., Elander R.P. (1999) The β-lactam antibiotics: Past, present, and future. Antonie Van Leeuwenhoek. 75, 5–19

[11] Chakravarti R., Sahai V. (2004) Compactin—A review. Appl. Microbiol. Biotechnol. 64, 618–624

[12] Nicoletti R., Ciavatta M.L., Buommino E., Tufano M.A. (2008) Antitumor extrolites produced by *Penicillium* species. Int. J. Biomed. Pharm. Sci. 2, 1–23

[13] Frisvad J.C., Smedsgaard J., Larsen T.O., Samson R.A. (2004) Mycotoxins, drugs and other extrolites produced by species in *Penicillium* subgenus *Penicillium*. Stud. Mycol. 49, 201–241.

[14] Nicoletti, R., Trincone, A. (2016) Bioactive Compounds Produced by Strains of *Penicillium* and *Talaromyces* of Marine Origin.Mar Drugs. 14, 37; doi: 10.3390/md14020037.

[15] Strobel, G., Daisy, B. (2003) Microbiol. Mol. Biol. Rev. *67*, 491–502.

[16] Hassan, A. E. H. A. (2007) Dissertation, University of Düsseldorf, Germany, 2007.

[17] Shaaban, M., El-Metwally, M. M., Laatsch, H. (2016) New Bioactive Metabolites from *Penicillium purpurogenum* MM. Zeitschrift Naturforsch. *71B,* 287-296.

[18] Hayakawa, M. and Nonomura, H. (1987). Humic Acid-Vitamin Agar, a New Medium for the Selective Isolation of Soil Actinomycetes. J. Ferment. Technol. 65, 501-509.

[19] Raper, K. B.,Thom. (1949) A manual *of Penicillium*. Williams and Wilkins Baltimore. USA.

[20] Collins, C. H., Lyne, P. M. (1985) Microbiological methods, 5th ed. Butterworth and Co. Pub. Ltd., London and Toronto pp. 167-181.

[21] Mohammed, Y. S., Luckner,M. (1953) Tetrahedron Lett.*,* 4, 1953.

[22]   Wei,M.Y., Yang,R.Y., Shao, C.-L., Wang, C.Y., Deng, D.S., She, Z.G., Lin, Y.C. (2011) Isolation, Structure Elucidation, Crystal Structure, and Biological Activity of a Marine Natural Alkaloid, Viridicatol. Chemistry of Natural Compounds, 47, 322-325.

[23]   Ciegler,A., Kadis,S., Ajl, S. J. (1971) Microbial Toxins: A comprehensive Treatise; Fungal Toxins. Metabolites of *Penicillium viridicatum* and closely related species. Copyright by Academic press, INC 111 Fifth Avenue, New York 10003, Volume VI, 509.

[24]   Luckner, M., Mothes, K. (1962) Tetrahedron Lett., 3, 1035.

[25]   Austin, D. J., Myers,M. B. (1964) J. Chem. Soc*.,* 1, 1197.

[26]   Dictionary of Natural Products on CD-ROM, Chapman & Hall Chemical Database, 2016.

[27]   https://scifinder.cas.org/scifinder.

[28]   Morton, H , Kocholaty, W., Junowiczkocholaty, R. , Kelner, A. (1945) Toxicity, and Antibiotic Activity of Kojic Acid, Produced by *Aspergillus Leuteo-Virescens1,* J. Bacteriol,.50 ,579-584.

[29]   Uchida, R., Ishikawa, S., Tomoda, H. (2014) Inhibition of tyrosinase activity and melanine pigmentation by 2-hydroxytyrosol. Acta Pharmaceutica Sinica B., 4, 141–145.

Design and Implementation of Data Integration and Analysis System for Campus Card

Zhao Lei
Shandong University of
Science and Technology,
Qingdao 266590, China

Wang Le
Shandong University of
Science and Technology,
Qingdao 266590, China

Cong Liu*
Shandong University of
Science and Technology,
Qingdao 266590, China

Abstract: Campus card is used to record learning, living and entertainment data of students during school learning period. It contains numerous valuable information of student daily behavior. The effective integration of campus card data is the basis for further mining and analyzing of student behavior. This paper presents the design and implementation of a data integration and analysis system for campus card. This system is analyzed and designed by Unified Modeling Language (UML) and implemented by Java. The main input involves the records of consumption, entrance guard, library and other types of related data. This system can used to integrate common types of source data and provide several kinds of statistical charts and tables to illustrate the analysis results. Moreover, the system supports to add new functions in terms of plugins.

Keywords: Data integration and analysis; student campus card data; statistical charts; system analysis and design

1. Introduction

The rapid development of information and network technology inspired the widely application of campus cards in lots of schools, especially in universities and colleges. The campus card records almost all the information of students' daily living in school, including borrowing books from library, consumption, dining, shopping, surfing the Internet, and seeking doctors and so on. These data imply many valuable information that can assist the administration managers and teachers to improve the effectiveness and efficiency. Nowadays, the research on the management and mining of campus card data become a hotspot [1]-[2].

Bowman et al. [1] first introduce the development of campus card monitoring system using Cacti Framework and Nagios Plug-in. It can monitor network traffic to evaluate operation of network. *Li et al.* [2] introduce some ways to improve the campus card system's stability and reliability, taking the campus card system of Fudan University as the example. The ways includes hierarchical authority, improvement of system automation, and system monitoring etc. *R. Hu* [3] introduces the composition, function, system design, syestem security and network design of a campus card system. *X. Dong et al.* [4] analyze the effectiveness of some clustering algorithm, including *CobWeb* and *EM*, in mining campus card data, based on consumption data and *Weka* platform. *Xu et al.* [5] use decision-tree-classification algorithm to analyses campus card data about physical exercise. Through this way, students' situation of exercise could be evaluated. *Jiang et al.* [6] analyze students' consumption behavior and learning behavior based on campus card data. This paper uses K-means clustering algorithm to analyses the consumption location and consumption level about boys and girls in school, and it also analyses the relations among the number of using card, the number of borrowing books, scholarship and academic achievement. *Chen* [7] makes use of the campus card data to excavate and analyze the user' behavior of dining and consumption, and puts forward some suggestions on the operation of dining hall and management of school, taking Liaoning Medical University as an example. *Fei et al.* [8] optimize impoverished students determination by K-means clustering algorithm to analyze the consumption data of campus card. *Li* [9] analyses the students' consumption and behavior according to campus card data of *Beijing Jiaotong University*. *Jin et al.* [10] expound the support of campus card data to schools' decision making based on the analysis of dining records, consumer records, etc.

Most of the above-mentioned references mainly focus on data mining and analysis for the campus card data, and lack of the universal platform of campus card's data integration and analysis. To solve this problem, this paper describes the design and implementation of data integration and analysis system for campus card, to support the improvement of university information management.

2. Basic Introduction of Data Integration and Analysis Systems for Campus Card

This system mainly realizes the integration, statistics and analysis of campus card log data. The detailed requirement are as follows.

(1) *Data Integration*. The system integrates data from diverse data sources into system. The supported data sources include Excel, HTML, CSV, structured text, XML and database. During integration, the users choose the file to store the source data in first, and then specify relationship between target data format and source data format. Finally, the system will check the format and import the source data if the source data is okay.

(2) *Data Query*. It supports basic query, complex query and expert inquiry. The basic query allows users to select one or a few conditions for a query. The complex query enables users to query by multiple conditions and combination of multiple tables. The expert inquiry supports users to write *Structured Query Language* (*SQL*) statement autonomously or load files containing queries to query.

(3) *Data Statistics*. It supports simple and comprehensive data statistics. The former implements simple statistics of the frequency of consumption, amount of consumption, number of access dormitory and so on. Comprehensive statistics realizes the statistical analysis of multiple data.

(4) *Data Presentation*. For statistical results, it supports the ordinary ways to display, such as line chart, histogram, scatter plot and other types of charts and tables. As for the students' activities, it can show them by two ways.

a) *Directed Graph*. It includes static and dynamic illustrations. Static display shows students' activity trajectory by weighted directed graph, which uses edge weight and thickness of line to represent the number of times through the corresponding path. Similar to "step into" debug method of a program, the manual dynamic display allows users to follow the student's trajectory step by step by an interactive way. Instead, the automatically dynamic display method illustrates students' trajectory automatically according to the user-defined time interval.

b) *Sequence Diagram*. It shows the changes of students' location according to the time sequence. Each point of a sequence diagram represents a location where the given student used his/her campus card.

(5) *User-defined Data Mining and Analysis Plugin*. The system provides interfaces to import user-defined plug-ins. Through this way, the users can realize special query and statistics functions.

3. System Design
3.1 Overall Design of System
As we can see, the system adopts *Model-View-Controller* (*MVC*) model. MVC model achieves the separation of logic, data and user interface, and it integrates the business logic into one component. Therefore, it is not required to rewrite the business logic while improving customized interface and user interaction. The basic process is:

(1) The controller creates the model and view, and the model is associated with the view.

(2) The user interacts with the system through the controller to change the state of the model.

(3) When the state of the model changed, the controller automatically refresh the related view to show query results.

3.2 Metadata Definition
In order to support diverse data sources, the metadata should be given in first. Taking consumption record, accessing dormitory records and library records of campus card as examples, the metadata standards are defined as follows.

(1) Consumption Record. Name, POS (Point of Sale) number, system code, operation type, balance of the card, account balance, occurrence time, recorded time, account number, account of the other side, number of using card, consumption type, serial number, student number.

(2) Access Dormitory Record. Date, time, student number, name, location, card number, department, event description.

(3) Library Record. Number, date, time, card number, card version, the type of card, controller number, read head number, door number, event code, illegal card number, controller number, door ID, work number, department ID, work number version, photo tagging. The first twelve fields are not allowed to empty, the rest of the fields are allowed to be empty.

3.3 Database Design
Converting consumption records, entrance guard records and library borrowing records to entities, the entities are consumption entity, dormitory building entity and library entity.

In order to support the above entities, it need the following entities and their attributes:

(1) POS entity. Number, IP address, system code, area code, Merc code (Merchant code), POS type, location.

(2) Student entity. employee number, first name, last name, ID department, work number, traffic password, telephone number, room number, whether to display name, whether to allow the dial, display name, employee version number, department effective time.

(3) Plug-in entity. Plug-in number, plug-in name, plug-in path, plug-in package name, and class name.

4. Key Technologies of Systems
4.1 Data Integration and Query
The source data may be raw or wrong, so the system use the *Extract-Transform-Load* (*ETL*) method to integrate a given data source into the system. The main techniques include:

(1) Different procedures are used to import different data sources into this system. Formatted text files can be direct imported into database using SQL statements. Excel files should be read into list, and then are imported to the database. XML files are imported into database after being traversed by the data in each node. The other database exported files should be used character stream to parse, and then, extract SQL statement and use addBatch operation method to realize the bulk insert.

(2) Some views of the data in database should be constructed to improve the speed of query in the cascade query. These views include Iodetails, PaymentAndPos, PositionChanged, etc.

(3) For analyzing students' behavior trajectory, the semantics of consumption locations are not clear. Hence, the consumption locations are classified into restaurants, hospitals, Internet bars, supermarkets, water room and other places according to their properties. User can view students' track in the some locations where they want to see, and ignore the meaningless sites for them. It is easy for user to analyze students' behavior.

4.2 Display of Statistical Analysis Using Directed Graph
Based on the campus card data, a student's trajectories during a certain period can be acquired. There are three kinds of display directed graphs: summary display, interactive single step display and automatic single step display.

To realize this function, the observer pattern is used to achieve a one-to-many dependency between objects. The class BTGraphY is observed, and observers include BTView, BTNext, BTDynamic and so on. The method "GraphArray" of class BTGraphY obtains the results by querying the database, then the method "NotifyGraphBT" notifies its boservers. The observers automatically update their views in next. This design pattern is easy to extend new function to reduce the times of data query. The class diagram is shown in *Fig*. 1.
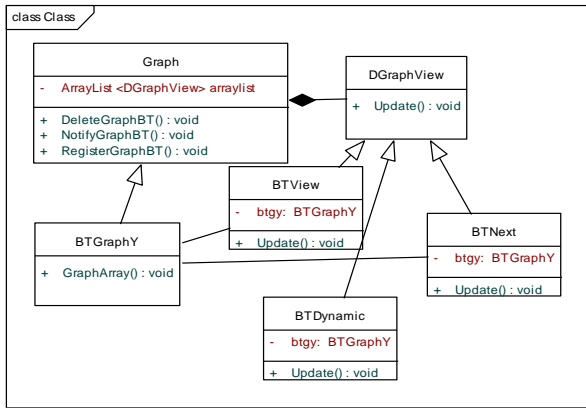
Figure 1. Class Diagram

Besides, we adopt the following methods to show the directed graph. The direct graph is shown *Fig*. 2.
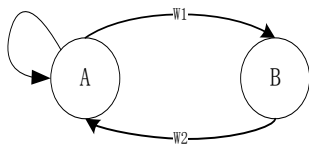

Figure 2. Directed Graph

(1) The point *A* and *B* represent location where student used campus card. The edge from *A* to *B* (or from *B* to *A*) represents student moved from position *A* to position *B* (or from position *B* to position *A*).

(2) Edge weight. The weight w_1 (or w_2) is the number of times of a student's going through the corresponding path.

(3) Edge thickness. We set an upper and lower bounds for the thickness of edges. The Edge thickness is given by equation

$$ET_i = \frac{(W_{max} - W_{min} + 1)}{(W_n - W_0 + 1)} \times (W_i - W_{min}) + W_0$$

(1).

ET_i is the thickness of edge *i*. W_{max} is the maximum of weight. W_{min} is the minimum of weight. W_n is the upper bounds for the thickness of edges. W_0 is the lower bounds for the thickness of edges. W_i is the weight of edge *i*.

4.3 Demonstration

The system can automatically demonstrate not only the trajectory of a single person, but also trajectories of multi-persons. The realization of automatic demonstration needs to add a timer to achieve a regular task. For the single step automatic display of directed graph, the differences of demonstrations between single person and multi-persons are as follows:

(1) For multi-persons' results, each person is allocated to a random color based on the person's ID, and the colors are stored in a hash table to prevent duplication of color.

(2) The color of an edge is the color of the corresponding person.

(3) The results are sorted according to two keys: The first is persons' *ID*, and the second is record's time.

4.4 User Defined Plug-in

The key to achieve the user defined plug-in accessing are as follows:

(1) Configuration plug-in of main program: It uses plug-in ID, plug-in name, plug-in address, package name and class name in database.

(2) Instantiate object of plug-in: Dynamic loading of Java is mainly achieved by ClassLoader. Java has provided a ready-made ClassLoader to implement class URLClassLoader, which can load the jar specified by URL .

This system provides the interface of basic data integration, query and statistics, so that the user can integrate the data mining analysis plug-in into system.

5. Implementation and Application

We can analyze the relationship between student's achievement and the number of times of accessing library, using the records of the second semester in 2012-2013 school year. The students are selected from different classes of different professions, and half of them are the top 30% in the class and the remains are the bottom 30%. *Fig.3* shows the query result of the library records of 90 students, and we can draw the following conclusions based on these results:

(1) Generally speaking, more than 82% excellent students go to the library more frequently than the others. However, there also a few excellent students rarely go to the library. Numeric result.

(2) More than 84% poor students barely even went to library. However the times of 15.5% poor students to go to library is more than 20 times. There also exists poor students who like to study in the library, but the number of them is smaller.

(3) Overall speaking, There not many students frequently go to the library whether poor students or eugenics.
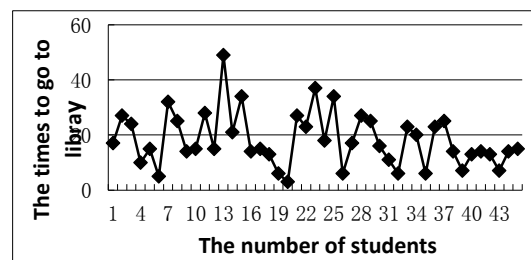
(a) Outstanding studens
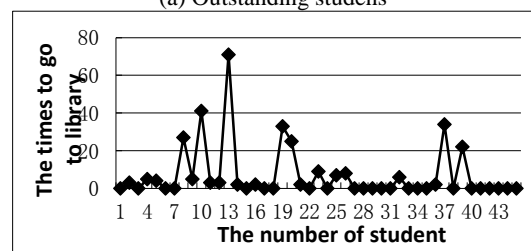
(b) poor performance students

Figure 3. Statistical results of library entry and exit of outstanding students

6. Conclusion and Future work

This paper presents a comprehensive introduction of the design and implementation of a Data Integration and Analysis System for Campus Card. The system is designed based on *UML* (*unified modeling language*) and is implemented using Java. We adopt MySQL as our database. The main functions of the system include data integration, data statistics, data query, data mining and data analysis. Compared with the other similar software, the proposed system has the following advantages: (1) Our focus is the campus card data, so it can meet the

requirements of universities better than the other general data integration system; (2) The system uses dynamic graphs to show the results, which are more vividly and easily to acquire model for the behavior of the students; (3) With classified and inductive data, we can get the student's behavior regularity more easily; and (4) It has better scalability with the open interfaces to allow users to import their own plugins.

However, this system can be improved at least in the following aspects: (1) we plan to provide more open interfaces to import plugins based on the other programming languages besides Java; and (2) we are trying to build the behavior model of students with cluster analysis and analyze their typical behavior patterns.

7. Reference

[1] Bowman, M., Debray, S. K., and Peterson, L. L. 1993. Reasoning about naming systems. .[1] W. Zhou, C. Chen, S. Li: Experimental technology and management, 2011, 28(4): 246-249.

[2] S. Li, W. Zhou, Y. Mi: Experimental technology and management, 2011, 28(6): 7-10.

[3] R. Hu: Research in laboratory, 2012, 31(7): 404-420

[4] X. Dong, H. Zhang: Computer systems & applications, 2014, 23(1):158-161.

[5] H. Xu, B. Sun, J. Gao, et al, in: Proc. Int. Conf. Comp. Sci. Sports Engin., 2010: 583-586.

[6] N. Jiang, W. Xu: Microcomputer Applications, 2015, 31(2):35-38.

[7] F. Chen: The Chinese Journal of ICT in Education, 2012, 9: 47-49.

[8] X. Fei, X. Dong, H. Zhang: Computer Knowledge and Technology, 2014, 10(20): 4934-4936.

[9] S. LI: Railway computer application, 2010, 19(6) : 55-58.

[10] P. Jin, X. Wang: Journal of East China Normal University (Natural Science), 2015, S1:525-529.

Designing of Trajectory Similarity Calculation System Based on Latitude and Longitude

Le Wang
Shandong University of
Science and Technology,
Qingdao 266590, China

Lei Zhao
Shandong University of
Science and Technology,
Qingdao 266590, China

Cong Liu*
Shandong University of
Science and Technology,
Qingdao 266590, China

Abstract: This paper explains the algorithm of similarity calculation: K-means [1] clustering algorithm and Apriori algorithm, achieves the similarity calculation between two or more trajectories by using these two algorithms. Developing this system in the b / s structure model based on similarity calculation. This paper mainly describes system general design and system implementation. Among them, system general design includes the main business design of the system, system function module and database design; system implementation includes the setting of software and hardware and system function.This paper ends with a conclusion and prospect, pointing out problems and inadequacies, discussing the questions of the system functions, algorithm optimization and the like.

Keywords: similarity calculation; K-means clustering algorithm; Apriori algorithm; trajectory; b / s

1. INTRODUCTION

With the rapid development of mobile communication equipment facilities and positioning service, resulting in lots number of trajectories which based on the latitude and longitude, the vast number of data has attracted the attention of people. How to manage and use the information effectively has become a hot research currently. The trajectory model is a popular and promising approach which is used to find the similarity of the trajectory, that is also used to recommend interest points.

This paper implements the trajectory similarity calculation using the K-means clustering algorithm and Apriori algorithm, and displays trajectories visually by baidu gis map. Designing the trajectory similarity calculation system based on latitude and longitude in the b/s structure mode.

2. RELATED WORK

Ying et al. [2] propose a method to compare user similarity semantically on the level of frequent patterns. They use PrefixSpan to mine frequent patterns and develop a similarity measure, called maximal semantic trajectory pattern similarity (*MTP similarity*). Maximal trajectory patterns, or maximal patterns, are those patterns that are not contained in any other frequent patterns. In the MTP similarity measure, the comparison between users is based on the comparison between maximal patterns. *Chen et al.* [3] improve the MTP similarity measure by remedying a defect which is that when comparing two identical users using the similarity measure the similarity value is not necessarily one, and extend it to take temporal information into account. *Ruipeng LU* [4] proposes a new similarity measure called the *CPS-based similarity measure*, by directly comparing two users' frequent pattern sets instead of being based on the comparison of patterns. This measure eliminates the above-mentioned defects of the (improved) MTP measure.

3. ALGORITHM OF SIMILARITY CALCULATION

3.1 K-means Clustering Algorithm

The basic philosophy of K-means algorithm is taking *k* as a parameter, and dividing *n* object into *c* cluster, making the similarity in the cluster high while between clusters low. It uses the mean value of objects in each cluster to calculate the center of the cluster to make it as similarity standard. When the algorithm starts, it selects *c* point from all point as initial central point of each cluster, then iterates the residual point one by one, and entrusts its recent cluster with it, then revises the center of gravity of the cluster. Repeat like this, until the criterion function restraining.

The distance between each point is indicated by the distance between two points, which usually uses Euclid distance, and its formula is as equation (1):

$$D(i, j) = \sqrt{\left|x_{i1} - x_{j1}\right|^2 + \left|x_{i2} - x_{j2}\right|^2 + \ldots\ldots + \left|x_{ip} - x_{jp}\right|^2} \quad (1)$$

P is the number of the attributes of each point.
The value of the cluster center is the mean value of the cluster, its formula is as equation (2):

$$m_i = \frac{1}{n} \sum_{i=1}^{n_i} x_i \quad (2)$$

In the formula, *n* is the number of mid-value of the cluster.
Its computational process is as follows:
The input: the number of the clusters (k) and the number of its data object (n).
The output: *k* cluster, which make the criterion function the smallest.
(a) *K* data object are selected stochastically, and these *k* object initial represents the centre of each cluster.
(b) According to the mean value of the data object of the cluster, each data object is assigned to the nearest cluster.
(c) Renew the mean value of each cluster.
(d) The second and third one is executed repeatedly, until the criterion function is restraining.

3.2 Apriori Algorithm

Among the best known algorithms for association rule induction is the apriori algorithm. This algorithm works in two steps: In a first step the frequent item sets are determined. These are sets of items that have at least the given minimum support. In the second step association rules are generated from the frequent item sets found in the first step. Usually the first step is the more important, because it accounts for the greater part of the processing time.

In order to make it efficient, the apriori algorithm exploits the simple observation that no super set of an infrequent item set can be frequent .

The original pseudocode by Agrawal is offered in Algorithm1.

Algorithm 1. The Apriori algorithm.

(1) L_1 -{large 1-itemsets};

(2)**for** $(k = 2; L_{k-1} \neq \Phi; k++)$ **do begin**

(3) C_k = apriori-gen(L_k -1); //New candidates

(4)   **forall** transactionst $t \in D$ **do begin**

(5) $C_t = subset(C_k, t)$; //Candidates contained int t

(6)   **forall** candidates $c \in C_t$ **do**

(7)   c.count++
(8)   **end**

(9) $L_k = \left\{ c \in C_k \middle| c.count \geq \min\sup \right\}$;

(10)   **end**

(11) $\text{Answer} = \cup_k L_k$;

In the algorithm *apriori-gen* is a function made up of two phases: union and pruning. In the union phase (see Algorithm 2), all k item sets candidates are generated.

Algorithm 2. Union phase of the Apriori.

insert into C_k

select $\text{p.item}_1,..., \text{p.item}_{k-1}, \text{q.item}_{k-1}$

from $L_{k-1} p , L_{k-1} q$

where $p.item_1 = q.item_1,..., p.item_{k-2} = q.item_{k-2}$,

 $p.item_{k-1} < q.item_{k-1}$;

Now in the pruning phase (see Algorithm 3), which gives the name to the algorithm, all candidates generated in the union phase with some non-frequent (k-1) item set are removed.

Algorithm 3. Pruning phase of the Apriori.

forall   item sets $c \in C_k$

    forall (k-1)-subsets s of c **do**

        If $s \notin L_{k-1}$ **then**

        Delete c form C_k ;

4. OVERALL DESIGN
4.1 The Main Business Of the System

Firstly, the user logins the system. If the user does not have an account, the system will prompt the user to register in the system. After that the user can use the system properly. Secondly, the user needs to import trajectories into the system, and then selects two or more trajectories. Thirdly, the user can compare the similarity of trajectories to find useful rules and find similar trajectories. What' s more, the system will also recommend interest points to users according to two or more trajectories that users have already chosen. In the end, the user exits or logs off the system if the user has finished all operations.

The similarity calculation is the main business of the system that is ran automatically in the background. Above all, the data should be divided into k classes through k-means algorithm. Then, the frequent sequence is obtained according to apriori algorithm. Finally, the similarity calculation is

achieved according to the concept of the longest common sub sequence .

4.2 System Function Module

This system can realize these functions: registration, login, exit, query and modify the user information, import the trajectory data, select the trajectory, view the trajectory through different ways (form ,text and baidu gis map), show the result of the similarity calculation and interest points and so on.

The structure of the main function module of the system is shown in *Fig.1*. The main function module of the system includes three parts: import data module, process data module and display data module. Import data module imports the trajectory to the database in order to use the data efficiently. Process data module clusters the trajectory, and calculates the similarity of the corresponding trajectory by using the similarity algorithm. Display data module displays the trajectory through baidu gis map or table. It's also shows the result of the similarity calculation to users in a friendly way.
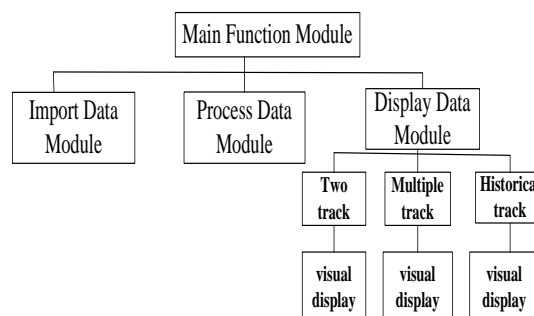

Figure 1. **Main Function Modules Of the System**

Import data module includes three parts: select data, read data and save data to database. The user will be alerted to select a file again, if the imported file format is wrong or the user does not choose a file.

The main function of process data module is to achieve cluster algorithm, and find the longest common sub sequence by using apriori algorithm.

Display data module is to show the similarity value and the trajectory information friendly.

4.3 Database Design

The system uses MySQL database, because MySQL database is multi thread, multi user operating SQL database server, and supports windows, Linux, UNIX, SUNOS etc. multiple operating system platforms, and is beneficial to transplant the system, could reduce pressure on the server. The database of this system is divided into four tables: User information table (user), trajectory file table (path), table associated with user table and trajectory file table(user_data), track detail information table (track_data). Among them, the user table stores user *ID,* user name, user password. Path table stores file *ID,* file name, file imported time. The user_data table stores the user *ID* and the file *ID* by associating the user table with the path table.Track_data table stores file *ID*, longitude, latitude, time, serial number, cluster mark, in which the path file *ID* and the user *ID* is auto-incremented .

5. SYSTEM IMPLEMENTATIONS

5.1 Software And Hardware Environment

This system runs on Windows 7. The following is the hardware and software environment of the system :

Hardware Environment:

CPU : Intel (R) Core (TM) -2450M CPU i5 @ 2.50 GHz

Memory: 4GB or above

Operating System: 64 bit

Broadband: 10.0M or above

Software Environment:

Compiler: MyEclipse 10.5

Java Running Environment: java 7.0

DataBase: MySQL 5.5

JSP Server: tomcat 6.0

5.2 System Function

The main function of the system is to obtain the longest common sub sequence through the frequent sub sequences that are based on the Apriori algorithm. The paper defines a method to calculate the similarity between two trajectories through the concept of the longest common sub sequence. The similarity between the two trajectories P and Q is defined as

$$sim(P,Q): sim(P,Q) = \frac{2 \cdot lenLCS(P,Q)}{len(p) + len(Q)} .$$

The *lenLCS (P,Q)* is the length of the longest common sub sequence of P and Q, *len(p)* is the length of the trajectory *p*. The *sim(P, Q)* is used to get the similarity of the trajectory. Finally, the system will present the trajectory map and the result of the similarity calculation to users in a friendly way.

6. SUMMARY

The structure of B/S (*browser/server*) was used in this system, because of its merits, which made the system developed easily, made the system more flexible, and made the maintenance and the management of system easier in the later stage.

The system has obtained the result of the user trajectory similarity calculation according to the K-means algorithm and Apriori algorithm. Meanwhile, for the result, this paper puts forward an recommendation mechanism to recommend interest points.

The design of the system can also be used to manage students to access to the Internet healthy and safe. As of way of learning and communicating for students, the internet and its management will play a vital role in students' physical and mental health as well as the improvement of their academic performance. Discovering the hiding and potential rules through computing the similarity of the Internet log for good students and bad students. Improving the poor performance of students in grades during the result of similarity calculation.

From an overall viewpoint of the system is successful, but also certain features of the complex operation, the user interface is also beautiful enough, the similarity calculation is not compared with other similarity calculations, these questions need to be improved in the future design and improvement. In the future work, we will try to find an efficient similarity calculation method through comparing with other various similarity algorithm. The results will be applied to calculate the similarity of trajectories and the similarity of the network log data more precisely.

7. REFERENCES

[1] Jiawei Hua, Michelline Kamber, Jian Pei. Data mining concepts and techniques: China Machine Press. 2012. 161-175. 288-306.

[2] Josh Jia-Ching Ying, Eric Hsueh-Chan Lu, Wang-Chien Lee, Tz-Chiao Weng, and Vincent S_ Tseng. Mining user similarity from semantic trajectories. In Proc, International Workshop on Location Based Social Networks (GIS-LBSN), pages19-26. ACM, 2010.

[3] Xihui Chen, Jun Pang, and Ran Xue. Constructing and comparing user mobility profiles for location-based services. In Proc. 28th Annual ACM Symposium on Applied Computing (SAC), pages 261–266. ACM, 2013.

[4] Ruipeng LU. New User Similarity Measures Based on Mobility Profiles. ShanDong: ShanDong Univerity, 2014.