CONTENTS

Data mining for providing personalized learning material with interactive whiteboard technology

Zhu Ke
Department of Educational Technology, Henan Normal University, XinXiang, China
Email: zhuke@htu.cn

***Abstract*—Customizing learning material to optimize personal learning has recently become a popular trend in Educational Technology. Because learning material has become an essential resource in the learning active, this study aims to develop a personalized learning system that is based on the data mining technique to provide personalized learning material for optimizing the learner's performance. The purpose of this research is to examine the design and development process of instructional materials that can be used through interactive whiteboards in terms of three critical aspects – software selection, and teaching methods and techniques.**

***Index Terms*—educational technology, interactive white board, data mining, computer assisted learning**

I. INTRODUCTION

Application of information technology can be found in every field of our life at any time. Ignoring these applications is almost impossible. Education is also the filed that is trying to catch up with these developments. New instructional technologies can be seen as an effort to keep up with new information technology [1]. It is known that the effectiveness of education increases as the information technology, which is called the "instructional technology", is used consciously. The information technologies used have an important role in the teaching-learning environment and offer different opportunities for the teacher and student during the teaching and learning process. Avoiding the use of these facilities means ignoring the quality and benefits of technology in education [2]. One of the newest examples of technological application in the field of education is interactive whiteboards. This device that is known as the "IWB", "electronic whiteboard" is a device used in presentations. It is sensitive to touch and used by connecting to digital projector. The interactive whiteboard is a touch-sensitive electronic presentation device. Fully functioning interactive whiteboards usually comprise four components: a computer, a projector, appropriate software and the display panel [3].

PanGu Presentation (PGP for short, the brand of IWB used in this research) is designed by the National Engineering Research Center for E-Learning, facing the education needs, according to subject characteristics, with "student-centered" Presentation. Based on the current multimedia classroom for configuration, especially the growing popularity of electronic devices such as a whiteboard, in the classroom teaching of information and communication technology the current focus on the reform of teaching field, the National Engineering Research Center for E-Learning started the "PanGu plan". PGP emphasizes modernization of educational content, teaching means and methods, and strives to improve the level of teachers' information technology application and update the teaching ideas, and promotes students' active learning and independent study by means of information, and improves our classroom equipment level and leads the direction of the reform of classroom teaching. PGP platform is designed to double innovation of technologies and concepts as the following aspects:

- Widely combined with various types of resources to resolve the status quo of the current teaching platform more than 80% of the PPT as a script;
- Realize the teaching content and the relationship between the reference and an end to the current situation of the present teaching content isolated;
- Create rich and varied teaching models, different from "teachers say, students listen to " the tradition of teaching through classroom teaching;
- Attach importance to the classroom demonstration and the basis of teach, blackboard writing, calculus, reasoning, keep traditional classroom teaching advantages;
- Provide unique new teaching activity design, to make up for the traditional q&a, lack of interactive teaching mainly process;
- The automatic recording and save the teaching process, a complete record of teachers in the traditional classroom blackboard writing, remarks, calculus, reasoning and handwritten mark;
- Provide open source, based on the resources sharing of teaching design platform, create and share with teachers' unique teaching ideology, teaching style of teaching design.

PGP platform provides easy courseware manufacture tools for teaching activity design and teaching resources organization provides production function, the user need only will related resources components into courseware area can be produced personalized teaching activity sequence, really realize the teaching resources organization and to create again.

PGP platform support resources including office, pictures, animations, Web, audio, PDF, 3 D, commonly used in teaching the document format, and also specially developed the test editing tools, topic selection, judgment, and other types including, to adapt to the needs of teaching immediately in the test.

As shown in figure 1 shows, platform through the organization of the classroom activities, fully respect teachers teaching design personalized and creative, and at the same time embodies the regularity of teaching activities. And based on this standard and realize the sharing of resources and the teaching design. Platform provides the diversity of teaching way, can satisfy different teachers' teaching style, different teaching design requirements. According to the teaching process and before the actual need, have a writing pad book based on the teaching, teaching, and activities such as teaching way to choose from. Dual coding theory suggests that information at the same time with visual (image) form and language (text) form two ways is present, can great enhancement memory and recognition. Two teaching is illustrated that show, meaning association the form of knowledge and to activate the learners with the language system language system, convenient for the organization of the classroom teachers, promote the students' cognitive construction. Two display can effectively solve the order of the existing in the teaching of PPT play, unable to control, lack of meaning and association, students didn't do disadvantages notes, etc. At the same time two display teaching model, can provide the many kinds of teaching resources and present platform, and to realize the resource of the layout of the diversification. Platform will provide including window switch, Windows, Windows, side by side, the vertical window cascade. PGP platform derive the traditional teaching "chalk + the blackboard" advantage, combined with today's advanced computer technology in the pioneering innovation, realize the traditional the blackboard and electronic board of the perfect fusion, will the blackboard writing, calculus, reasoning, tagging etc. to traditional classroom activities into the modern classroom, to the "everywhere, ready to write" write and class writing according to the effect of trace are kept. PANGU teaching platform configuration interactive terminal module, provide and network two interactive way, for classroom build vivid and active classroom atmosphere, to realize "teach" give priority to the teaching mode to "learn" based teaching mode change; The interactive data acquired platform can be automatically analysis, realize the interaction results present vivid, can direct convenient from the classroom test students' learning effects, easy to adjust the teaching progress. Cooperate with PANGU teaching platform, relying on central China normal university, no.1 middle, and other famous education institutions design and development of high quality, professional standard teaching database, for the courseware manufacture, the lectures to provide abundant resources and convenient tools.

Teaching resources includes public resource and subject resource database two parts. Public resource database provides graphs, routine identification, 3 d model and other public resources; Subject to the discipline of resource to provide teaching resources, meet a taught, should, quiz, teaching mode and theory teaching, experiment teaching, practice teaching and the needs of the class.

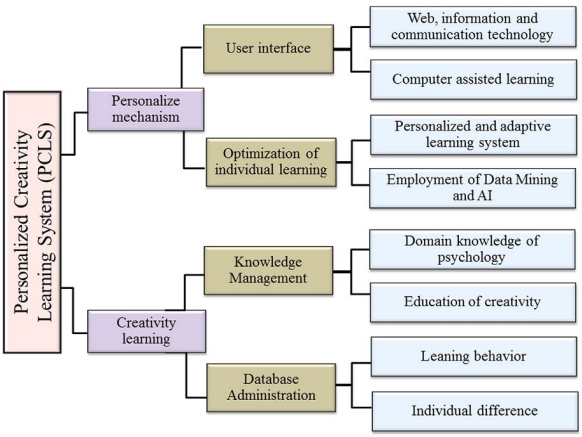


Figure 1. The development framework of the research system

II. METHODS

A. Participants

In this study, the IWB student survey was distributed to 110 middle school students (Grade 6 and 7). In total, 95students properly responded to the survey, resulting in a response rate of 86.4%, while only 90 responses were valid for the analyses. In order to analyze students' perceptions towards the IWB use, several questions regarding students' personal information and their IWB experiences were also included in the questionnaires.

B. Integration and checking data

Databases are highly susceptible to noisy, missing, and inconsistent data due to their typically huge size and their likely origin from multiple, heterogeneous sources [4]. Low-quality data will lead to low-quality mining results. Data integration merges data from multiple sources into a coherent data store such a data warehouse. Careful integration can help reduce and void redundancies and inconsistencies in the resulting set. This can help improve the accuracy and speed of the subsequent data mining process.

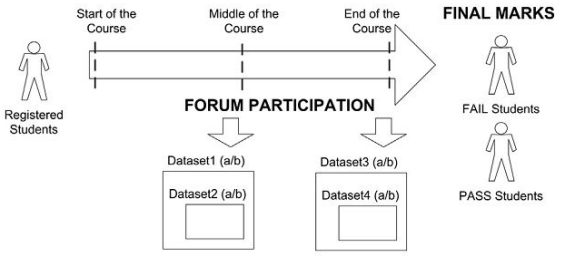


Figure 2. Data gathering framework

C. Data cleaning

Data cleaning routines attempt to fill in missing values, smooth out noise while identifying outliers, and correct inconsistencies in the data. We use a global constant to fill in the missing value. Use a measure of central tendency for the attribute.

D. Data Transformation and Data Discretization

In this preprocessing step, the data are transformed or consolidated so that the resulting mining process may be more efficient, and the patterns found may be easier to understand. Data transformation routines convert the data into appropriate forms for mining. For example, in normalization, attribute are scaled so as to fall within a small range such as 0.0 to 1.0.Data discretization transforms numeric data by mapping values to interval or concept labels. For nominal data, concept hierarchies may be generated based on schema definitions as well as the number of distinct values per attribute.

E. Apply Data Mining Result

The range of possible solutions to the above problems is vast, extending far beyond our approach to instructional technology [5]. While our approach to instructional technology alone cannot resolve all of the problems, it can have a significant impact on all of them. For this stage, we used SPSS（Statistical Product and Service Solutions）that provide data mining algorithms for clustering, classification, and association. Cluster analysis itself is not one specific algorithm, but the general task to be solved. Various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them can achieve it.

III. Helpful Hints

In this section, for each algorithm used in the study, the test characteristic and results obtained are shown.

In each test, the number of clusters was calibrated to generate the greater amount of clusters having mutually exclusive attributes. In Table 1 some clusters obtained are presented.

TABLE I. SOME OF THE BEST RULES OBTAINED WITH THE CLUSTERING ALGORITHMS

Cluster 1	Cluster 2	Cluster 3
Interface describes competencies: well Component structure: sequenced Knowledge transfer: by web quest Designer's profile: Computing: advanced	Interface describes competencies: very good Component structure: compact Knowledge transfer: by demonstrations Designer's profile: Computing: average	Interface describes competencies: scantly Component structure: isolated Knowledge transfer: by problem solving Designer's profile: Computing: average

Various tests were verified with ID3 algorithms with the already mentioned datasets. We obtained a set of IF-THEN-ELSE rules from the algorithms. After an analysis, we eliminated those rules that were with irrelevant information .Table 2 shows some of the best rules obtained.

TABLE II.SOME OF THE BEST RULES OBTAINED WITH THE CLASSIFICATION ALGORITHMS

Rules—Generated	Rules—Interpretation
format =doc; knowledge transfer =demonstrations; media balance = regular => high	The LO has a high semantic density if it does fulfill the next requirements: has a doc format contains demonstrations, and the media balance is regular.
format =swf; media balance = basic; structure =atomic =>low	The LO has a high semantic density if it does fulfill the next requirements: has an swf format, the media balance is basic, and the structure is atomic.
format =rmvb; educational resource type = reading => high	The LO has a high semantic density if it does fulfill the next requirements: has a rmvb format and is used as a reading in class.

The Apriori algorithm is a seminal algorithm for mining frequent item sets for Boolean association rules It explores the level-wise mining Apriori property that all nonempty subsets of a frequent item set must also be frequent At the kth iteration, it forms frequent k-item set candidates based on the frequent (k-1), and scans the database once to find the complete set of frequent k-item sets.

TABLE III. SOME OF THE BEST RULES OBTAINED WITH THE ASSOCIATION ALGORITHMS

Rules—Generated	Rules—Interpretation
version = final; environment = classroom => aggregation level = 1	If the LO consists of the final version and is used in perennial learning, then aggregation level is basic
structure = atomic; version =final; environment=classroom=>aggregation level = 1	If the LO consists of atomic structure, final version, and is used in perennial learning, then aggregation level is basic
aggregation level= 1; version= final; easy navigation = security => environment = classroom	If the LO is basic and security navigation then it is used in perennial learning

IV. Recommendations for Future Research

In this paper, we presented an adapted methodology for the application of data mining techniques to Los, trying to discover relevant rule in its design and usage characteristics. This can help teachers to assess learning resource precisely utilizing learning objects usability rules in a web-based learning environment. Additionally, teachers can devote themselves to teaching and designing courseware since they save a lot of time in evaluating learning objects. More significantly, teachers could understand the factors influencing learning performance in a web-based learning environment according to the

obtained interpretable learning objects usability rules. Students will have the benefit of access to the highest-quality learning resources available, making a significant impact on the quality of their learning experience and their learning outcomes.

Further collect data from other channel are required in future work. For example, it would be interesting to consider the situational of learning as an element to recommend teachers and students suitable learning objects with real environment. Also, it would be consider the location, equipment, even weather and something else related with learning.

ACKNOWLEDGMENT

This work is supported by Science and technology research projects of Educational Commission of Henan Province (No. 14A880019).

REFERENCES

[1] Aldrich, F., Rogers, Y., & Scaife, M. "Getting to grips with ''interactivity'': helping teachers assess the educational value of CD-ROMs". British Journal of Educational Technology, vol.29, pp.321–332, 1998.

[2] Draper, S. W., & Brown, M. I. "Increasing interactivity in lectures using an electronic voting system". Journal of Computer Assisted Learning, vol.20, pp.81–94, 2004.

[3] Murcia, K., Multi-modal Representations in Primary Science. What's offered by Interactive Whiteboard Technology? Teaching science, vol.56 , pp.23-29, 2010.

[4] Zucker. A.A., Transforming Schools with Technology. Cambridge: Harvard Educational Press, 2008.

[5] Hwang, W., Chen, N.& Hsu, R., Development and evaluation of multimedia whiteboard system for improving mathematical problem solving. Computers & Education, vol.46, pp.105–121, 2006.

Zhu ke, born in 1982, he received his PHD in computer science from the Huazhong Normal University.

His research interests include learning analysis, Computer education.

Motivational Factors that Influence the Acceptance of Online Course Using TAM

Zhang Jin
Department of Educational Technology, Henan Normal University, Xinxiang, China
Email: zj@htu.cn

***Abstract*—Technology acceptance is an interesting topic for researches that belong to different areas of scientific scopes, the phenomenon of technology adoption or rejection was studied by many fields' researchers. To predict user's adoption of technology, many theories were developed for this purpose. Existing literature in the field of e-learning technology acceptance reflects a significant number of independent studies that primarily investigate the causal relationships proposed by technology acceptance theory, such as the technology acceptance model (TAM). Online Course and other virtual teaching platforms have bolstered the ability and motivation of universities to support distance learning or traditional classroom learning. The aim of our study is to improve understanding of the motivational factors behind student satisfaction, or dissatisfaction, with the Online Course. Our study extends the Technology Acceptance Model (TAM) to include technical support and perceived self-efficacy, with the expectation that they influence usage of Online Course. We surveyed 102 students of Henan Normal University. The data showed that training has a direct effect on perceived ease of use and perceived usefulness. Perceived ease of use and attitude was also directly influenced by perceived compatibility with student tasks. The results reveal the importance of training and perceived compatibility with student tasks.**

***Index Terms*—Online Course, TAM, Perceived Usefulness**

I. Introduction

Information technology (IT) adoption and diffusion has been studied in great detail by researchers in the information system area. IT acceptance in education remains a central concern of information systems research and practice [1]. Although IT is playing an increasingly important role in contemporary education, resistance to IT remains significant in the education sector. Understanding the conditions under which IT are or are not accepted and used by students continues to be an important issue.

Online Course is an inclusive term that describes educational technology that electronically or technologically supports learning and teaching [2]. Depending on whether a particular aspect, component or delivery method is given emphasis, a wide array of similar or overlapping terms has been used. As such, Online Course encompasses multimedia learning, technology-enhanced learning (TEL), computer-based training (CBT), computer-assisted instruction (CAI), internet-based training (IBT), web-based training (WBT), online education, virtual education, virtual learning environments (VLE) which are also called learning platforms, m-learning, digital educational collaboration, distributed learning, computer-mediated communication, cyber-learning, and multi-modal instruction. Students who have access to an E-learning system can get course materials in different formats (text, image, sound, etc.), and can interact with their colleagues and lecturers individually and simultaneously via message boards, forums, chat rooms, videoconferencing, etc. This can be done wherever there is an Internet connection. Students can learn with their pace, and shape the learning process to suit their own needs [3].

Although many colleges currently use the Web for teaching and learning, little research has been done to identify the factors that influence students' acceptance of Online Course. The TAM model was first introduced by Davis et al. to explain the acceptance and usage of information technologies. It was based on the ‘‘Theory of Reasoned Action" developed in Social Psychology by Hamilton [4]. The Theory of Reasoned Action (TRA) is a general system designed to explain almost every type of human behavior, and part of the importance of individual beliefs, in order to predict human conduct. TAM models center exclusively on the analysis of information technology and, as opposed to TRA models, pre-establish those factors that condition user attitude towards innovation, behavioral intention and intensity of system usage [5]. The two key factors in determining intention, which predict the development of an innovation and are present in all studies of TAM model development, are: perceived usefulness (PU) and perceived ease of use.

The purpose of the present study is to:

1. Understand the users' expectations or perception of Online Course;
2. Develop and test the constructs of the factors of related studies;
3. Investigate the relationship between perceived usefulness, perceived ease of use, and training, perceived usefulness for professors, perceived compatibility with student tasks and actual usage behavior;
4. Explore how the role of risk perception on satisfaction and intentions to change improve the teaching–learning process.

II. METHODOLOGY

A. Technology acceptance model

Fig. 1 shows the research model. User-perceived web quality and habit are modeled as exogenous variables, while perceived usefulness and trust are modeled as mediators between the exogenous variables and the learner's behavioral intention to continue using the Online Course. The model has eight hypotheses [6]. Each is listed with arguments and justification.

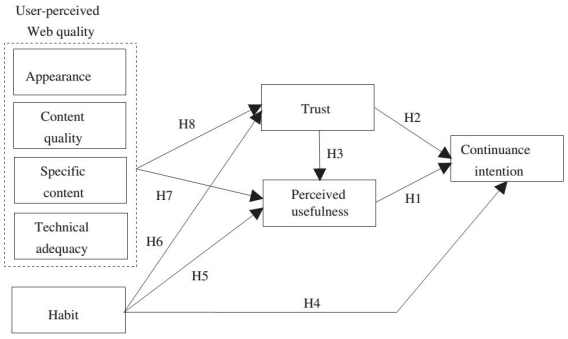


Figure 1. Proposed model for the acceptance of Online Course

B. Hypotheses

Perceived ease of use is defined as the degree to which an individual believes that using a particular system would be free of physical and mental effort. A considerable amount of prior studies supported the significant effect of perceived ease of use on behavioral intention, either directly or indirectly through perceived usefulness [7]. Although TAM posits that perceived ease of use is another determinant, Karahanna, Straub, and Chervany found that perceived ease of use has no significant effect on behavioral intentions to use for experienced IT users. Perceived ease of use usually is shown as an important predictor for potential adopters of systems, because it reflects users' computer self-efficacy when they adopt a new system [8]. Therefore, the first set of hypotheses for this study is stated as follows:

H1. Perceived ease of use has a significant effect on the perceived usefulness.

H2. Perceived ease of use has a significant effect on the intention to use Online Course.

H3. Perceived usefulness has a significant effect on the intention to use Online Course.

Recent studies acknowledged the importance of individual differences such as motivations and intentions of usage in determining users' acceptance of new technologies. Perceived compatibility with student tasks is a significant factor that explains the adoption and use of communication technologies [9]. Past studies on the role of motivation in Internet use confirmed that perceived compatibility with student tasks has a positive impact on new technology acceptance and usage (e.g., Lin, 1998; Stafford & Stern, 2002). We can state these two hypotheses:

H4. Perceived compatibility with student tasks has a significant effect on perceived usefulness of Online Course.

H5. Perceived compatibility with student tasks has a significant effect on perceived ease of use of Online Course.

In the above discussion of perceived usefulness by users, the possibility of altering their intention of use was mentioned because users consider the technology use to be necessary.

Perceived usefulness is operationalized as the degree to which an individual believes that using a particular system would enhance his or her job performance. Users' intention to use an information technology is expected to be greatly affected by their perceived usefulness of the system. There is also extensive empirical evidence that supports the significant effect of perceived usefulness on behavioral intention. Based on prior research, this study hypothesized the following.

H6. Perceived usefulness for professors has a significant effect on perceived usefulness of Online Course.

H7. Perceived usefulness for professors has a significant effect on the intention to use Online Course.

H8. Training has a significant effect on perceived usefulness of Online Course.

H9. Training has a significant effect on perceived ease of use of Online Course.

The proposed model has six constructs and nine hypotheses have been generated from the relations of these six constructs.

III. DATA ANALYSIS AND RESULTS

In order to empirically test the proposed model, we give a survey. Through the investigation of scale reliability and validity, we first fine-tune to scale, eliminating some not reliable item, according to the feedback; we also adjusted the meaning expression, which is not clear item.

In confirmation of the questionnaire can be clear understand by general users, we finally got the formal questionnaire. Formal questionnaire consists of two parts:

(1) The basic personal situation, including user's gender, age, formal Education Information.

(2) Six latent variables and 18 observation variables constitute the questionnaire.

The questionnaire has several items related to each of the constructs included in the model. The survey items were measured using a five-point Likert scale. All items ranged from 1 (strongly disagree) to 5 (strongly agree). Theoretical constructs were operationalized using validated items from prior re-search.

The reliability and validity of the research instrument was first assessed using item reliability, internal consistency, and discriminant validity. Individual item loadings were used to evaluate individual item reliability. According to Chin (1998), individual items with loadings greater than 0.505 are considered acceptable, implying the item explained about 50% of the variance in a specific measure and ensured that the items in the measurement model measured the same construct. Results of item reliability indicated that all items exceeded the threshold, indicating that the survey

instrument was sufficient for measuring each construct individually. Internal consistency for each construct was evaluated using Cronbach's alpha.

Discriminant validity, using the average variance extracted (AVE), was tested [10]. Discriminant validity is the lack of a relationship among measures, which theoretically should not be related. The AVE measures the variance captured by the indicators relative to the measurement error, which should be greater than 0.5 in order to justify the use of a construct. The latent variables should not exceed the AVE to justify the discriminant validity. This is achieved by calculating the square roots of the average variance extracted (AVE) values, which measure the average variance shared between a construct and its measures, and by calculating the correlations between different constructs. A matrix can then be constructed where the square root of AVE is in the diagonal, and the correlations between the constructs are in the off diagonal.

Table I. Items loading

Construct	Indicator	Loading
Perceived usefulness for professors (PUP)	PUP1. Integrated online survey and testing tools in Online Course can help teachers to analysis students	0.7304
	PUP2. Online Course can help teachers use and organizational learning resources efficiently	0.6045
	PUP3. Online Course can help teachers to develop teaching activities flexibly	0.7122
Perceived compatibility with student tasks (PC)	PC1. Online Course in accordance with student's study habits	0.6822
	PC2. Online Course can support the user mode of learning	0.7031
	PC3. Online Course has no difference of the operation with the other learning software	0.6888
	PC4. Online Course can aid the learning better	0.7011
Training (T)	T1.the training of how to use Online Course is comprehensively	0.7231
	T2. Through the training can improve the understanding of the Online Course	0.6903
	T3. Through the training can raise the proficiency of operating	0.7376
	T4. Training teacher can answer questions skillfully	0.7655
Perceived usefulness (PU)	PU1. Use Online Course can get more opportunity to exchange ideas	0.6432
	PU2. Use Online Course make higher scores	0.6899
	PU3. Sakai can solve the problem of learning which deal with difficulty before	0.7210
	PU4. Online Course can improve the learning efficiency	0.7800
Perceived ease of use (PEU)	PEU1. Online Course can be operated correctly	0.6122
	PEU2. Use Online Course can interchange with classmates conveniently	0.7033
	PEU3. Use Online Course without help	0.6384
	PEU4. Online Course can meet need outside course	0.5988
	PEU5. Willing to explore the function of Online Course	0.7601

Ⅳ. Discussion

This study investigated the factors that influence behavioral intention to use Online Course in different user through the TAM framework developed by Davis (1986). The findings generally supported the hypotheses derived from the model as well as earlier empirical studies. As mentioned earlier, since the TAM is a general and parsimonious model, the key variables (i.e., perceived ease of use and perceived usefulness) may function differently depending on the external variables of specific research settings. The current study employed three external variables representing individual differences and system characteristics. As can be seen from the survey, experience in computer use had a direct effect on perceived ease of use, while it indirectly affected both perceived usefulness and behavioral intention to use. However, those effects were quite small.

The finding that perceived compatibility with student tasks is a significant factor. Considering that most of the participants in the current study had been computer and Internet users for more than 1 year, the significant effect of on ease of use and usefulness suggests that more sophisticated instruction on using the digital system in learning is necessary.

The findings also indicate that while training quality is an important driver of perceived usefulness for professors behavioral intentions, its indirect effect on user satisfaction is equally important, if not more so, in promoting the efficiency of Online Course.

ACKNOWLEDGMENT

This work is supported by soft-scientific project of Henan Province (No. 132400410982) and Science and technology research projects of Educational Commission of Henan Province (No. 14A880019).

REFERENCES

[1] Ajzen, I. Perceived Behavioral Control, Self-Efficacy, Locus of Control, and the Theory of Planned Behavior. Journal of Applied Social Psychology, 32, 665–683,2002
[2] Allen, I. E., & Seaman, J. Learning on demand: Online education in the United States, 2009. http://sloanconsortium.org/publications/survey/pdf/learningondemand.pdf Retrieved from the Sloan Consortium website.2011.
[3] Lin, C. S., Wu, S., & Tsai, R. J. Integrating perceived playfulness into expectation-confirmation model for web portal context. Information & Management, vol. 42, pp.683–693, 2005.
[4] Hamilton, M. Adding contextual specificity to the technology acceptance model. Computers in Human Behavior, vol.22, pp.427–447, 2006.
[5] Sugimoto, C., Pratt, J., & Hauser, K. Using field co-citation analysis to assess reciprocal and shared impact of LIS/MIS fields. Journal of the American Society for Information Science, vol.59, pp.1441–1453, 2008.
[6] Van Raaij, E. M., & Schepers, J. J. L. The acceptance and use of a virtual learning environment in China. Computers and Education, vol.50, pp.838–852, 2008.
[7] Benbasat, I., & Barki, H. Quo vadis, TAM? Journal of the Association of Information Systems, vol.8, pp.211–218, 2007.
[8] Yiu, C. S., Grant, K., & Edgar, D. A. Adoption of the Internet for retail bank customers in Hong Kong – Technology adoption model. International Journal of Information Management, pp.336–351, 2007.
[9] Van Raaij, E. M., & Schepers, J. J. L. The acceptance and use of a virtual learning environment in China. Computers and Education, vol.50, pp.838–852, 2008.
[10] Sussman, S. W., & Siegal, W. S. Informational influence in organizations: An integrated approach to knowledge adoption. Information Systems Research, 2003.

Zhang Jin, born in 1983, she received her Master degree in educational technology from the Yunnan Normal University. His research interests include Multicriteria Decision Analysis, Intelligent decision support systems.

Time based Routing Protocol in Opportunistic Networks

Li Liu
College of Information Science and Electrical Engineering ,ShanDong Jiaotong University, China
liuli_free@126.com

***Abstract*—Opportunistic Networks are used in intermittent connected networks by use of store-carry and forward fashion. Social properties are usually used as an efficient method to improve the routing performance. The people meet different kinds of other people in different time. In this paper, we present TBR routing protocol and utilize the irregularity of mobility about meet people set and time to predict the future meet opportunities. We establish experiment based on ONE and simulations shown that the efficiency of TBR outperforms Epidemic and PROPHET.**

***Index Terms*—Opportunistic Networks; Routing Protocol.**

I. Introduction

More and more mobile devices such as iPhone, PDA etc. are implemented by people. People can communicate and transmit messages at any place and any time. The increasing applications and services based on mobile devices induce large number of overhead to 3G. Opportunistic networks [1] provide a flexible manner to support intermittent connection transmission by use of short range wireless transmission technologies such as Bluetooth, WiFi. It utilizes opportunistic contact opportunity to deliver messages using store-carry – forward fashion.

A node (device) wants to deliver a message to other node, however, an end-to-end path between two nodes is difficult to find. In this intermittent connect scenario, source node will select appropriate intermediate nodes as relay nodes to forward the message to destination node. Usually, the device which has more probability to meet the destination device is prior selected. Most recent researches focus on the prediction algorithm to future meet opportunities.

On the same time, researchers found most of mobile devices are carried by human, which makes the networks are characteristic with the social relationships. Many recent researches are inspired by the social reality such as [2-4]. Human with social relationship such as family, classmate, colleague, always form a community. People in community generally meet frequently than out of community. However, compared to large number of mobile devices, community members are so little that lots of opportunities are wasted. Several researchers exploit the strangers to help forward between different communities and obtain better performance such as [5-8].

On the other hand, Human's movement is usually regular through time. For instance, as Fig 1 shown, Mark gets bus to work at 8:00 am. He works at office between 9:00 am and 5:00 pm. Then he spends an hour to take sports and gets bus to go home at 7:00 pm. At the end, he arrives home at 7:30 pm and stay home till next day. From Monday to Friday, Mark's time schedules are usually similar. And at special time period, he may meet several people frequently, such as colleagues and his friends. In particularly, some strangers who usually get same bus with Mark are period meet. These information can be used to predict the future contact opportunities.

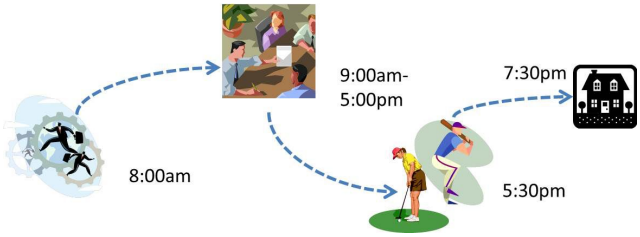


Figure 1. One day of Mark

In this paper, we present a time based routing protocol (P3), which is based on the irregular meet people set according to different time period in human's mobility to predict the future meet opportunities. We establish simulation and evaluate the performance of time based routing protocol (TBR) through compare to Epidemic routing protocol and PROPHET routing protocol. The results demonstrate the higher efficiency of time based routing protocol.

II. Related Work

In Opportunistic Networks, routing protocol is a challenge problem due to high mobility and scare connection. Recent few years, lots of researches about social related strategies have been proposed to solve the routing problem. Community is the basic concept which is popularly used in social based routing protocols. People who meet frequently form a community. Physical proximity is usually utilized as community detection principles. Physical proximity is related to nodes' encounter history. By use of encounter history record, contact graph can be constructed which is mostly used to community detection. Bubble Rap [2] defines k-clique community from contact graph and hierarchical ranking

tree to guide routing. Nodes which can reach each other through a series of adjacent k-cliques form a community. In addition, MOPS [3] and LocalCom [4] construct communities based on neighboring relationship from nodes' encounter histories.

In addition, lots of researches take human's behavior regularity into consideration. Habit [5] constructs regularity graph and interest graph for routing protocol by leveraging information of nodes' regularity of movement and their social networks. HiBOp [6] automatically learns and represents context information, the users' behavior and their social relations, and exploits this knowledge to drive the forwarding process. Zhou et al. [7] analyze spatial and temporal characteristic of people and construct history temporal community at different time period to predict future contact probability. Nguyen and Giordano present CiPRO in [8] which uses BackPropagation Neural Network (BNN) model to predict the context of nodes, further decides when and where to start the routing process in order to maximize the transmission delay and minimize network overload.

Our work in this paper aims to utilize people's regular mobility properties in order to predict the future contact opportunities stably. TBR records frequently meet people set according to different time periods and use them to compute the future meet probability in the future.

III. System Model

We assume there are N mobile devices and they transmit with each other through Bluetooth. And the mobility of mobile devices is regular through time. We can imagine a people's work days. People meet different people set in different time periods. For example, when people are in work, the people meet set consists of their colleagues. When people are out of work, the people set usually some strangers or some friends.

In this scenario, mobile devices record the meet people set at different time period. We divide the day time into 12 parts from 7:00am to 7:00pm in one hour interval. The night time is discarded because the mobility is rare in night.

IV. Implement of TBA

In TBR, mobile devices record and maintain frequently meet people set according to different time periods.

We assume that transmission tasks generally happen between friends. And for friends, they are familiar of their mobility regularity with each other and they are familiar to meet people set with each other. When a mobile device wants to transmit a message to its friend, the destination id and its people set list are bind into the message. For the message in one mobile device, it will compute the meet probability of the destination and store it in message list.

Thus, the TBR routing protocol are presented as follows, which consists of 4 steps.

1. The mobile devices record and maintain the people set they meet frequently in fixed intervals.
2. When two mobile devices are in their transmission range. Two devices (called A and B for simply description) exchange the message list with each other.
3. For each message in message list, TBR first find its destination, noted as D. Then TBR computes the meet probability P(A,D) between A and the destination. Then A compares P(A,D) to P(B,D) in message list. If P(A,D>P(B,D), the message will be delivered by A. Otherwise, the message is stayed in B. B has the similar process with A.
4. Start the transmission process.

The value of meet probability P(A,D) between A and D is the sum of number of similar people set at future corresponding time, according to following (1).

$$P(A,D)=\sum_{i=1}^{n} p_i \tag{1}$$

Where Pi represent the number of similar people in A and B's meet people set in i time period from current time period. When i is 1, it represent current time period. N represent the number of next future time period.

V. Simulation and Evaluation

In the simulation, we use real data set collected by an opportunistic mobile social application MobiClique during SIGCOMM2009 conference [9]. Around 100 smart phones were distributed to a set of volunteers during the first two days of the conference. The experiment is carried out through the Opportunistic Network Environment (ONE) Simulator [10]. We handle proximity information of devices to satisfy external movement requirement of ONE. The messages with 10 hours TTL are generated randomly.

We compare the effectiveness of TBR with two 'non-oblivious' routing protocols: Epidemic and PROPHET. Epidemic adopts a simple flooding method which copies the messages to every encounter that has not received a copy. PROPHET is prediction-based routing which predicts and selects the forwarders by the use of history encounter records. For PROPHET, we use the default parameters provided by ONE.

In Fig 2, we show comparison of all algorithms in terms of delivery ratio, overhead ratio and average latency under different simulation times. The simulation times are designed at 6 hour interval. As shown, the performance of TBR outperforms Epidemic and PROPHET obviously. In 16 hours, P3 forwards 27.31% messages with overhead ratio of 14.3 and average latency of 3063. While the delivery ratios of Epidemic and PROPHET are 13.03\% and 10.93% respectively with overhead ratio of 853 and 923respectively. The average latency of Epidemic and PROPHET are 2841 and 2207.

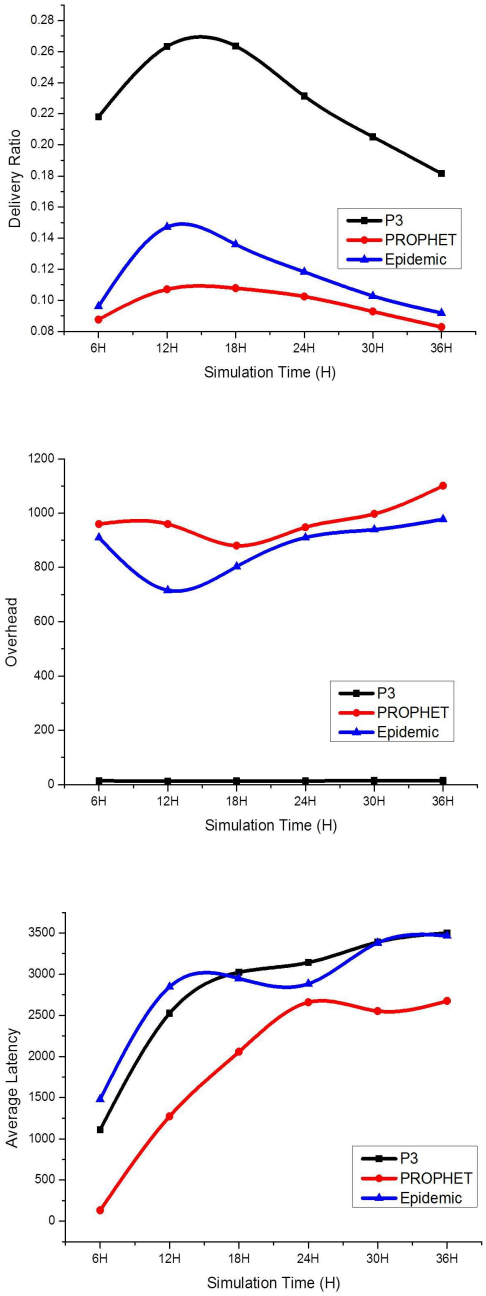


Figure 2. Simulation and Results

VI. CONCLUSIONS

In this paper, we present a routing protocol in Opportunistic Networks, named TBR. In TBR, each mobile device records and maintains frequently meet people set according to different time periods. And based this information, TBR selects the higher meet probability to the destination device as forwarder in order to improve the efficiency. The simulation shows that TBR obtains the higher forwarding performance than Epidemic and Prophet.

ACKNOWLEDGEMENTS

This work was financially supported by the Shandong Jiaotong University Science Research Foundation (Z201212).

REFERENCES

[1] M. Conti, S. Giordano, M. May, and A. Passarella, “From opportunistic networks to pportunistic computing,” Communications Magazine, vol. 48, no. 9, pp. 126–139, 2010.

[2] H. Pan, J. Crowcroft, and E. Yoneki, “Bubble rap: Social-based forwarding in delay-tolerant networks,” Mobile Computing, vol. 10, no. 11, pp.1576–1589, 2011.

[3] F. Li and J. Wu, “Mops: Providing content-based service in disruptiontolerant networks,” Distributed Computing Systems, 2009. 29th IEEE International Conference on, pp. 526–533, June 2009.

[4] L. Feng and W. Jie, “Localcom: A community-based epidemic forwarding scheme in disruption-tolerant networks,” 6th Annual IEEE Communications Society Conference on, pp. 1–9, 2009.

[5] A. Mashhadi, S. Ben Mokhtar, and L. Capra, “Habit: Leveraging human mobility and social network for efficient content dissemination in delay tolerant networks,” IEEE International Symposium. WoWMoM 2009. pp. 1–6, 2009.

[6] B. Chiara, C. Marco, and P. Andrea, “Exploiting users social relations to forward data in opportunistic networks: The hibop solution,” Pervasive and Mobile Computing, vol. 4, no. 5, pp. 633–657, Oct 2008.

[7] T. ZHOU, H. XU, and Y. CHEN, “Spatial and temporal communities based routing algorithm in human social network,” Journal of Computational Information Systems, vol. 8, no. 18, pp. 7823–7830, Oct 2012.

[8] N. Hoang, Anh and G. Silvia, “Context information prediction for social based routing in opportunistic networks,” Ad Hoc Networks, vol. 10, no. 8, pp. 1557–1569, 2012.

[9] A.-K. Pietil¨ainen, E. Oliver, J. LeBrun, G. Varghese, and C. Diot, “Mobiclique: middleware for mobile social networking,” in Proceedings of the 2nd ACM Workshop on Online Social Networks (WOSN’09), Barcelona, Spain, August 2009, pp. 49–54.

[10] A. Ker¨anen, T. K¨arkk¨ainen, and J. Ott, “Simulating mobility and dtns with the one,” Journal of Communications, vol. 5, no. 2, pp. 92–105, 2010.

The Establishment of Female Attractiveness Rating Predicted Model

Zhai Lili
Institute of Textile, Tianjin Polytechnic University, Tianjin, China

***Abstract*—WHR, BMI and VHI have therefore been proposed to be the most important determinant of female body attractiveness by different researchers. The above key body indices, although simple and easy to use, may have overlooked the complexity of body geometry or phenotypes, as the various body proportions are interrelated. It is therefore necessary to investigate the various body proportions as a whole to see whether there is an optimum body phenotype for body attractiveness. Based on the analysis of 39 mainly female body ratios, the present study established a predicted model for famale attractiveness using Principle Component Analysis, Canonical Analysis and Regression Analysis and suggested the optimum body ratios.**

***Index Terms*—Female attractiveness, body proportion, BMI, WHR, VHI**

I. INTRODUCTION

In women, three potential cues to physical attractiveness have been proposed, viz. body mass index (BMI), volume height index (VHI) and waist-to-hip ratio (WHR). A lot of research work has focused on the waist-hip ratio (WHR), the ratio of the circumference (or width) of the waist to the circumference (or width) of the hips. A low WHR is believed to correspond to the optimal fat distribution for high fertility [1]. Furnham [2] supported and extended Singh's research and predicted that the effect of breast size on attractiveness judgments depended on overall body fat and WHR. However, two components of the WHR may have different effects on female attractiveness. The optimal WHR for attractiveness was 0.7[3]. The waist conveying information about fecundity and health status was more important than hip size in assessing the female attractiveness [4].

Tovée and his co-workers proposed that BMI (body mass index, weight scaled for height) is a far more important factor than WHR in determining the attractiveness of a female body [5]. They also indicated that there were no significant differences in the rating of attractiveness by male and female raters [6-7].

Fan et al [8-9] showed that VHI, viz. the body volume divided by the square of the height, is the most important and direct visual determinant of female physical attractiveness by analyzing the attractiveness of three-dimensional (3D) scanned images. They found that about 90% of the variance of the attractiveness ratings could be explained by VHI alone compared with BMI which only explained less than 80% of the variance.

The above mentioned key body indices, although simple and easy to use, may have overlooked the complexity of body geometry or phenotypes, as the various body proportions are interrelated. It is therefore necessary to investigate the various body proportions together to see whether there is an optimum body phenotype and evolution of body shape in a particular direction would enhance female body attractiveness.

II. METHODS

A.Participants

35 male and 45 female students with a mean age of 22.86 years (SD=2.33), who were enrolled at Tianjin Polytechnic University, were invited to view and rate the attractiveness of the body images in a scale from 1 (least attractive) to 9 (most attractive). All participants took part in this study on a voluntary basis.

B. Stimuli

Ninety six females from Tianjin Polytechnic University whose average age was 22.1 with SD=±1.04 were scanned using a Lectra 3D body scanner. These models wore tight-fitting body underwear that did not alter the shape of the body. 39 key body proportions were obtained from the body scan data. Movie clips of 3D scanned female images (an example is shown in Figure 1) were made in the same way as described in Fan et al's previous paper [16].

C. Procedure

We presented stimuli and recorded responses using MediaLab psychology research software running on a personal computer with a 1280x1024 resolution screen. Raters were asked to rate each model as quickly as possible on a nine point Likert scale that asks raters to specify the extent to which they find the model attractive. Raters took 6.29 s, on average, to rate each model. Once a rater entered a rating, the video stopped and the next model was presented.

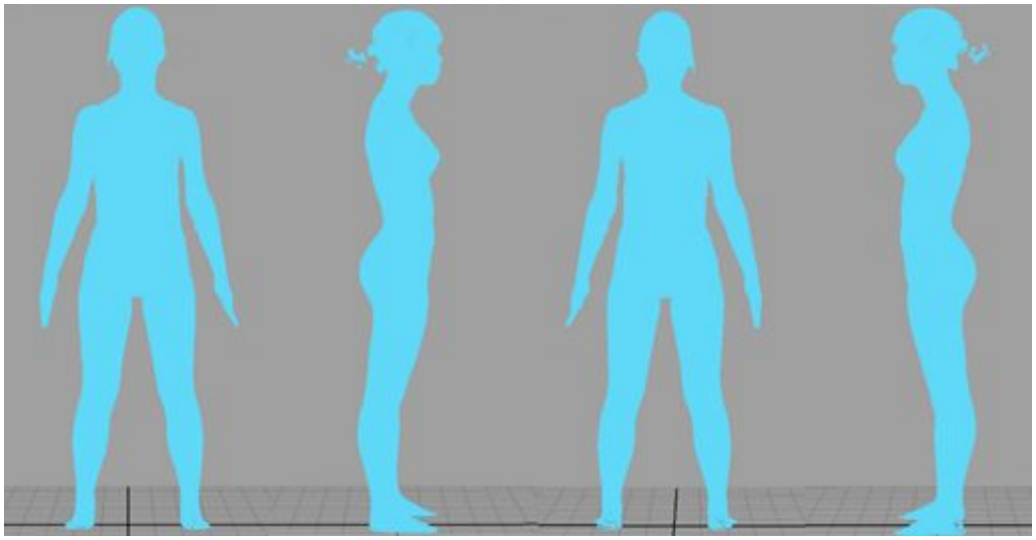


Figure1. An example of a movie clip in the stage of starting, 3s, 6s and 9s, respectively

III RESULTS AND ANALYSIS

A. Principle Component Analysis (PCA) and Quadratic Model

Since the 39 body proportions may be correlated, Principle Component Analysis (PCA) was applied to find independent principle components analysis.

Before applying PCA, the values of the 39 body proportions were first normalized. For each of the body proportions, the normalized body proportion is calculated using the following formula:

$$pn = \frac{(p - meanp)}{stdp} \quad (1)$$

Where, pn is the normalized body proportion, p is the original body proportion, meanp is the mean value of the body proportion, stdp is the standard deviation for the body proportion. After normalization, each body proportion has the same mean value of zero and standard deviation of unity.

Principle Component Analysis (PCA) was carried out using SPSS. Six independent components with eigenvalue greater than 2.0 were extracted (See Table I). The top six components, which cumulatively explained 77.8% of the total variance, were chosen for further analysis. We also use the "Sorted by Size" function in SPSS to identify the significant relationship between the principle components and the body proportions.

TABLEI. Total Variance Explained from PCA

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	12.316	31.579	31.579
2	5.448	13.970	55.178
3	3.755	9.629	63.686
4	3.318	8.508	77.773
5	2.811	7.207	82.951
6	2.683	6.881	89.921
7	2.020	5.179	92.272
8	1.639	4.204	94.041

A quadratic regression model was established using the 'rstool' function in MATLAB to predict the attractiveness ratings from the six independent components. The quadratic regression model includes constant term, linear terms, quadratic terms including the square terms and cross terms. The quadratic regression model can be illustrated as follow:

$$y = \alpha_m + \sum_{i=1}^{6} \beta_i C_i + \sum_{i=1,j=1}^{6} \gamma_{ij} C_i^2 + \sum_{i<j}^{6} \sum_{j=2}^{6} \gamma_{ij} C_i C_j \quad (2)$$

Where, y is attractiveness rating, α_m , β_i and γ_{ij} are coefficients in the constant term, linear terms, square terms and cross terms, respectively.Table II lists the coefficients of all quadratic regression model.

TABLEII The linear gradients (β) and the quadratic gradients (γ) of the quadratic model. Randomization test:* p<0.05, ** p<0.01, *** p<0.001.

	β	γ					
		C1	C2	C3	C4	C5	C6
C1	0.19	-0.00					
C2	-0.03	0.02	0.002				
C3	-0.06	0.008	-0.026	0.011			
C4	0.13	-0.022	0.037	0.036	-0.020		
C5	-0.11	0.022	0.026	-0.04	0.05	-0.03	
C6	0.001	-0.01	-0.01	-0.02	0.013	0.038	-0.05

Its r^2 is 0.746 with the p<0.05, which suggests that the regression model is significant. However, no individual independent principal component is significantly related to dependent variable (attractiveness ratings), which may be a result of underestimation due to the fact that the principal components are not at the major axes of the multivariate response surface.

B. Canonical analysis

Canonical analysis was carried out to find the major axes of the response surface by rotating the axes and transforming the quadratic gradients into canonical form (viz. getting rid of the cross terms). The transformation matrix between a new set of independent variables mi and the principal components are listed in TableIII.

With the new set of independent variables mi the quadric regression model after canonical transformation can be expressed as follow:

$$y = b_0 + \sum_{i=1}^{6} \theta_i m_i + \sum_{i=1}^{6} \lambda_i m_i^2 \qquad (3)$$

Where, y is attractiveness rating, mi are newly transformed variables, b_0 , θ_i and λ_i are contant, linear coefficients and quadratic coefficients, respectively.

TABLEIII. Transformation matrix and Eigenvalue from canonical analysis

	C1	C2	C3	C4	C5	C6
m_1	-0.12	-0.50	0.68	0.08	-0.43	-0.23
m_2	0.25	0.55	0.55	0.52	0.12	-0.14
m_3	0.80	0.07	-0.04	-0.47	-0.11	-0.33
m_4	0.43	-0.50	0.19	0.15	0.44	0.55
m_5	-0.03	0.40	0.26	-0.41	-0.34	0.68
m_6	0.29	-0.11	-0.32	0.53	-0.68	0.18

TABLEIV. The linear and quadric coefficients of the regression equation. Randomization test: * p<0.05, ** p<0.01, *** p<0.001.

	Linear coefficient	Quadric coefficient
m_1	-0.014	0.035**
m_2	0.051	0.016*
m_3	0.103***	0.006
m_4	0.060*	-0.013
m_5	-0.055	-0.045**
m_6	0.234***	-0.061***

Constant term: 5.35.

TableIV lists the linear and quadratic coefficients of the quadratic regression model. As highlighted, m_3, m_4, m_6 , m_1^2 , m_2^2, m_5^2 and $m_6^2$are significantly related to attractiveness ratings. The attractiveness ratings predicted using Equation (3) are highly related to the evaluated attractiveness ratings. The predicted attractiveness ratings are plotted against the evaluated attractiveness ratings in Figure 2. The liner regression equation is y = 0.946x + 0.277 with $R^2 = 0.939$.

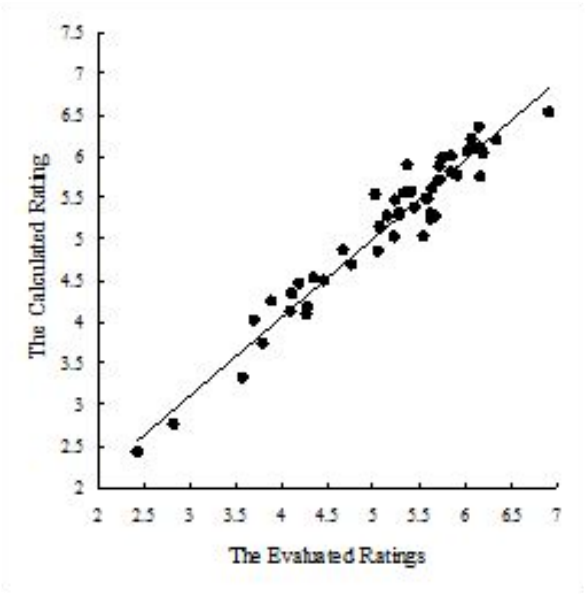


Figure2. Plots of calculated vs evaluated attractiveness ratings.

C. optimal body proportions

As listed in TableIV, three of six quadratic coefficients were negative indicting convex surface and other three quadric coefficients were positive indicting concave surface. For the three variables (m4, m5 and m6) having negative quadratic coefficients, the each of the stationary points correspond to the maximum value of the dependent variable (viz. body attractiveness rating). The respective stationary points of these three variables corresponding to maximum body attractiveness are at m_4=2.292, m_5=-0.605, and m_6=1.939.

For the three variables (m_1, m_2 and m_3) having positive quadric coefficients, the stationary points correspond to the minimum value of dependent variable (viz. body attractiveness rating). The values of m1, m2 and m3, which result in the greatest value of the dependent variable, must be the ones most distant from the respective stationary points, but still within the range of the variable under investigation. For these three variables, the respective stationary points are at m_1= 0.204, m_2=-1.604, and m_3=-8.436. Since for the stimuli investigated in this study, the range of m_1, m_2 and m_3 are [-4.475, 4.984], [-4.540, 6.270], and [-9.357, 4.687], respectively, the value of m_1, $m_2$and $m_3$resulting in maximum value of dependent variable are m_1=-4.475, m_2 =6.270, and m_3=4.687.

From the m values corresponding to the maximum female body attractiveness, we can get the optimum value of body proportions for body attractiveness. TableV lists the derived optimum value of body proportions together with the the percentile of the optimum body proportions.

IV.DISCUSSION

The percentile within the half standard deviation of a mean is between 30.85% and 69.15%. By examining the percentiles of the optimum body proportions in TableV,

we can see that 15 of the 39 optimum body proportions are within the half standard deviation, 14 are more than half standard deviation below the mean and 10 are more than half standard deviation above the mean.

TABLEV. The optimum body proportions and percentiles

Proportion	Min	Max	Mean	Optimum	Percentile（%）
Ratio of bust height over chin height (BHCH)	0.817	0.880	0.847	0.840	29.51
Ratio of waist height over chin height(WHCH)	0.720	0.767	0.742	0.737	32.61
Ratio of hip height over chin height(HHCH)	0.559	0.610	0.580	0.570	21.35
Ratio of crotch height over chin height(CRCH)	0.488	0.535	0.513	0.510	40.61
Ratio of leg length over chin height(LHCH)	0.670	0.739	0.692	0.690	44.94
Ratio of waist girth over chin height(WGCH)	0.453	0.681	0.515	0.472	16.90
Ratio of bust girth over chin height(BGCH)	0.582	0.775	0.653	0.612	17.08
Ratio of hip girth over chin height(HGCH)	0.622	0.756	0.681	0.671	38.39
Ratio of waist girth over hip girth(WHR)	0.683	0.901	0.755	0.704	10.20
Ratio of waist width over hip width(WWHW)	0.649	0.920	0.759	0.778	63.13
Ratio of waist depth over hip depth(WDHD)	0.750	1.000	0.855	0.781	10.73
Ratio of waist height over hip height(WHHH)	1.215	1.320	1.278	1.294	77.54
Ratio of waist girth over bust girth(WGBG)	0.723	0.879	0.787	0.774	35.18
Ratio of waist width over bust width(WWBW)	0.712	1.177	0.900	0.914	56.45
Ratio of waist depth over bust height(WDBH)	0.142	0.248	0.172	0.174	53.54
Ratio of waist height over bust height(WHBH)	0.848	0.904	0.876	0.878	58.53
Ratio of waist girth over abdomen girth(WGAG)	0.780	1.027	0.912	0.917	54.52
Ratio of waist width over should width(WWSW)	0.428	0.711	0.532	0.537	53.60
Ratio of waist width over waist depth(WWWD)	1.090	1.535	1.345	1.386	63.50
Ratio of BP distance over bust height(BPBH)	0.130	0.173	0.151	0.146	29.80
Ratio of hip width over hip depth(HWHD)	1.315	1.710	1.513	1.643	93.90
Ratio of BP distance over bust width(BPBW)	0.500	0.772	0.593	0.621	78.46
Ratio of BP distance over bust girth(BPBG)	0.177	0.208	0.196	0.200	72.89
Ratio of back width over back length(BWBL)	0.723	1.047	0.907	0.858	26.00
Ratio of waist height over body height(WHH)	0.598	0.646	0.625	0.621	33.61
Ratio of waist depth over bust depth(WDBD)	0.702	0.957	0.802	0.781	36.52
Ratio of waist width over bust depth(WWBD)	0.813	1.313	1.075	1.026	29.18
Ratio of waist width over hip depth(WWHD)	0.982	1.477	1.146	1.133	44.13
Ratio of bust width over waist depth(BWWD)	1.244	1.774	1.500	1.597	76.05
Ratio of hip width over waist depth(HWWD)	1.341	2.058	1.778	2.082	97.33
Ratio of bust width over bust depth(BWBD)	1.027	1.350	1.196	1.157	28.28
Ratio of bust depth over hip depth(BDHD)	0.941	1.222	1.068	1.081	59.48
Ratio of hip width over bust depth(HWBD)	1.254	1.635	1.419	1.526	88.53
Ratio of bust width over hip width(BWHW)	0.702	0.971	0.845	0.753	4.15
Ratio of bust width over hip depth(BWHD)	1.118	1.420	1.277	1.245	33.42
Ratio of bust width over should width(BWSW)	0.657	0.971	0.792	0.702	5.33
Ratio of hip width over should width(HWSW)	0.763	1.090	0.939	0.932	45.30
BMI	15.700	27.300	20.438	17.740	12.85
VHI	13.787	25.915	18.832	16.031	12.03

For the body proportions, which have the optimum values close to the means, averageness is attractive. For facial attractiveness, it was first discovered by Galton through photographic composites and later confirmed by Langlois & Roggman through digital composites that averageness is attractive [10-11]. It appears that this principle also applies to some female body attractiveness. The preference for averageness can be explained by the stabilizing selection principle that natural selection

imposes pressure against the two extremes in a population.

Form TableV, it is apparent that slimness, tallness, and hour-glass shape (viz. small waist, narrow bust width and wider hip) are preferred for female attractiveness. Slimness and tallness are related to fitness and health, and hour-glass shape (viz. small waist, narrow bust width and wider hip) is related to youth and fertility. It appears that there is a directional selection pressure in body attractiveness towards fitness, health, reproductively potential.

Based on the finding that majority of the optimally attractive body proportions are close to the population means, but some are close the extremes, it may be speculated that, female body attractiveness is a result of coupled effects of stabilizing and directional selection in natural evolution. From the study of female facial attractiveness, Perrett suggested that averageness is attractive, but not optimally attractive, and there could be a directional selection pressure on the evolution of human face shape [12]. It seems that the principle for face attractiveness discovered by Perrett et al also applies to the body attractiveness. Furthermore, the present study shows that the directional selection pressure in body attractiveness is there to improve fitness, health, and reproductive potential in evolution.

REFERENCES

[1] Zaastrabm, Seidell J. C., Van Noord, et al., "Fat and fecundity: prospective study of effect of body fat distribution on conception rates", British Medical Journal, pp.484-487, 1993.

[2] Singh, D., "Is thin really beautiful and good? Relationship between waist-to-hip ratio (WHR) and female attractiveness", Personality Individual Difference, vol.16, pp.465–481, 1994.

[3] Furnham, A., Tan, T., McManus, C., "Waist-to-hip ratio and preferences for body shape: a replication and extension", Personality and Individual Differences, vol.38, pp. 327–336, 1997.

[4] Malgorzata R. W., Boguslaw P., "Men's ratings of female attractiveness are influenced more by changes in female waist size compared with changes in hip size", Biological Psychology, vol.68, pp.299-308, 2005.

[5] Tovée, M. J., Maisey, D. S., Emery, J. L. & Cornelissen, P. L., "Visual cues to female physical attractiveness", Proc. R. Soc. Lond. B, no.266, pp.211–218, 1999.

[6] Tovée, M. J., Cornelissen, P. L., "Female and male perceptions of female physical attractiveness in front-view and profile", British Journal of Psychology, no. 92, pp.391–402, 2001.

[7] Tovée, M. J., Hancock, P. J. B., Mahmoodi, S., Singleton, B. R. R., Cornelissen, P. L., "Human female attractiveness: waveform analysis of body shape", Proc. R. Soc. Lond. B, no. 269, pp.2205–2213, 2002.

[8] Fan J., Liu F., Dai W., "Visual perception of female physical attractiveness", Proceedings of Royal Society of London Series B- Biological Sciences, no.271, pp.347-352, 2004.

[9] Fan J., Dai W., Qian X., "Effects of shape parameters on the attractiveness of a female body", Perceptual and Motor Skills, no.105, pp.117-132, 2007.

[10] Jintu Fan, K.P.Chau, XianfuWan, LiliZhai, EthanLau, "Prediction of facial attractiveness from facial proportions, Pattern Recognition, vol.45, pp.2326-2334, 2012.

[11] Langlois, J. H. and Roggman, L. A., "Attractive faces are only average", Psychol Sci, vol.1, no.2, pp.115-121, 1990.

[12] Perrett, D. I., May, K. A. and Yoshikawa, S., "Facial shape and judgements of female attractiveness", Nature, vol.368, no.6468, 239-242, 1994.

A Flocculation Simulation Method Based on Two-dimensional DLA Model

Chang Xu
Zhejiang Business Technology Institute, Ningbo, China
Email: netame@126.com

***Abstract*—By analyzing the classical DLA flocculation growth algorithm, the paper tries to improve the ideal hypothesis that the initial particle is distributed evenly. Through the MATLAB simulation platform, we do research on different particle motion step length, the maximum motion step number and collision probability, and their effects on morphology and structure of flocculation, when the total flocculation particle number is fixed. After comparing with the actual flocculation process, the paper raises group parameters which can describe the flocculation process accurately. It also provides a more effective method to make quantitative description of the flocculation process and prediction of the flocculation results.**

***Index Terms*—flocculation simulation, two-dimensional DLA, water treatment**

I Introduction

A The necessities of flocculation simulation

Flocculation is one of the important methods in water treatment. It can help migrate, transform and remove the particulate matter in water. The improvement of the flocculation technology can have direct effect on the subsequent deposition and filtration in water treatment [1].

Recently, there is a great development on the flocculation technology in both theories and applications. But most research work focuses on the amount of flocculation and the water quality of the flocculation effluent. In the real flocculation process, the water particles' shapes are changed complexly, the other factors which can influence the flocculation's formation and results include the quality of raw water, the hydraulic conditions during flocculation, the flocculants' types and dosage, the technology's operation conditions and etc [2,3,4]. Due to the limitations of the experimental conditions and analysis methods, these factors mentioned above are hard to be considered in the lab studies.

Therefore, it is necessary to conduct a simulation study on the flocculation's unstable, irregular and complicated motion process by using fractal theory and simulation technology. The study will establish a flocs fractal dynamics growth model, simulate the flocs growth process, and even breakthrough the limitations of experimental conditions by adopting computer simulation method. This study can help the interrelations between the dynamic growth model parameters and the influencing factors to the flocculation process. It also provides references to the quantitative descriptions of flocculation processes and characteristics under different influencing factors and to the modification of the flocculation technology [5, 6].

B Fractal flocs dynamic model

There several more mature fractal flocs dynamic growth models: Ballistic Aggregation (BA model) [7], Diffusion-Limited Aggregation (DLA model) [8] and Reaction-Limited Aggregation (RLA model) [9]. In BA model, particles and clusters are moving according to linear path, it can form dense flocs. DLA model pioneers the fractal aggregation simulation, in this way, the aggregates have obvious geometric center. RLA model supposes in the flocs growth process, particles and clusters can only have one success adhesion after many times' contact. On the contrary, BA model and DLA model both suppose particles and clusters will have adhesion immediately after the collision [10].

DLA model is considered to be the most important growth model. It has simple kinematics and dynamics equations. But it can produce fractal structure with the scale invariance and self-similarity. The model can fully reveal the fractal flocs growth mechanism in real process with low computation loads, thus it is regarded as the most usually used and practical flocs fractal simulation model.

C The shortages in the classic two-dimensional DLA model

The fractal evolution process of the classic two-dimensional DLA model is shown in Figure 1. In the two modeling process of the DLA model, it supposes the initial particle is distributed evenly [11], which is the ideal distribution extremely rarely seen in nature. On the other side, the DLA model only makes research on the condensed particles number's effect on the simulation's results, other parameters are set fixed. This method does not reflect other parameters in the model, such as particles motion step length, the motion step number and collision probably, and these factors' effects on flocculation are neglected. But actually, those factors have significant effects on flocculation process.

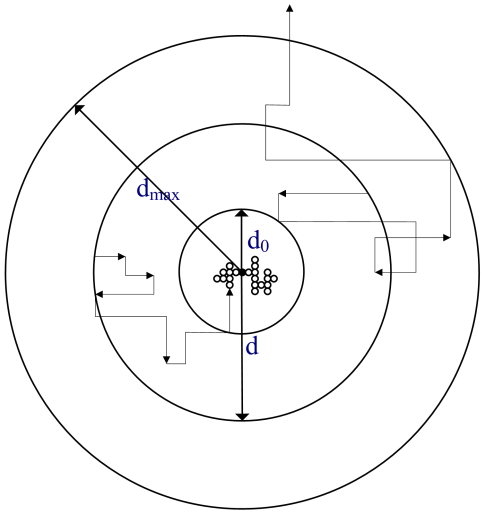


Figure1. The classic two-dimensional DLA model.

Hence, the classic DLA fractal dynamic model [12] can not get comprehensive and accurate simulation to flocculation procession and results. This improvement to the algorithm in the study can make more accurate mathematic description to flocculation process, and then improve the simulation's reliability and true scale.

II THE MODIFIED TWO-DIMENSIONAL DLA MODEL SIMULATION METHOD

A The modified simulation based on the classic DLA model

To the problems mentioned above in the classic DLA model, the modified one, according to the flocs particles distribution characters, states that the flocs particles is distributed normally [13]. Then based on the classic DLA model's simulation process [14], suppose the total flocs particles' number is fixed, the new model will do research on motion step length, step number and reaction probability respectively, get the fractal simulation results of these factor to other model parameters, such as the flocs fractal structure, density, strength and surface characteristics. Usually, the flocs fractal structure can be described by a non-integer fractal dimension. This research is adopting the radius of gyration to calculate the fractal dimension of simulated flocs.

B The simulation process of the modified DLA algorithm

The simulation process of the modified DLA modeling method is described as the follows: the motion step length, step number and reaction probability will be designed first, then the flocs simulation will be made according to different model's parameters' node, that is, to fix the particle number, place the seeds by DLA model, the particles are produced in turn. The average particle size is chosen to be μ , variance to be σ , all particles are normally distributed. After the collision between the particles and seeds, if the distance d between the particle center and seed center is smaller than the condensation radius d_0 , the condensation can be confirmed. If d is larger than the escape radius d_m , new particles will be produced until the simulation of fixed particle number is finished. Finally, we use the radius of gyration to calculate the fractal dimension of simulated flocs.

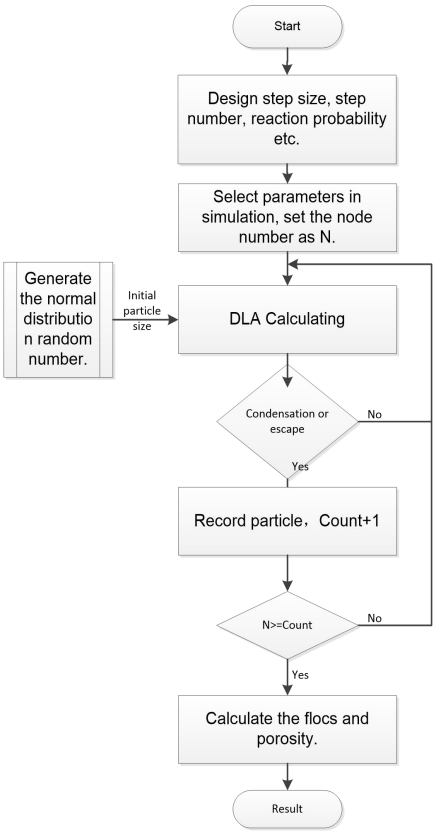


Figure2. The simulation process based on modified DLA

III ANALYSIS OF SIMULATION RESULTS

The following simulation values are all set in the drawing area of 200 × 200, the total flocs particles number N_n is 3000, the initial condensation particle radius d_0 is 2; the particle escape radius d_m is 10. The model parameters node can be seen in TABLE I.

TABLE I.
THE DESIGN OF DLA MODEL PARAMETERS NODE.

Step length Ns	0.5, 1.0, 1.5, 2.0, 2.5, 3.0
Max step number MS	50, 100,150, 200, 250, 300
Collision probability（Px，Py）	(0.1,1.0), (0.2,0.9), (0.3,0.8)

A Simulation with the changed particles motion step length

In this stage, the flocculation simulation is done according to different particles motion step length Ns (see Table 1) by using the radius of gyration to calculate the fractal dimension. After each Ns simulation, the flocs porosity's (Vf) and fractal dimension's (Df) changes will be recorded. Ns's optimal value will be confirmed according to the comprehensive analysis of simulation results. The results are shown in Figure 3 and TABLE II.

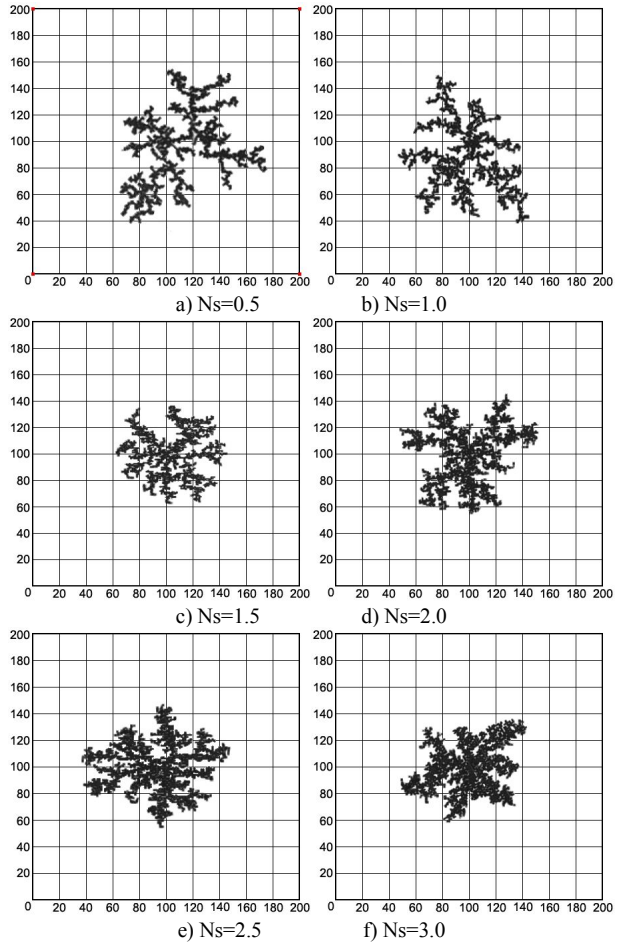


Figure 8. Simulation results with different Ns.

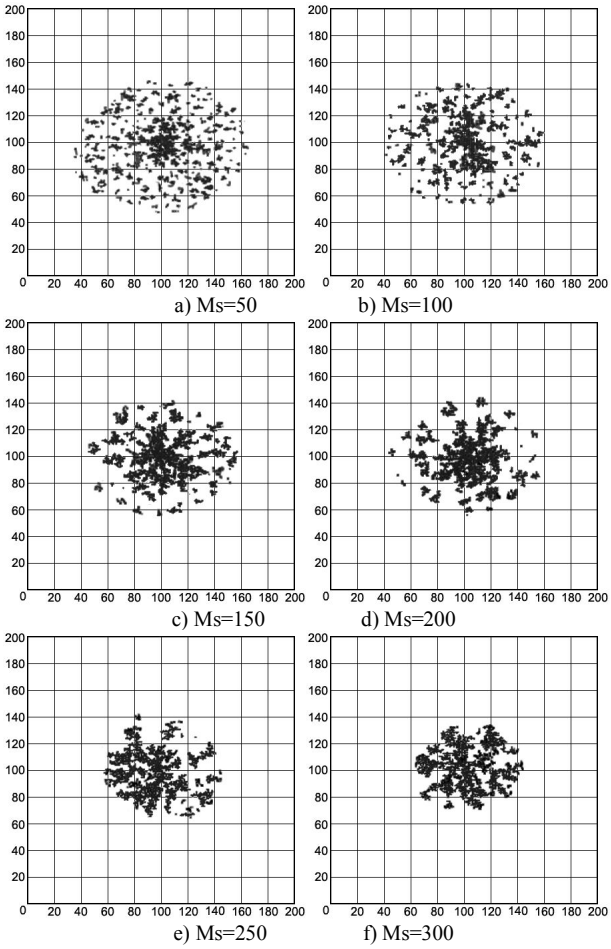


Figure 8. Simulation results with different Ms.

TABLE II.
SIMULATION RESULTS WITH FIXED PARTICLE NUMBER AND CHANGED MOTION STEP LENGTH.

Ns	0.5	1.0	1.5	2.0	2.5	3.0
Vf	0.91	0.89	0.86	0.85	0.75	0.77
Df	1.6	1.51	1.65	1.648	1.75	1.74

With the increase of Ns, the porosity of flocs decreases, fractal dimension increases, the flocs become dense and branching decreases gradually. At this time, the simulated flocs will not bear the fractal structure anymore, and it has bigger deviations with the real flocs pattern. It will be found that, if we put the simulation results and experimental results together, when Ns=2.0, the simulated flocs has obvious fractal structure characteristics, and it has the smallest deviation with the actual flocs pattern.

B Simulation with the changed particle motion step number

In this stage, the maximum simulated particle motion step number is Ms (see TABLE I). When the random particle motion step number reaches Ms, the particle will stop and become a new seed. The simulation results are shown in Figure 4 and TABLE III.

TABLE III.
SIMULATION RESULTS WITH FIXED PARTICLE NUMBER AND CHANGED MAXIMUM MOTION STEP NUMBER.

Ms	50	100	150	200	250	300
Vf	0.91	0.875	0.835	0.841	0.83	0.872
Df	1.432	1.51	1.6	1.625	1.65	1.725

Form the above results, with the increase of Ms, the flocculation porosity decreases, the fractal dimension increases. When Ms >300, the flocculation fractal dimension keeps unchanged. When Ms>400, the flocculation does not have obvious clusters anymore, it is growing like continuous branches and has obvious fractal structure characteristics. So Ms can be chosen between 300 and 400 in this flocculation simulation.

C Simulation with the changed collision probability

In the real flocculation process, the reaction flocculation can not be made every time of the particle collision. Considering the particle collision probability's effect on the simulation results, we can change the particles reaction probability (Px) [15] in the X-direction and the particles reaction probability (Py) in the Y-direction, make records of each group, we can get the relations between the flocs porosity (Vf) and fractal

dimension (Df), which are shown in Figure 5 and TABLE IV.

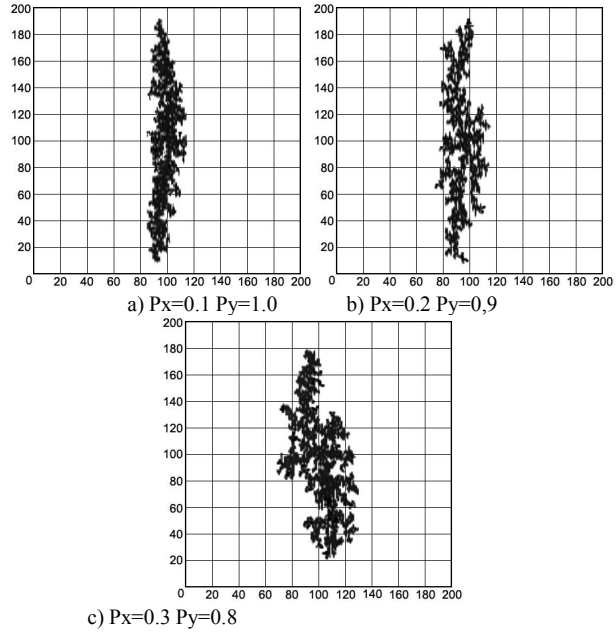


Figure 8. Simulation results with different collision probability.

TABLE IV.
SIMULATION RESULTS WITH FIXED PARTICLE NUMBER AND CHANGED PARTICLE COLLISION PROBABILITY.

（Px,Py）	(0.1,1.0)	(0.2,0.9)	(0.3,0.8)
Vf	0.98	0.95	0.93
Df	1.08	1.26	1.4

From the data, with $|Px-Py|$, the absolute value of the probability difference in X and Y direction, increases, the flocs fractal dimension decreases and porosity increases gradually. It is because that, when the reaction probability is bigger, it needs less collision times between particles and seeds or flocs. When the reaction probability is bigger in one direction, flocs will grow toward this direction; the flocs will become fluffier and has bigger porosity and smaller fractal dimension. This is consistent with the real flocculation process. It can be concluded that the particle collision probability is one of the influencing factors to flocs pattern.

IV CONCLUSIONS

The research on the flocs pattern is the most important part in flocculation study. The modified DLA model simulation method can reproduce or predict the flocculation process and results more reliably. The research on flocs pattern will become more effective by the new model. In the further study, we can make comprehensive choice in the DLA model parameters, through scientific experimental method, such as uniform design, to develop the Monte Carlo simulation and finally get the best parameter combination in the simulation.

ACKNOWLEDGMENT

This work was supported by Educational Commission of Zhejiang Province (Y201224212).

REFERENCES

[1] Zheng, D. M., Zhu, S. G., Jiang, W. T., and DAI, J, "Simulation on Flocculation Process Based on DLCA Model and Its Application," Journal of Nanchang University (Engineering & Technology), vol. 2, pp. 006, 2011.

[2] Zhang, J. Z, Fractal, Beijing: Tsinghua University Press, 1995

[3] Qiu, Zumin, Dongjing Liu, and Ru Zhang, "Simulation of two-dimensional floc growth using improved DLA model," 2011 IEEE 2nd International Conference on Computing, Control and Industrial Engineering (CCIE),IEEE, vol. 1, 2011.

[4] Zhang Jian-hua, Liu Zhen-hua, "Study of The Relationship between Fractal Dimension and Viscosity Ratio of Viscous Fingering with a Modified DLA Model," Journal of Petroleum Science and Engineering, vol. 21(1), pp. 123-128, 1998.

[5] Yu Zhen-xun, Tan Kai-xuan, and Yi Zheng-ji. "Advance in Simulation of Flocs Growth and Their Fractal Properties," Journal of University of South China (Science and Technology), vol. 21(2), pp. 46-48, 2007.

[6] Jin Peng-kang, Liu Huan, Zhang Jing, and Wang Xia-chang. "Kinetic Simulation of Coagulation by Monte-Carlo Method," Journal of Xi'an University of Architecture & Technology (Natural Science Edition), vol. 40(3), pp. 398-402, 2008.

[7] Dey, Supravat, Dibyendu Das, and R. Rajesh, "Lattice models for ballistic aggregation in one dimension," EPL (Europhysics Letters), vol. 93(4), 2011.

[8] Menshutin, Anton, "Scaling in the Diffusion Limited Aggregation Model," Physical Review Letters, vol. 108(1), 2012.

[9] Zhu, Wancheng, "Hierarchical mesoporous SrCO3 submicron spheres derived from reaction-limited aggregation induced "rod-to-dumbbell-to-sphere" self-assembly," CrystEngComm, vol. 12(6), pp. 1795-1802, 2010.

[10] Williams H.T. Goodwin L. and Steven G, "Two-Dimensional Growth Models," Physics Letter A, 1998, 250(2): 105-110.

[11] Guo Hang, Study on The Fractal Dynamic Simulation Technology of Floc in Water, Harbin: Harbin Institute of Technology, 2012.

[12] Zhang, J., Zheng, Y., Li, C. Y., & Chai, L. H, "A confined DLA model of microbial biofilm in wastewater treatment systems," 2011 International Conference on Electrical and Control Engineering (ICECE), IEEE, vol. 1, 2011.

[13] Yang, Y. Y., Ren, X. J., Li, S. C., Zhu, L. M., & Tian, S. Y, "Quantitative Investigation of Flocs Flotation Removal and Hydrocyclonic Separation for the Washing Water Treatment," Applied Mechanics and Materials, vol. 333, pp. 1857-1861, 2013.

[14] Cha, S., Park, J., Hwang, J., and Noh, J, "An efficient diffusion model for viscous fingering," The Visual Computer, vol. 28, pp. 563-571, 2012.

[15] Iliadis, C., Longland, R., Champagne, A. E., Coc, A., and Fitzgerald, R, "Charged-particle thermonuclear reaction rates: II. Tables and graphs of reaction rates and probability density functions," Nuclear Physics, vol. 841(1), pp. 31-250, 2010.

The Off-line Fire Truck Data Collecting and Analyzing System Based on PC

S.Li
Railway police college , Zhengzhou City, China

***Abstract*—The fire truck is an important equipment for the fire alarm, which provides mobility support for the fire officers and soldiers to fight against fire. This paper introduces a set of off-line fire truck data collecting and analyzing System, for the collection and storage of fire truck process data in the operation, and this system is capable of doing multi-display of data and fault analysis of the fire truck for its potential engineering problems. System uses a collection of flash memory, the host computer display and analysis of the structure, the paper gives the overall design and the final part of the software renderings.**

***Index Terms*—fire truck, off-line data collecting, data analysis, fault diagnose**

I. Introduction

Fire-truck, known as the rescue train, is special for use as a fire or other emergency use of the vehicle, which is a general term for all kinds of fire fighting vehicles equipped with a variety of fire equipment, and it is the basic mobile fire fighting tool of fire fighting forces. The quality level of fire truck reflects the national fire equipment level, even the whole level of the fire protection industry and government. In the process of daily maintenance of fire truck, previous operation data of some devices in the fire truck is the focus of attention of the mechanic, but most of the current domestic fire truck is not equipped with a data acquisition and analysis system, which makes the mechanic, during the maintenance time, can not understand past operation status of the key component of the vehicle when treating the fire cases. This paper introduces a set of off-line fire truck data collecting and Analyzing System, for the collection and storage of fire truck process data in the operation, and the data stored is able to be transferred to flash or other mobile storage by USB protocol. The data can be read by the PC software, which does the display, graph, and fault analysis. With the application and development of automotive electronic technology, automotive electronic control system is becoming more and more complicated. Diagnostic method and device of traditional both accuracy and ease of use, and adaptability to the development of automobile technology, can not meet the needs of users. In order to improve the fault diagnosis technology, the diagnosis theory and method must be enhanced, and latest achievements are widely applied in various disciplines, which is with the aid of mathematical tools and computer. The design of hardware and software will be proposed in this paper[1-3].

II. Design of Hardware

The hardware is based on AVR microcontroller, and the key data for the fire truck includes engine speed, pump speed, water tank level, nozzle pressure and the engine temperature. Thus, the sensors are carefully selected including 0-2600r/min engine speed sensors with minimum indexing unit 1r/min, 0-1480r/min pump speed sensor with minimum indexing unit 1r/min, 0-1600mm water tank water level sensor with minimum indexing unit 1mm, 0-0.9MPa nozzle pressure sensor with minimum indexing unit 1KPa, and 0-200C temperature sensor with minimum indexing unit 0.1C. Because the battery voltage of fire truck is 24V, and the hardware all work under 5V, the 24V to 5V transistor is employed. In the auto ignition operation, the battery began to supply the slave system, and the system begins to collect fire engine run time data. The hardware structure is shown in figure 1.

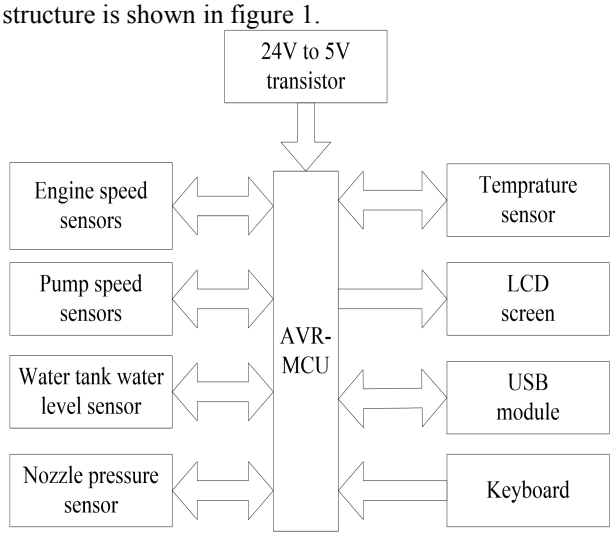


Figure1. Hardware structure

After collecting running data of fire truck, the data is stored in form TXT files, with data format shown in table I. The time in expressed in hour, minute and second format, and each of them occupied 1 bytes. The water level, nozzle pressure and temperature occupied 2 bytes, same as engine speed, pump speed and car speed. If the high position is 1, it means minus data. @@is synchronization.

All the data is store in sequence from low position to high position 30 seconds per time. If the data of water temperature, water level, pressure and flow is 0xFFFF, it means that the device is power-off. The data file is name as xxyearxxmonthxxday.txt such as 070314.txt.

Table.I
Data Format

Name	Symbol	Unit	Range
hour	Hour	1 hour	0-23
minute	Min	1 minute	0-59
second	Sec	1 second	0-59
temperature	Temperature	0.1°C	-20.0-200.0°C
engine speed	Engineering Speed	1r/min	0-2600 r/min
pump speed	Pump Speed	1r/min	0-1480r/min
car speed	CarSpeed	0.1km/h	0-250km/h
water level	WaterLevel	1mm	0-1600mm
water press	WaterPress	1KPa，即 0.001MPa	0-0.9MPa
water flow	WaterFlow	0.01m3/h	0-50m3/h
Syn-chronization	@@		

III. Design of Software

The software structure is shown in figure 3, which contains functions： login function, main menu function , data importing function, data query and display function, data image display function, fault diagnosis function and result output function.

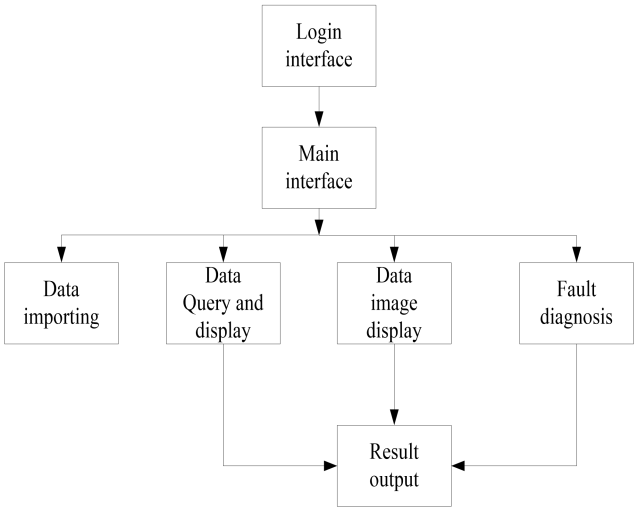


Figure.2 Software structure

The interface of software is programmed by VB, and the Access is employed as database tool. The main interface is shown in figure 3, and the data can be imported from the TXT files into database.

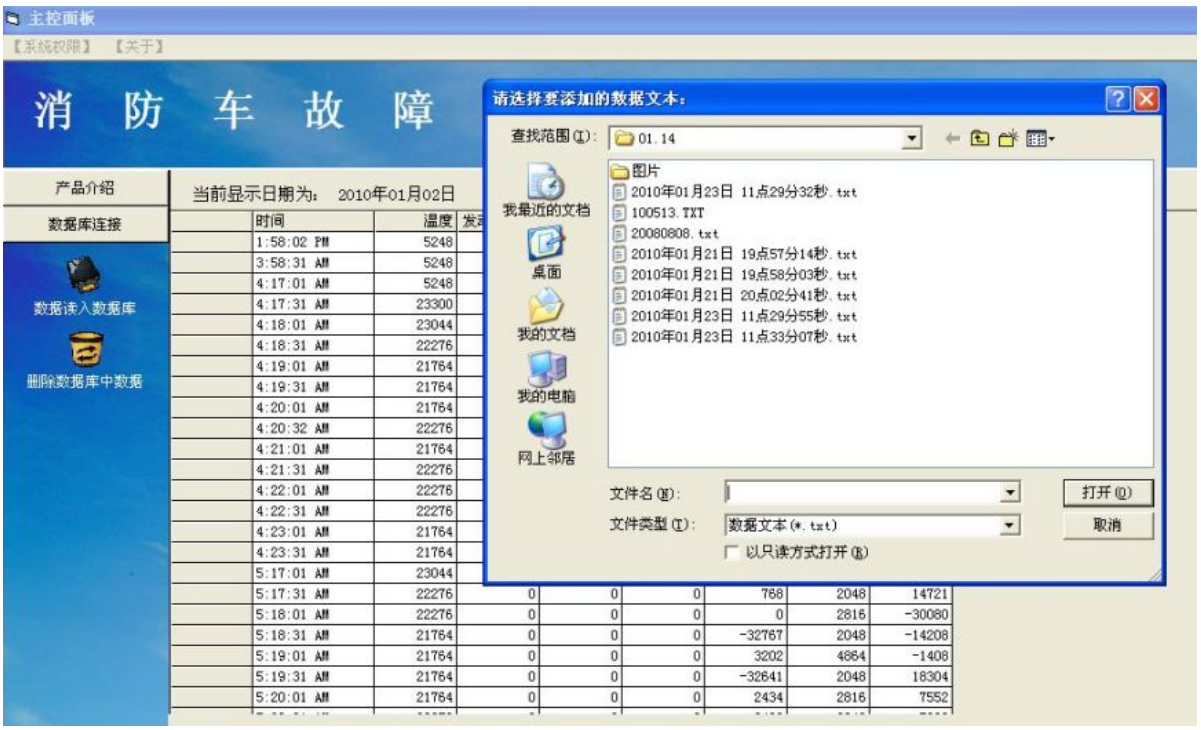


Figure.3 Main Interface

The data in database can be queried and image displayed by condition, and the image display can shown the diagnose result of the fire truck by analyzing the data.

The fault diagnose method is implemented by experts algorithm, and the diagnose result also can be shown in form of data, shown figure4 and figure5[4-5].

Figure.4 Diagnose in Image

Figure.5 Diagnose in files

IV. Conclusion

The fire truck data acquisition and the analysis system introduced in this paper, is capable of solving the problem of better vehicle operation data storage, and it can do display and analysis according to the data, with the intuitive way to show it. The system obtains the characteristics of simple operation, friendly and practical interface. By using he fault diagnosis system, we can quickly, accurately diagnose the common faults of engine, thereby reducing the probability of fault misdiagnosis caused by lack of vehicle maintenance and repair personnel level and experience to a certain extent.

ACKNOWLEDGMENT

This work was supported by Railway Police College funding(JY2013068).

REFERENCES

[1] XIN Hui-juan, "Development and Research of Automotive Engine Fault Diagnosis Expert System of Network Based on ASP" *Journal of agriculture mechanization research*, vol. 6, pp. 190–193, 2006.

[2] Guo rong chun, "Study on Engine Remote Fault Diagnose based on Virtual Instrument" *Journal of Shandong University of technology*, vol. 17, pp. 13–16, 2003.

[3] Alex Toussaint, "Java rule engine API specification." *JCP final review* , pp. 22–28, 2003.

[4] XU Li-chao, MACong-bing, FANQin-man, BAO Xu, "Fault Diagnosis Method Study for Automobile Engine Based on BP Neural Network" *Tractor& Farm Transporter*, vol.34, pp. 101–104, 2007.

[5] WU Yu-xiang, ZHANG Ze-jian "Design of the diagnostic system of automobile electronic control fail based on personal computer" *Manufacturing auotomation*, vol.12, pp. 51–53, 2006.

Analysis of Building Electrical Lighting Energy Saving Based on Artificial Neural Network

YU.Chen
Zhengzhou Vocational University of Information and Technology,Zhengzhou,China
Email:181817440@qq.com

***Abstract*—According to the actual situation in an office building electrical lighting in Zhengzhou area, this study gives a solution of energy saving, and energy saving calculation, establishes an evaluation model of artificial neural network based on the lighting energy conservation office. The calculation results show that the effect of energy saving, the method is obviously, energy saving assessment model is correct, and it has good application value.**

***Index Terms*—building electrical, lighting energy saving, artificial neural network, model**

I. INTRODUCTION

In recent years, energy conservation and sustainable development has increasingly become the focus of attention of the whole society, also caused the human reflection to increasingly serious environmental problems. Aiming at the energy situation in China, building energy conservation has become an important part of a series of measures in our country, the national and local successively promulgated a number of related to building energy efficiency standards. In China, the lighting electricity accounted for total electricity consumption of about 12%, so energy-saving lighting is an important part of building energy saving. 。 In the lighting design, in addition to the correct use of lighting, lighting type, meet the lighting quality requirements, but also pay close attention to energy-saving lighting.

According to an office illumination conditions, studied illumination energy saving and design optimization, set up energy-saving electrical lighting model. The model was trained using artificial neural network, model evaluation accuracy, provides a scientific basis for the implementation of energy-saving, help to realize intelligent lighting design[1-2].

II. ELECTRIC LIGHTING ENERGY SAVING

Energy saving lighting design should follow the following principles: in the premise with lighting quantity and quality sufficient condition, should try to save electricity. Energy saving lighting should improve the efficiency of the entire lighting system, and can not lose the lighting quality and one-sided emphasis on energy saving.

Lighting design should consider the measures of energy saving lighting from the following aspects:

1) Should be selected in the workplace according to national standard luminance. And a suitable light source is selected according to different occasions.

2) The selection of lamps reasonable light distribution, high efficiency, to meet the glare restriction conditions, should be preferred to open direct lighting (efficiency is not lower than 75%).

3) Rational use of local lighting. The same room with small range high luminance requirements, priority should be given to the local lighting to meet.

4 ） Select the ballast inductance ballasts, should choose the factor of electronic ballasts and energy-saving high power, and the energy efficiency grade products.

5 ） Distribution line indoor lighting should be used copper conductor, conductor cross-section distribution lines should be selected reasonably, and it may be appropriate to increase, in order to reduce the line impedance.

Lighting design for a certain office building in Zhengzhou was investigated; the data reflects the level of design office building lighting power consumption. According to the classification, the focus of the investigation of the survey found, office power density distribution more in $10 \leqslant LPD < 18$, accounted for the sample 61.7%; ordinary conference room divided power density distribution more in $10 \leqslant LPD < 18$, accounted for the sample 54.1%; office for the power density distribution more in $LPD < 10$.

III. BUILDING ELECTRICAL LIGHTING ENERGY-SAVING EVALUATION MODEL

The comprehensive benefit of technology of electric lighting energy saving can be divided into economic benefits, social benefits, ecological benefits. Economic benefits directly reflect the energy saving effect; social benefits including human comfort, project aesthetics; ecological benefit mainly refers to whether it is conducive to environmental protection. Therefore, building energy efficiency evaluation index of electric lighting energy saving effect are the main index, the index of human body comfort, appearance of the indicators, environmental indicators project. Between these 4 indicators is a nonlinear relationship[3].

BP neural network algorithm is a set of input / output problem into a learning method for nonlinear

optimization problems and the gradient algorithm using iterative value to solve the problems of power.

From the past research has proved, with Sigmoid nonlinear function of 3 layers BP neural network can approximate any continuous function with arbitrary precision. BP network with gradient descent algorithm, the learning rate is fixed; the network training time is long, and the possible local convergence, the learning and memory network with instability. Improved BP algorithm and Levenberg-Marquadt back propagation algorithm with momentum factor (L-M back propagation algorithm) converges faster than the basic BP algorithm, fast convergence and good stability.

And then analyze the BP momentum factor method and L-M back propagation algorithm. 4 input hypothesis of neural network respectively by a1, a2, a3, a4 says, the hidden layer neuron input

$$S_{pj} = \sum_{i=1}^{4} w_{pij} \times a_{pj} - \theta_{pj} \text{ j=1, 2…n} \quad (1)$$

In the formula,

n- The number of neurons in hidden layer

w_{pij} - The connection weights between the hidden layer of the j neurons and I neurons in the input layer

a_{pj} - The hidden layer of the j neurons of input values

θ_{pj} - The hidden layer of the j neuron threshold

The hidden layer activation function was tansig, so the output of hidden layer j neurons is

$$b_{pj} = \tan sig(s_{pj}) \quad \text{j=1,2,…n} \quad (2)$$

The output layer activation function was purelin, the number of neurons was 1, so the input and output of neurons in the output layer is

$$L_p = \sum_{j=1}^{n} v_{pj} \times b_{pj} - r_p \quad (3)$$

$$M = purelin(L_p) \quad (4)$$

In the formula,

v_{pj} - The output connection weights between the hidden layer neurons and j neurons

r_p - The output layer neuron threshold

The error function is

$$MSE = \frac{1}{PL} \sum_{P=1}^{P} \sum_{k=1}^{L} (T_{pk} - O_{pk})^2 \quad (5)$$

In the formula,

P- The number of sample

L- The number of output neurons

T_{pk} - Expected output neuron k

O_{pk} - The actual output neuron k

Every time in the standard BP algorithm does not consider other modifications are modified training output error is reduced to solve this method, leading to an increasing number of iterations problem.

The right connection coefficient training constantly adjust the network, so that the value of < permissible error network error function. BP momentum will last modified weight a portion of the amount added to the error calculation of the weights of correction, as the actual weight adjustment this time.

$$\Delta\omega = \eta G + \alpha \Delta\omega(n-1) \quad (6)$$

In the formula,

G- The negative gradient MSE

α - Momentum coefficient，$0 < \alpha < 1$

This reduces the oscillation trend in the learning process, improve the speed of convergence, basically solved the problem of local minima. L-M back propagation algorithm with the gradient descent method and Newton method advantage, the network weights less number of convergence is very fast, better performance.

A BP neural network is 3 layer neural network of a single hidden layer. Input as evaluation index, a total of 4 inputs, so the input neural element into 4. The output for the evaluation results, the project is feasible or not. Set the output value of 1 indicates that the project is feasible, 0.5 representing the project improved feasible, 0 said the project is not feasible, so the output neural element into 1. Determine the number of hidden layer neurons number is 4, the newff () function to create a network, setting the target error is 5e-3, the maximum training step 2000.

Because the input sample is 4 indicators, so be normalized. These 4 indicators are as follows: 1) Energy saving effect index. The sample of saving energy consumption energy consumption and energy consumption compared to, for a maximum of 1.

2) Human comfort index. Don't feel comfortable 0, feeling a little bit of comfort for 0.2, feeling a little comfort for 0.4, can also feel as comfortable as 0.6, 0.8, feels very comfortable for 1.

3) Project appearance index. Don't feel beautiful 0, feeling slightly appearance of 0.2, feeling a little appearance of 0.4, can also be felt for 0.6, feel beautiful 0.8, feel very beautiful 1.

4) Environmental protection index. No environmental protection was 0, with a little green is 0.2, a little green is 0.4, also can be 0.6 to 0.8, environmental protection, environmental protection 1.

In the above mentioned office as an example, the learning samples by learning algorithm of BP neural network, the results as shown in the table, where L-M back propagation algorithm faster training speed, the training step number <10, the network generalization performance is good, not local convergence, the training sample network output value and the maximum relative error is less than 6% of the expected value the test sample network, output value and the maximum relative error of expected value is less than 4%. Lighting energy saving scheme is feasible using in this area, the results shows that the feasibility of energy saving evaluation to improve the neural network and effectiveness[4].

IV. CONCLUSION

In this study, the use of artificial neural network to evaluate the energy-saving electrical lighting, the weight factor implied in the network, the learning, memory, pattern recognition function of neural network, and has the advantages of simple, practical. Therefore, the lighting design based on artificial neural networks theory calculation method is a more excellent energy saving lighting design method.

REFERENCE

[1] Su,J.Y.(2007). “Comparison and fusion of computational intelligence algorithms”, Journal of Chinese Electronic Research Institute.2, 52-56, 61

[2] Francis Rubinstein， Michael Siminovitch ,and Rudolph Venderber , “Fifty Percent Energy Saving with Automatic Lighting Controls”， IEEE Transactions ON INDUSTRY APPLICATIONS V0L． 29， N0.4 JULY/AU—GUST 1993

[3] Simon Haykin.(2001). Neural network comprehensive basis (Second Edition) Beijing: Tsinghua University press.

[4] H. Kumagawa: "The Lighting Control Techniques and the Applications to Realize Energy Saving", J. Illum. Engng. Jpn., Vol. 90 No.9, pp.661-665 (2006-9) (in Japanese)

Determination Method of Piecewise Linear Membership Function Based on the Interval Density Cluster

Qi. Liu
School of Physics and Electronic Engineering, Zhoukou Normal University, Zhoukou, Henan, China

Shaohui. Wang
School of Physics and Electronic Engineering, Zhoukou Normal University, Zhoukou, Henan, China

Honghui Zhang
School of Physics and Electronic Engineering, Zhoukou Normal University, Zhoukou, Henan, China

***Abstract*—Triangle and monotone membership functions have simple forms and clear physical meaning, which are commonly used in the research work. In this paper, a density clustering method based on the interval is brought forward, which has avoided the cluster results falling into a local minimum, also by using this method, the cluster center can be quickly obtained. Through simulation, the piecewise linear membership function of variation of total volume in petroleum drilling is determined, which solves the problem that the function is hard to be defined.**

***Index Terms*—piecewise linear membership function; interval; density cluster; cluster center; petroleum drilling; variation of total volume**

I. Introduction

The research of phenomenon and things with uncertainty is significant[1]. Doctor Zadeh put forward the membership degree[2] to express the uncertainty. There are two main forms of membership function[3]: piecewise linear function and nonlinear function. But how to determine the membership function[4] has not been fundamentally resolved, in practice, it is unreliable to determine the membership value subjectively, if using the statistical method, it will not only cost too much, but also can not be achieved sometimes.

Fuzzy clustering analysis is the foundation of classification and system modeling problems. The purpose of fuzzy cluster is to extract the inherent characteristics from a large data set, and obtain the compact representation of the system behavior. Doctor Bezdek put forward a Fuzzy C-mean Clustering method, which is a mature method at present[5]. But there are some shortcomings in this algorithm, for example, it is particularly sensitive to cluster center, and is easy to fall into a local minimum[6]. If the cluster center is chosen randomly, the objective function is obtained by multiple exercises, but which still costs much time and can not completely avoid the local optimization. In this paper, an interval density clustering method based on interval density is put forward, by using this method, the cluster center of the data is quickly obtained, and the piecewise linear membership function is determined too, in a word, this is a relatively simple and fast algorithm. [7]

II. Piecewise Linear Membership Function

There are some piecewise linear membership functions, such as triangle, monotone, ladder, in which, triangle and monotone membership functions have simple forms and clear physical meaning, and they are commonly used in the research work. In this paper, the determination of triangle and monotone membership function is mainly studied. [8-23]

The expression of the triangle membership function is shown as equation (1), which is suitable for the middle linguistic value modified by middle and more, here, the function is determined by parameters a, b, c, shown as Fig.1.

$$f(x,a,b,c)=\max\left\{\left[\min\left(\frac{x-a}{b-a},\frac{c-x}{c-b}\right)\right],0\right\} \quad (1)$$

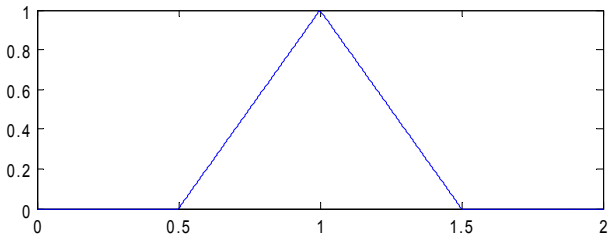
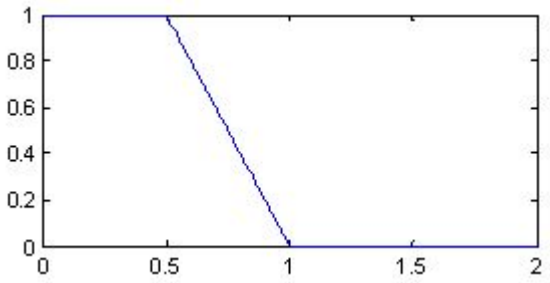


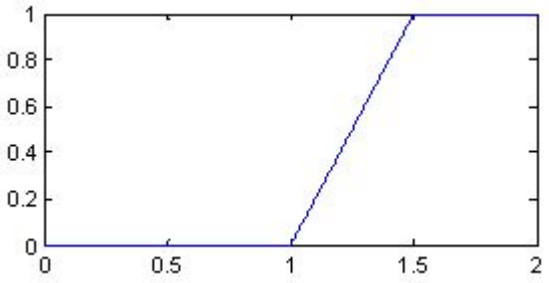
Figure 1. Triangle membership function

The expression of the monotone membership function is shown as equation (2), which is suitable for the bipolar linguistic value modified by very and very not, here, the function is determined by parameters d, e, shown as Fig.2.

$$f(x,a,b)=\begin{cases}\max\left\{\left[\min\left(\frac{x-d}{e-d},1\right)\right],0\right\} & \text{Single increasing}\\ \max\left\{\left[\min\left(1,\frac{e-x}{e-d}\right)\right],0\right\} & \text{Single reducing}\end{cases} \quad (2)$$



(a)Single reducing



(b)Single increasing

Figure 2. Monotone membership function

III. Density Clustering Method Based on the Interval

In this paper, a density clustering method based on the interval is brought forward, first, according to the interval density, the data is classified, and then the clustering center is obtained by the density, which effectively avoids the cluster results falling into a local minimum.

For the data $X=\{x_1,x_2,...,x_n\}\subset R$, the density clustering method is expressed as follows:

Step1: Set threshold τ of the interval density, and number of cluster center q;

Step2: Set width d of the interval, and divide the data set X into interval S_i with width d;

Step3: Scan the data, and get the density of the interval, that is the number of the data in each interval;

Step4: Select the interval S_k with the largest density, and search S_{k-1} from the left, if the density threshold is larger than τ, search S_{k-1} until the density threshold is smaller than τ; search S_{k+1} from the left, if the density threshold is larger than τ, search S_{k+2} until the density threshold is smaller than τ; at last the continuous intervals l with density threshold larger than τ is obtained.

Step5: Calculate the density index of each data x_i:

$$D_i=\sum_{j=1}^{n}\exp\left[\frac{-\|x_i-x_j\|^2}{(0.5r_a)^2}\right] \quad (3)$$

Select the data x_{c_1} with the largest density as a cluster center, and delete the continuous density interval l;

The field radius r_a in equation (3) is:

$$r_a=\frac{l}{2}\times d \quad (4)$$

Step6: Execute step 4 and 5 continuously until the cluster center q is obtained, and define the cluster center as w_m, $m=1,2,...,q$.

In the above method, r_a represents radius of the continuous interval l, and is always decided by equation (4). In practical analysis, the cluster center q has no effect on the cluster results, as a result, radius r_a is determined by equation (4), and the computational complexity is smaller, the center appears earlier.

IV. Simulation

By using the density cluster method based on the interval to get the cluster center, the piecewise linear membership function can be obtained. In this paper, the total volume data EB_Da51 from the petroleum drilling[7] in Hubei province are selected, and the piecewise membership function of variation of the volume are determined by Matlab. Some data of EB_Da51 are shown in Fig.3, in which the ordinate represents variation of the total volume, and the unit is m^3.

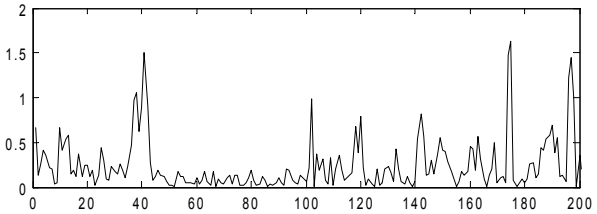


Figure 3.Change of volume data EB_Da51 from the petroleum drilling

On the basis of the initial cluster center q obtained by the density clustering method, the cluster center w_m and some other parameters, the concrete process of piecewise linear membership function is described as follows:

Step1: Unify universe of the data, and determine the number of the linguistic value. In this paper, membership function of the change of the total volume in petroleum

drilling is seek, and data EB_Da51 is mapped to a unified universe space, there are four linguistic values: none, small, middle and large, that is to say, c, $1 \le k \le q$.

Step2: Determine the type of the membership function, that is to say, Small and Middle mean the middle linguistic value, which are expressed by triangle membership function, None and Large mean the bipolar linguistic value, which are expressed by monotone membership function.

Step3: Calculate the clustering center w_m by the piecewise density cluster, and the obtained centers are $w_m = 0.0786$, w_m=0.3044 , $w_m = 0.6607$, $w_m = 1.4180$.

Step4: According to the vicinity principle, the parameters of the triangle and monotone membership function are shown in Tab.1.

TAB.1 PARAMETERS OF PIECEWISE LINEAR MEMBERSHIP FUNCTION

	Linguistic value	Type of the function	Parameter
1	None	monotone	$d = 0.0786$ $e = 0.3044$
2	Small	triangle	$a = 0.0786$ $b = 0.3044$ $c = 0.6607$
3	Middle	triangle	$a = 0.3044$ $b = 0.6607$ $c = 1.4180$
4	Large	monotone	$d = 0.6607$ $e = 1.4180$

Step5: According to the parameters in Tab.1, the membership function is shown in Fig.4.

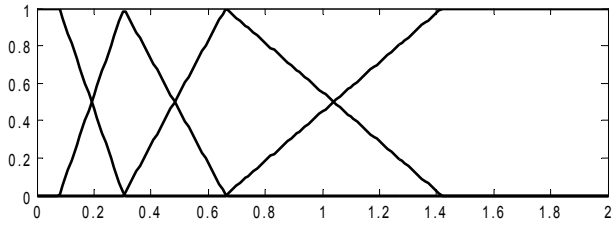


Figure 4.Piecewise linear membership function of the total volume

V. CONCLUSION

According to the case study, the density cluster method brought forward in this paper has no iteration, in this method, the initial cluster center is obtained by the density of interval, which can effectively solve the problem that fuzzy c-means method heavily depends on the initial cluster center randomly generated, and obtain the center rapidly. By using the density cluster method proposed in this paper, the triangle and monotone membership functions of the change of total volume in petroleum drilling are obtained, which is a simple and fast method to get the piecewise linear membership function.

ACKNOWLEDGMENTS

The research of this paper has been sponsored by Henan Provincial Research Foundation for Science and Technological Breakthroughs, China (Grant No. 132102210179), Natural Science Foundation of Henan Educational Committee, China (Grant No.2011B510021).

REFERENCE

[1] Li Deyi, AI research and development in the network age [J]. Caai Transactions on Intelligent Systems,2009,2:1-6(In Chinese).

[2] Zadeh LA. Fuzzy sets [J]. Information and Control,1965,8:338-353.

[3] WU Zhong-dong, GAO Xin-bo, XIE Wei-xin, A study of a newfuzzy clustering algorithm based on the kernel method [J]. Journal of Xidian Normal University,2004,4:533-537(In Chinese).

[4] LIU Qi, DU Yuan-dong. Application Research of PID Control Optimized by GBLB-PSO in HVAC System [C]. Advanced Materials Research,2011,219-220:1325-1328.

[5] LIU Qi, ZHANG Hong-hui, HUANG Zhen-zhen. Analysis of oil drilling working state based on fuzzy clustering method(FCM)[J]. Journal of Zhoukou Normal University,2009,5:47-49(In Chinese).

[6] LIU Qi, DU Yuan-dong. The normal cloud model of fuzzy inference prediction method for petroleum drilling accident [C]. 2011 International Conference on Electric Information and Control Engineering,2011,Volume2:1816-1821.

[7] WANG Jie, WU Guo-zeng, ZHU Xiao-dong. Dynamic Processing of the Right Border of Real Time Logging Signal in Wavelet Analysis[J]. Journal of Zhengzhou University(Natural Science Edition),2010,3:84-89(In Chinese).

[8] Aminzadeh F.. Applications of AI and soft computing for challenging problems in the oil industry [J]. Journal of Petroleum Science and Engineering, 2005,47(1):5-14.

[9] Han P, Wang G.Y., Wang D.F.. "On the application of predictive function control in steam temperature system of thermal power plant". IEEE Proc-Control Theory, 2004, 148(6): 135-138

[10] Smith O J M. "A controller to overcome dead time".ISA, 1959,6(2):28-33.

[11] Venkatashankar V, Chidambaram M. "Design of P and PI controllers for unstable first-order plus time delay systems". INT. J. Control, 1994, 60(1): 137-144.

[12] ZHANG Jian-hai, ZHANG Sen-lin, LIU Mei-qin. "Robust stability analysis of delayed discrete-time standard neural network".Journal of Zhejiang

University: Engineering Science, 2009,43(8): 1383-1388 (In Chinese).
[13] Park J H, Sung S W, Lee I B. "An enhanced PID control strategy for unstable processes". Automatica,1998, 34(6):751-756.
[14] HAN An-tai, WANG Shu-qing. "Decentralized fuzzy control for a class of nonlinear interconnected large-scale systems with time-delay based on LMI approach".Control and Decision, 2004,19(4): 416-428 (In Chinese).
[15] YUAN Yu-hao, ZHANG Qing-ling, CHEN Bing. "Delay-dependent fuzzy control for nonlinear descriptor systems". Acta Automatica Sinica, 2006,32(5): 824-828 (In Chinese).
[16] SUNG S W, LEE I. "Limitations and countermeasures of PID controllers". Industrial & Engineering Chemistry Research, 1996, 35(8):2596-2610.
[17] KWAK H J, SUNG S W， LEE I B. "On-line Process identification and autotuning for integrating Processes". Industrial Engineering Chemistry Research, 1997, 36(12): 5329-5338.
[18] LUYBEN W L. "Tuning Proportional-Integral-Derivative Controllers for Integrator/Deadtime Processes". Industrial Engineering Chemistry Research, 1996, 35(10):3480-3483.
[19] WANG L, CLUETT W R. "Tuning PID controllers for integrating processes". IEE Proceedings-Control Theory & Applications, 1997, 144(5):385-392.
[20] SONG S H, CAI W J, WANG Y G."Auto-tuning of cascade control systems". ISA Transaction, 2003, 42(l):63-72.
[21] M. Mahlouji and A. Noruzi, "Human Iris Segmentation for Iris Recognition in Unconstrained Environments", IJCSI International Journal of Computer Science Issues, Vol. 9, No 3, 2012.
[22] Garcia C E, Morari M. "Internal model control 1. A, unifying review and some new results". Ind Eng Chem Process Des Dev, 1982, 21(2): 308-323.
[23] S. Nithyanandam, K. S. Gayathri, P. L. K. Priyadarsini, "A New IRIS Normalization Process For Recognition System With Cryptographic Techniques", IJCSI International Journal of Computer Science Issues, Vol. 8, No 4, 2011.

Computer Numerical Simulation of Stability Analyzing of Unsaturated Slope with Rainfall Infiltration

Wen Zhong
State Key Laboratory of High-Efficient Mining and Safety of Metal Mines (University of Science and Technology Beijing), Ministry of Education, Beijing 100083, China.
Email: zhongwenbj@yeah.net

Zhuoying Tan*(corresponding author)
State Key Laboratory of High-Efficient Mining and Safety of Metal Mines (University of Science and Technology Beijing), Ministry of Education, Beijing 100083, China.

***Abstract*—Rainfall infiltration is a large effect on the stability of soil slopes. Based on the actual rainfall data, saturated-unsaturated seepage theory and the mechanical theory of unsaturated porous media, numerical simulations were conducted to examine the seepage field of saturated-unsaturated slope during rainfall infiltration in this paper. The influence of the coefficient with rainfall duration, rainfall patterns, rainfall intensity and soil saturated permeability on seepage field were studied, and the relationship between the factor of safety and the position of sliding surface is obtained. The simulated results show that the saturated hydraulic conductivity has a large effect on slope stability. If the permeability of soils is relatively large, the changing range and velocity of the factor of safety increase with the rainfall intensity; and if it is relatively small, the influence of the rainfall intensity on the slope stability is also small.**

***Index Terms*—rainfall infiltration, numerical simulation, saturated-unsaturated seepage, stability analyzing**

I. Introduction

The application of the limit equilibrium analysis method and the seepage finite element approaches to the analysis of slope stability has been widely used for many years. Many numerical simulations with related software should be used for during the slope stability analysis. To investigate the influence of coupled numerical analysis on the stability of unsaturated soil slopes, the limit equilibrium analysis method and the seepage finite element approaches should be used.

Soil slope failures can broadly be attributed to the convergences of three factors, i.e. rainfall, steepness of slope, and soil geological profile [1]. Unsaturated soil slopes can remain stable for a long time and then fail under rainfall infiltration, and in many cases these failures cause loss of life and great economic losses [2]. Experience has shown that many slope failures occurred during or shortly after rainfall [3]. The rainfall infiltration produces a downward flux that changes the water content and pore-water pressure gradients with depth, hence reduces the soil shear strength and subsequently triggers the slope failure [4]. Rainfall infiltration and suction variation in unsaturated soils must be taken into consideration in the analysis of most slope stability problems, particularly in the southern rainy area where the annual precipitation is high [5].

Rainfall infiltration has long been a topic of interest in geotechnical engineering [6]. The process of rainfall infiltration into a soil slope is an extremely complex problem, involving numerous parameters such as rainfall duration [7], rainfall patterns [8], rainfall intensity [9], soil permeability, soil initial moisture condition, soil water retention ability, soil porosity, and evaporation rate etc[10]. And the problem becomes more complicated when dealing with the unsaturated soils because the hydraulic properties of the soils are strongly non-linear functions [11]. Many researchers suggested that conventional methods which were based on the assumption of saturated behavior cannot be applied successfully for slopes under unsaturated conditions [12]. Slope stability analysis of unsaturated slopes requires an extensive and detailed saturated-unsaturated transient seepage analysis, because slope failures in unsaturated conditions are closely related to rainfall infiltration [13]. Numerous researchers have worked on the analysis of rainfall-induced instability of saturated-unsaturated soil slopes and concluded that rainfall infiltration is an important factor triggering the instability of slopes [14]. And numerous studies have also been conducted on the infiltration characteristics of soils [15]. An efficient global optimization algorithm with unsaturated conditions is proposed for critical slip surface searching of slope stability analysis [16].

In this paper, based on the actual rainfall data, saturated-unsaturated seepage theory and the mechanical theory of unsaturated porous media, numerical simulations were conducted to examine the seepage field of saturated-unsaturated slope during rainfall infiltration. The influence of the coefficient with rainfall duration, rainfall patterns, rainfall intensity and soil saturated permeability on seepage field were studied, and the

relationship between the factor of safety and the position of sliding surface is obtained.

II. Optimization Theory and Method

A. Partial differential water flow equations

The general governing differential equation for two-dimensional seepage can be expressed as:

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial x}\left(k_x \frac{\partial H}{\partial x}\right) + \frac{\partial}{\partial y}\left(k_y \frac{\partial H}{\partial y}\right) + Q \quad (1)$$

where: H = the total head, k_x = the hydraulic conductivity in the x-direction, k_y = the hydraulic conductivity in the y-direction, Q = the applied boundary flux, θ = the volumetric water content, t = time.

This equation states that the difference between the flow (flux) entering and leaving an elemental volume at a point in time is equal to the change in storage of the soil systems. More fundamentally, it states that the sum of the rates of change of flows in the x-directions and y-directions plus the external applied flux is equal to the rate of change of the volumetric water content with respect to time.

Under steady-state conditions, the flux entering and leaving an elemental volume is the same at all times. The right side of the equation consequently vanishes and the equation reduces to:

$$\frac{\partial}{\partial x}\left(k_x \frac{\partial H}{\partial x}\right) + \frac{\partial}{\partial y}\left(k_y \frac{\partial H}{\partial y}\right) + Q = 0 \quad (2)$$

Changes in volumetric water content are dependent on changes in the stress state and the properties of the soil. The stress state for both saturated and unsaturated conditions can be described by two state variables [1]. These stress state variables are $(\sigma - u_a)$ and $(u_a - u_w)$ where is σ the total stress, u_a is the pore-air pressure, u_w and is the pore-water pressure.

The following derivation is formulated for conditions of constant total stress; that is, there is no loading or unloading of the soil mass. The following derivation also assumes that the pore-air pressure remains constant at atmospheric pressure during transient processes. This means that $(\sigma - u_a)$ remains constant and has no effect on the change in volumetric water content. Changes in volumetric water content are consequently dependent only on changes in the $(u_a - u_w)$ stress state variable, and with u_a remaining constant, the change in volumetric water content is a function only of pore-water pressure changes. As a result, the change in volumetric water content can be related to a change in pore-water pressure by the following equation

$$\partial \theta = m_w \partial u_w \quad (3)$$

where: m_w = the slope of the storage curve.

The total hydraulic head H is defined as:

$$H = \frac{u_w}{\gamma_w} + y \quad (4)$$

where: u_w = the pore-water pressure, γ_w = the unit weight of water, and y = the elevation.

Equation (4) can be rearranged as:

$$u_w = \gamma_w (H - y) \quad (5)$$

Substituting Equation (5) into Equation (3) gives the following equation:

$$\partial \theta = m_w \lambda_w \partial (H - y) \quad (6)$$

Which now can be substituted into Equation (1), leading to the following expression:

$$\frac{\partial}{\partial x}\left(k_x \frac{\partial H}{\partial x}\right) + \frac{\partial}{\partial y}\left(k_y \frac{\partial H}{\partial y}\right) + Q = m_w \gamma_w \frac{\partial (H - y)}{\partial t} \quad (7)$$

Since the elevation is a constant, the derivative of y with respect to time disappears, leaving the following governing differential equation used in finite element formulation:

$$\frac{\partial}{\partial x}\left(k_x \frac{\partial H}{\partial x}\right) + \frac{\partial}{\partial y}\left(k_y \frac{\partial H}{\partial y}\right) + Q = m_w \gamma_w \frac{\partial H}{\partial t} \quad (8)$$

B. Finite element water flow equations

Applying the Galerkin method of weighed residual to the governing differential equation, the finite element for two-dimensional seepage equation can be derived as:

$$\begin{aligned} &\tau \int_A ([B]^T [C][B]) dA\{H\} + \tau \int_A (\lambda < N >^T < N >) dA\{H\}, \\ &t = q\tau \int_L (< N >^T) dL \end{aligned} \quad (9)$$

where: $[B]$ = the gradient matrix, $[C]$ = the element hydraulic conductivity matrix, $\{H\}$ = the vector of nodal heads, $< N >$ = the vector of interpolating function, q = the unit flux across the edge of an element, τ = the thickness of an element, t = time, λ = storage term for a transient seepage equals to $m_w \gamma_w$, A = a designation for summation over the area of an element, and L = a designation for summation over the edge of an element.

In an axisymmetric analysis, the equivalent element thickness is the circumferential distance at different radius, R about the symmetric axis. The complete circumferential distance is 2π radian times R, since it is formulated for one radian, the equivalent thickness is R. Therefore, the finite element equation for the axisymmetric case is:

$$\begin{aligned} &\int_A ([B]^T [C][B] R) dA\{H\} + \int_A (\lambda < N >^T < N > R) dA\{H\}, \\ &t = q \int_L (< N >^T R) dL \end{aligned} \quad (10)$$

Note that the radial distance R is not a constant within an element as in the case of the thickness τ in the two-dimensional analysis; consequently, R is a variable inside the integral.

In an abbreviated form, the finite element seepage equation can be expressed as:

$$[K]\{H\} + [M]\{H\}, t = \{Q\} \quad (11)$$

where: $[K]$ =the element characteristic matrix, $[M]$ the element mass matrix, $\{Q\}$ the element applied flux vector

Equation (12) is the general finite element equation for a transient seepage analysis. For a steady-state analysis, the head is not a function of time and, consequently, the term $\{H\}$, t vanishes, reducing the finite element equation to:

$$[K]\{H\} = \{Q\} \tag{12}$$

which is the abbreviated finite element form of the fundamental seepage equation, Darcy's Law.

C. Temporal integration

The finite element solution for a transient analysis is a function of time as indicated by the $\{H\}$, t term in the finite element equation. The time integration can be performed by a finite difference approximation scheme. Writing the finite element equation in terms of finite differences leads to the following equation:

$$\begin{aligned}&\left(w\Delta t[K]+[M]\right)\{H_1\} = \\ &\Delta t\left((1-\omega)\{Q_0\}+\omega\{Q_1\}\right)=\left([M]-(1-\omega)\Delta t[K]\right)\{H_0\}\end{aligned} \tag{13}$$

where: t = the time increment, ω = a ratio between 0 and 1, H_1 = the head at end of time increment, H_0 = the head at start of time increment, Q_1 = the nodal flux at end of time increment, and Q_0 = the nodal flux at start of time increment.

We use the Backward Difference Method, a method that sets to 1.0, the finite element equation is then simplified to:

$$\left(\Delta t[K]+[M]\right)\{H_1\} = \Delta t\{Q_1\} = [M]\{H_0\} \tag{14}$$

As indicated by Equation (15) in order to solve for the new head at the end of the time increment, it is necessary to know the head at the start of the increment. Stated in general terms, the initial conditions must be known in order to perform a transient analysis.

D. Field variable model

To formulate a finite element analysis, it is necessary to adopt a model for the distribution of the field variable within the element. Since the field variable in the seepage analysis is the total head (H) , it is necessary to adopt a model for the distribution of H within the element.

We assume that the head distribution within the element follows the adopted interpolating functions. This means that the head distribution is linear when the secondary nodes are missing, and the head distribution is nonlinear when the secondary nodes are present.

In equation form, the head distribution model is:

$$h =< N > \{H\} \tag{15}$$

where: h = the head at any local coordinate, $< N >$ = a vector of interpolation function, and $\{H\}$ = a vector of heads at the nodes.

III. RAINFALL INFILTRATION METHODS

A. Partial differential water flow equations

It is not especially difficult to obtain a direct measurement of a volumetric water content function in a laboratory, but it does require time and it requires finding a geotechnical laboratory that performs the service. It is, however, standard practice to obtain a grain-size distribution curve and many companies have the capability and facilities to develop their own curves. The development of the grain-size distribution curve is inexpensive and can be quickly accomplished.

One of the required input parameters for a transient analysis is the volumetric water content function. Since it can sometimes be difficult or time consuming to obtain a volumetric water content function, it may be of benefit to be able to develop an estimation of the volumetric water content function using either a closed-form solution that requires user-specified curve-fitting parameters, or to use a predictive method that uses a measured grain-size distribution curve.

Aubertin et al (2003) presented a method to predict the volumetric water content function which is modified from the method proposed by Kovacs (1981). The modifications were made to Kovacs method to better represent materials such as tailings from hard-rock mines. A further modification extended the method for clay type soils. The Aubertin et al. method predicts the volumetric water content function using basic material properties which can be useful, particularly for preliminary analysis. It should be cautioned that, especially for clay type materials, it is critical to base a final design on measured material properties as opposed to estimated ones.

The function is initially determined as a degree of saturation function and then is later converted to a volumetric water content function. The function is developed by defining the degree of saturation for two main components. The first component contributes to the amount of water that is stored in a soil by capillary forces that exist at relatively small negative pore-water pressures. The second component contributes to the volumetric water content function at large negative pore-water pressures where the amount of water that exists in the soil is primarily a function of adhesion. Both of these components can be evaluated from the negative pore-water pressure and material property information such as particle-size, the shape of the particles and the porosity.

The degree of saturation as determined based on the capillary and adhesive components is as follows:

$$S_r = \frac{\Theta_w}{n} = S_c + S_a^*\left(1 - S_c\right) \tag{16}$$

where: S_r = the degree of saturation, Θ_w = the volumetric water content, n = the porosity, S_c = the degree of saturation due to capillary forces, and S_a^* = the bounded degree of saturation due to adhesion (S_a).

and where: $$S_a^* = \left(1 - S_a\right) + 1 \tag{17}$$

The adhesive component is a bounded value since it is possible at low suctions for the value S_a to be greater than 1. The bounded value ensures that for a S_a greater or equal to 1, S_a^* = 1 and if S_a is less than 1, then $S_a^* = S_a$.

The adhesion component is associated with the thin film of water that covers the surface of the soil grain and

depends on basic material properties such as the negative pore-water pressure in the soil and the particle-size, shape coefficient and porosity of the soil. It is determined by the following equation:

$$S_a = a\frac{\left(h_{co}/\psi_n\right)^{2/3}}{\left(\psi/\psi_n\right)^{1/6} e^{1/3}} C_\psi \tag{18}$$

where: a = a curve fitting parameter, ψ = the suction, ψ_n = a suction term introduced to ensure dimensionless component, e = the void ratio, h_{co} = the mean capillary rise (cm) determined for capillary soils by:

$$h_{co} = \frac{h(cm^2)}{eD_{10}(cm)} \text{ or } h_{co} = \frac{\xi w_L^{1.75}}{e} \tag{19}$$

For cohesion type soils where: D_{10} = the particle diameter (cm) corresponding to 10% passing on a grain size curve,

$$b(cm^2) \text{ is given by: } b(cm^2) = \frac{0.75}{1.17\log C_u + 1} \tag{20}$$

where: C_u = the coefficient of uniformity, w_L =the liquid limit (%), ξ = a constant approximately equal to 402.2 C_ψ = a correction coefficient that allows a progressive decrease in water content at high suctions, forcing the function through a water content of zero at one million kPa suction as initially proposed by Fredlund and Xing (1994) and described by:

$$C_\psi = 1 - \frac{\ln\left(1+\frac{\psi}{\psi_r}\right)}{\ln\left(1+\frac{\psi_o}{\psi_r}\right)} \tag{21}$$

where: ψ_r = the suction corresponding to the residual water content at which point an increase in suction will not effectively remove more liquid water from the soil and given by:

$$\psi_r = 0.86\left(\frac{\xi}{e}\right)^{1.2} w_L^{1.74} \tag{22}$$

The capillary saturation, which depends essentially on the pore diameter and the pore size distribution, is given by:

$$S_c = 1 - \left[\left(\frac{h_{co}}{\psi}\right)^2 + 1\right]^m \exp\left[-m\left(\frac{h_{co}}{\psi}\right)^2\right] \tag{23}$$

where: m = a fitting parameter that takes into account the pore size distribution and controls the shape and position of the volumetric water content function in the capillary zone.

For plastic-cohesive soils considered here, both the value of parameters m and a can be taken as constants with $m = 3\times10^{-5}$ and $a = 7\times10^{-4}$ in the predictive applications. For the capillary based soils, m and a can be taken as 1 and 0.01 respectively.

B. Rainfall Model

After rainfall, the distribution of the water content in the soil of the slope internal changes gradually over time (Fig.1) due to the gradual infiltration, and causes the wetting front move down constantly (Fig.2).

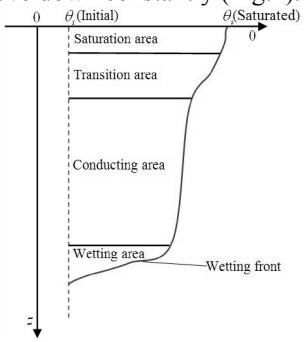


Figure 1. Distribution and partition with rainfall infiltration

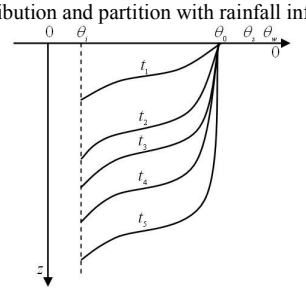


Figure 2. Rainfall infiltration -time history

- The water impinges upon the ground in a very short period of time when it rainfall, the water content in surface soil ($\theta(0,t)$) increases initial value to maximum value (θ_0) in a relatively short time. With the generally natural conditions, fully saturated state of soil is impossible, so the maximum value (θ_0) is below saturated water content θ_s .
- With gradual infiltration, wetting front will continue to move down to the inside of slope, the distribution curve of the water content from relatively steep gradually into the relatively flat.
- The absolute value of water content gradient ($\frac{\partial\theta}{\partial z}$) what is from large to small in surface soil. The $\frac{\partial\theta}{\partial z}\to 0$ when t is large enough, and that means the water content constant of near the surface is the same.

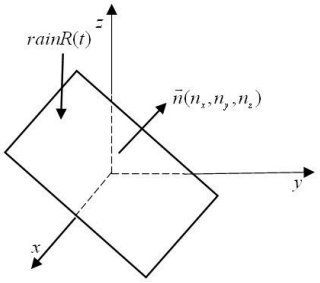


Figure 3. Schematic plot with rainfall infiltration of slope

The outgoing line direction on the surface of soil slope is $\vec{n}(n_x, n_y, n_z)$, If the $R(t)$ is vertical rainfall intensity, the component of rainfall on the slope surface normal direction is:

$$q_n(t) = R(t)n_z \tag{24}$$

According to the Darcy's law, the maximum infiltration capacity of the slope in all directions is:

$$R_j(t) = -k_j(h_w)\frac{\partial(h_w + z)}{\partial x_j} \tag{25}$$

Transformed into normal direction of the infiltration rate is:

$$R_n(t) = R_j(t)n_j \tag{26}$$

For slope, the actual infiltration flow is $q_s(t)$, and perpendicular to the slope direction, based on the foregoing analysis, we can deduce the relationship between rainfall intensity and actual infiltration.

$$q_n(t) \le R(t)n_z, q_s(t) = q_n(t) \tag{27}$$

$$q_n(t) > R(t)n_z, q_s(t) = R_n(t) \tag{28}$$

IV. Engineering Practice Analysis

A. Engineering Overview

The typical waste dump slope in southern rainy area is selected in this study object (Fig.4), and the rainfall of nearly five years in the region is shown in Table I. Slope material properties are three parts that divided into preliminary compacted soil, loose soil and bedrock. Three material parameters of calculation as shown in table II. The Water and soil characteristic curve of material ① as shown in Fig.5. The Water and soil characteristic curve of material ② as shown in Fig.6.

Mesh model of slope was divided into 16909 nodes, 16909 units (Fig.7). The model boundary above the underground water level is zero flow; the boundary below the groundwater level is the given head, and the slope bottom is impervious boundary. The soil slope surface is the infiltration border. This example takes the soil-water conditions after four steps for the initial conditions of the simulation in next phase. Through Once every other stability analyzing after rainfall seepage finite element analysis, find the changing range and velocity of the factor of safety increase with the rainfall intensity.

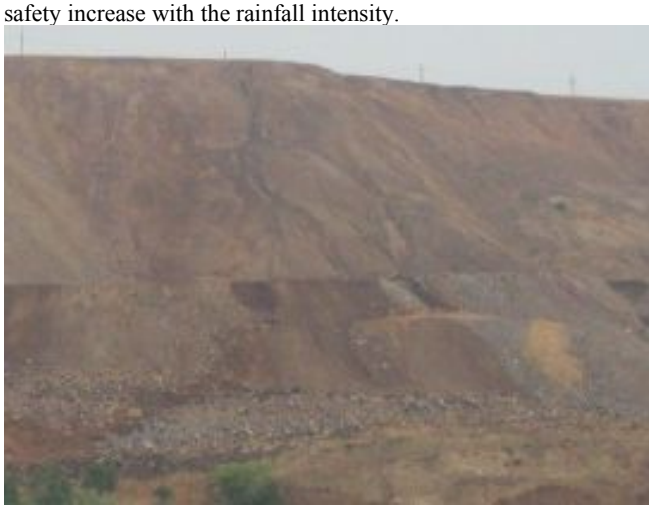


Figure 4. Location shooting of slope

B. Simulation Results

TABLE I.
NEARLY FIVE YEARS OF RAINFALL MONITORING

Rainfall	2006	2007	2008	2009	2010
Jan.	117.90	54.60	191.40	46.00	145.5
Feb.	17.40	71.10	80.20	23.20	73.1
Mar.	266.80	39.40	197.60	162.10	106.8
Apr.	102.10	151.60	148.10	350.90	126.3
May.	106.10	163.80	208.00	252.60	191.5
Jun.	264.60	195.40	429.90	719.60	132.8
Jul.	259.30	190.00	364.40	114.70	46.4
Aug.	166.90	207.10	89.70	275.90	94.8
Sep.	24.80	27.90	86.00	71.70	79.1
Oct.	93.80	28.30	76.70	31.90	93.6
Nov.	51.00	196.10	6.10	26.60	83.9
Dec.	14.50	127.20	19.60	6.70	22.9
Yearly	1485.2	1452.5	1897.7	2123.8	1196.7

TABLE II.
PARAMETERS OF CALCULATION

Material	Coefficient of permeability (cm/s)	Percentage of saturated water rate (%)	Residual water rate (%)	Natural bulk density γ'(kN/m^3)	Saturated bulk density γ_{sat}(kN/m^3)
Material ①	3.37e-002	25	10	17.6	17.9
Material ②	5.72e-002	40	20	16.9	17.5
Material ③	3.20e-005	5.2	5.2	30.6	30.6

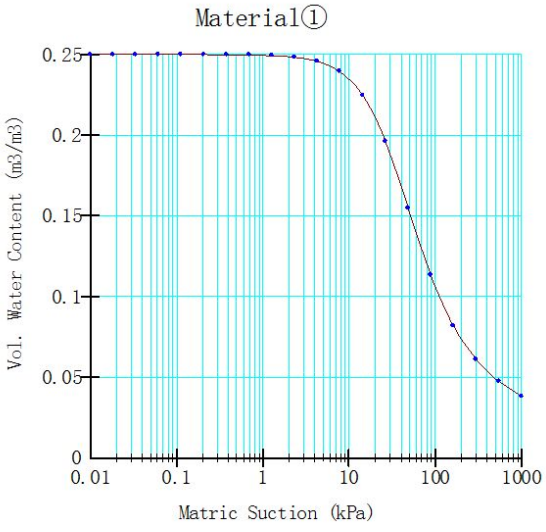


Figure 5. Water and soil characteristic curve

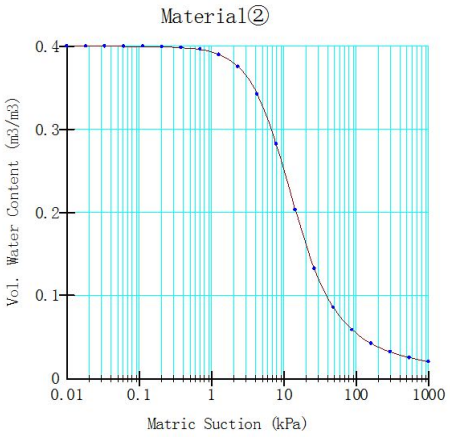


Figure 6. Water and soil characteristic curve

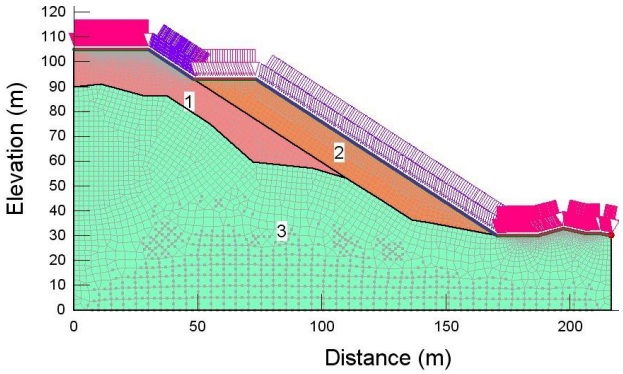


Figure 7. Simplified calculation model of the seepage finite element

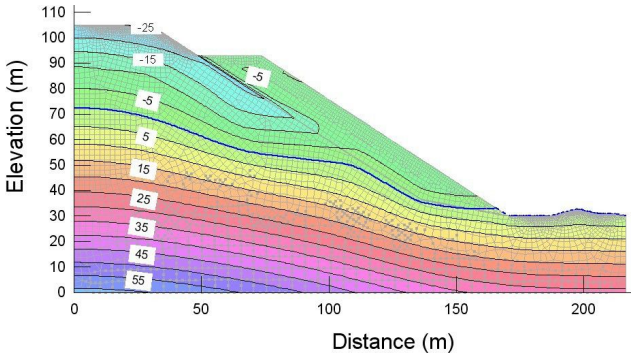


Figure 8. Simulation steady state pressure head contours

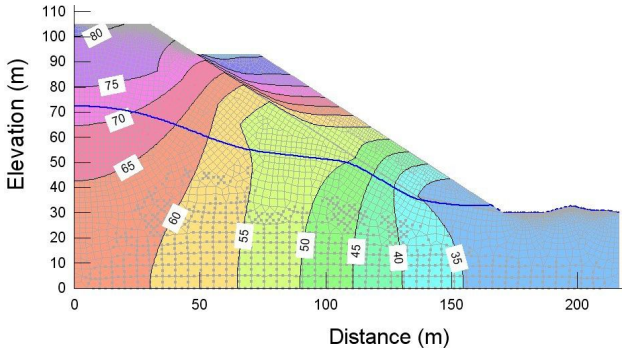


Figure 9. Simulation steady state total head contours

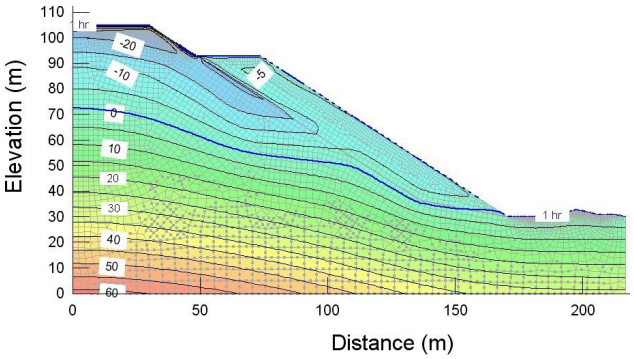


Figure 10. Simulation transient step1 pressure head contours

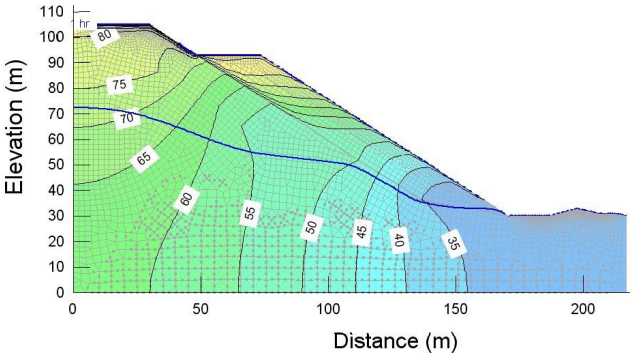


Figure 11. Simulation transient step1 total head contours

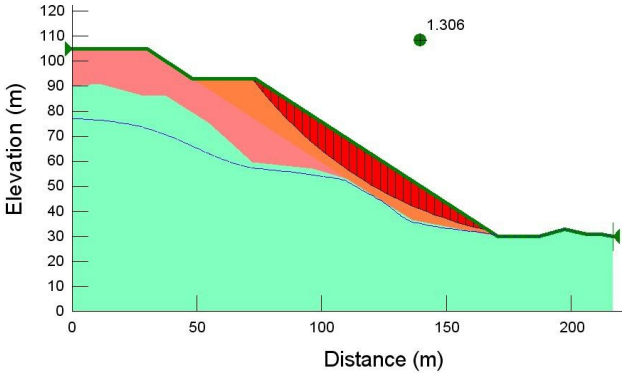


Figure 12. Simulation slope stability step1

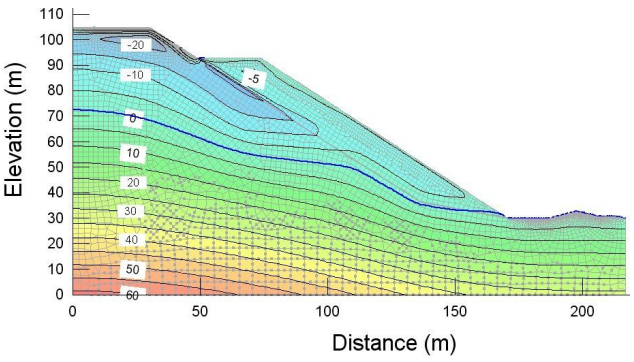


Figure 13. Simulation transient step2 pressure head contours

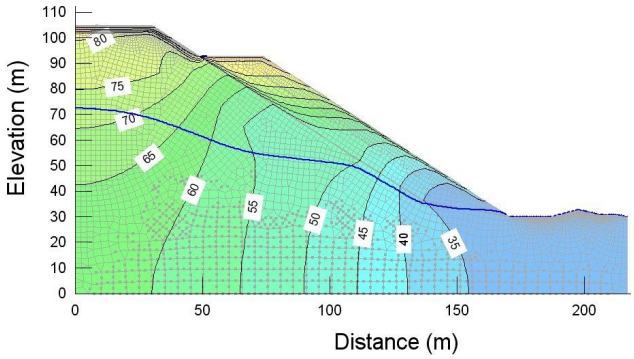


Figure 14. Simulation transient step2 total head contours

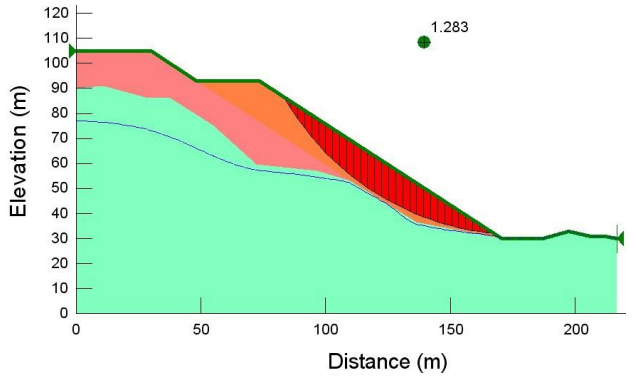


Figure 15. Simulation slope stability step2

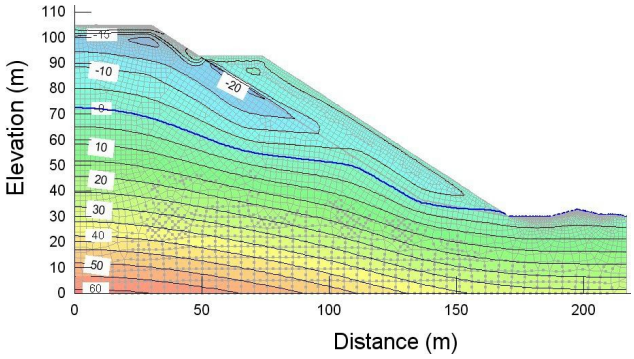


Figure 16. Simulation transient step3 pressure head contours

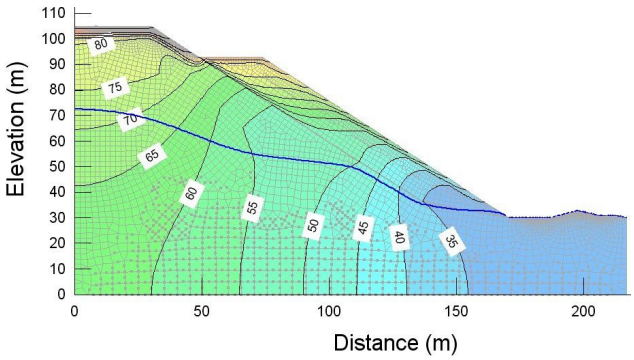


Figure 17. Simulation transient step3 total head contours

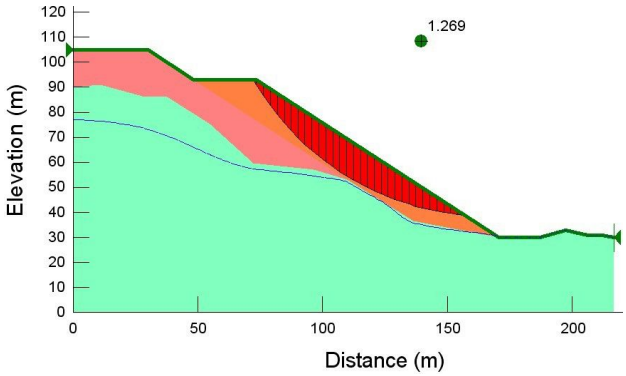


Figure 18. Simulation slope stability step3

C. *Discussion*

The relationship between rainfall intensity and saturated permeability also has a great influence on the variation of pore-water pressure and factor of safety. The presented model is reasonable in reflecting the actual behavior of slope under rainfall infiltration，and that the maximum wetting deformation occurs at both the corner of slope surface and the top of slope.

V. COCLUSION

Through the slope seepage and the stability analysis over these examples under the different rainfall time, we can conclude: the coefficient with rainfall duration, rainfall patterns, rainfall intensity and soil saturated permeability on seepage field were studied, and the relationship between the factor of safety and the position of sliding surface is obtained. The simulated results show that the saturated hydraulic conductivity has a large effect on slope stability. If the permeability of soils is relatively large, the changing range and velocity of the factor of safety increase with the rainfall intensity; and if it is relatively small, the influence of the rainfall intensity on the slope stability is also small.

ACKNOWLEDGMENT

This work was partly supported by the Key Program of National Natural Science foundation of China (No. 51034001) and the National Key Basic R & D Program of China (No.2010CB731501).

REFERENCES

[1] R. H. Chen, H. P. Chen, and K. S. Chen, “Simulation of a slope failure induced by rainfall infiltration,” Environ Geol, vol. 58, pp. 943–952, April 2009.

[2] D. G. Fredlund, and A. Xing, “Equations for the soil–water characteristic curve,” Can Geotech., vol. 31, pp. 521–532, May 1994.

[3] F. Kang, J. J. Li, “Artificial bee colony algorithm with local search for numerical optimization,” Journal of Software, vol. 6, pp. 490–952, March 2011.

[4] S. You, G. F. Zhao, and H. G. Ji, “Model for transversely isotropic materials based on distinct lattice spring model (DLSM),” Journal of Computers, vol. 6, pp. 1139–1144, June 2011.

[5] S. J. Li, Z. C. Shangguan, and J. Z. Wang, "Computer Simulation of sequential impoundment process of concrete–faced rock fill dam," Journal of Computers, vol. 7, pp. 1801–1808, August 2012.
[6] M. D. Fredlund, D. G. Fredlund, and S. L. Houston, "Assessment of unsaturated soil properties for seepage modeling through tailings and mine wastes," Proc., 10th Int. Conf. on Tailings and Mine Wastes Geol, vail, Colo., pp. 149–157, June 2003.
[7] D. G. Fredlund and M. D. Fredlund, "Numerical modeling for saturated-unsaturated soil systems," Short Course Notes, vail, Colo., pp. 149–157, June 2003.
[8] D. G. Fredlund, "The 1999 R. M. Hardy Lecture: The implementation of unsaturated soil mechanics into geotechnical engineering, R. M. Hardy Address," Can Geotech., vol. 37, pp. 963–986, March 2000.
[9] E. C. Leong, L. He, and H. Rahardjo, "Factors affecting the filter paper method for total and matric suction measurements," Can Geotech., vol. 25, pp. 321–332, April 2002.
[10] D. Tami, H. Rahardjo, and E. C. Leong, "Effects of hysteresis on steady-state infiltration in unsaturated slopes." J. Geotech. Geoenviron. Eng., vol. 130(9), 956–967, March 2004.
[11] K. Gavin and J. F. Xue, " A simple method to analyze infiltration into unsaturated soil slopes " Computers and Geotechnics, vol. 35 pp.223–230, July 2008.
[12] Y. Ma, S. Y. Feng， " Modeling water infiltration in a large layered soil column with a modified Green–Ampt model and HYDRUS–1D," Computer an Electronics in Agricuture, vol. 71 pp.40–47, August 2010
[13] E. C. Leong, L. He, H. Rahardjo and D. G. Fredlund, "A comparative study of constitutive models for unsaturated soils," Proc., 2nd Asian Conf. on Unsaturated Soils, Osaka, pp. 41–46, June 2003.
[14] N. T. M. Thieu, M. D. Fredlund, D. G. Fredlund, and H. Q. Vu, "Seepage modeling in a saturated–unsaturated soil system Assessment of unsaturated soil properties for seepage modeling through tailings and mine wastes," Proc., Int. Conf. on Management of the land and Water Resources, Hanoi, Vietnam, pp. 49–56, May 2001.
[15] H. T. V. Pham, and D. G. Fredlund, "The application of dynamic programming to slope stability analysis," Can Geotech., vol. 40, pp. 830–847, June 2003.
[16] H. T. V. Pham, D. G. Fredlund, and G. Jr. Gitirana, "Slope stability analysis using dynamic programming combined with finite element stress analysis," Proc., Int. Conf. on Management of the land and Water Resources, Hanoi, Vietnam, pp. 107–114, May 2001.

Wen Zhong was born in China in 1983, and he is now major in University of Science and Technology Beijing for his Ph.D. Her research interests are slope deformation analysis, slope stability evaluation, theory of unloading rock mass for joint rock mass and its application, and construction management of hydroelectric engineering.

Research on Private Cloud Storage Solutions for Meteorological Applications

Liu Xuelin
Henan Institite of Medical Device Testing, Zhengzhou, Henan, China

Xu Junfeng
Henan Institite of Medical Device Testing, Zhengzhou, Henan, China

Bai Jiefang
Henan Institite of Medical Device Testing, Zhengzhou, Henan, China

***Abstract*—In order to solve the problem of the frequent meteorological observation data-write-in and data storage, a private cloud storage model designed consists of four layers: user access interface layer, meta-data storage layer, entity data storage layer and a relational database layer. Adopting Hadoop Distributed File System implements the meteorological data storage as well as efficient query and search business; and using Relational database realizes frequent data writing of data collection extremity. Results from experiments illustrate that this system not only has achieved massive meteorological data storage, query, attribute management but overcome the shortcomings of single insertion inefficiency as well. It meets the requirements for the data storage and management of meteorological departments.**

***Index Terms*—Big Data, Cloud Storage, Distributed File System, Relational Database, Column oriented Database**

I. Introduction

Meteorological observation data are originally from the ground and airborne meteorological observatories. It is more than 2610 observatories that daily get various meteorological data whose volume is up to 300-500M and at least 150G data must be processed. Doppler radar probes each 6 minutes and gets the 20MB observation data once, the daily 4.5GB data and the yearly amount of data over 1.5TB. The central meteorological station receives materials up to PB magnitude, the data for operational application up to TB magnitude, and meteorological observation data increasing exponentially [1]. At present, these data mainly, stored in the form of files or database, are applied to (1) data analysis forming the information for weather forecast or disaster forecast etc. and providing decision support for other departments, and to (2) storage as materials furnishing data support for other researches.

Cloud computing, implementing convenient access, fast allocation, and releasing shared resource pool through the Internet, is a computing model developed from parallel computing, distributed computing and grid computing. It results from the complex gradual progress and leap of the conceptions including virtualization, utility computing, IaaS(Infrastructure as a Service), PaaS(Platform as a Service), SaaS(Software as a Service) etc[2-3]. Cloud storage, depending on the functions of cluster applications, grid technologies, distributed file systems, is a system to organize a large number of various storage devices in network through application software integration to cooperate and jointly provide data storage and business access [4-5].This paper mainly introduced how to use cloud storage to store and manage meteorological observation data.

II The Meteorological Private Cloud Storage Model

The main problems of meteorological data storage and management

Nowadays, the problems exist in meteorological data storage and management.

- Storage management. With the development of science and technology, meteorological attributes detected by human beings and the amount of data are constantly growing, so it is necessary to dynamically extend the storage capacity and attributes of meteorological database.
- Speed of response. The collected meteorological data are real-time ones that should be saved in time. While querying, making detailed statistical analysis and reading in massive meteorological data, users should be responded timely.
- Stability and safety of data Meteorological data are precious materials for science research and historical ones. You should guarantee the stability and safety in storage, and at the same time make full use of available computing devices and resources.
- Convenient operation and maintenance . Meteorological data are original materials. Effective handling can provide convenience for the following business dealing. Servers are centrally placed in order to facilitate maintenance.

S *Meteorological private cloud storage model*

Meteorological private cloud storage architecture model.

According to the above problems and requirements for data storage the meteorological departments proposed, we designed a private cloud storage model as shown in the Fig.1.The model consists of four layers: user access interface layer, meta-data storage layer, entity data storage layer and relational database layer [6-10].

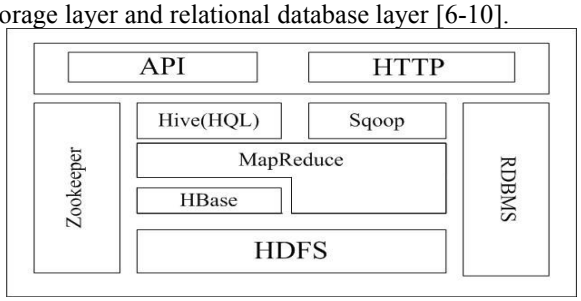


Figure 1.Hadoop-based model architecture of the meteorological private cloud storage.

- Entity data storage layer. HDFS, whose attributes are good expansibility, high reliability and low price, is able to store massive files on common clusters at a low cost [6-8]. Nevertheless, as a distributed file system on the ground platform of meteorological private cloud storage, it was adapted to administrate and store large files, so HBase was selected to be database. HDFS, a ground data storage container, realize the function of storing data. As HBase does not support SQL, Hive supporting Hsql is chosen a tool of data warehouse and convenient to query, administrate and operate database by API in the late development.
- Meta-data storage layer. The layer mainly achieves to manage the attributes of meteorological table in HBase and Oracle. By the way of administrating meta-data tables, you can implement the management of mapping on the attributes of meteorological tables.
- Relational database layer. Oracle was selected to store and manage real-time data and user data of cloud storage platform. Long respond time was a disadvantage of HDFS. The meteorological data was real-time data and amount of data each time was small. Single data insertion was inefficient, so you should save them into fast-response relational databases. Sqoop, used as data transmission and migration, periodically transfers the meteorological data from relational databases to HBase. This method can improve the system's storage efficiency
- User access interface layer. This layer provided HTTP and API interfaces to support access to WEB service and other service developments of meteorological cloud application.

III EXPERIMENTAL DESIGN

Hadoop-based cloud storage scheme

According to the above analysis, the overall structure could be designed for data acquisition module, the results show module, query and analysis module, memory module and data migration module[9-12].

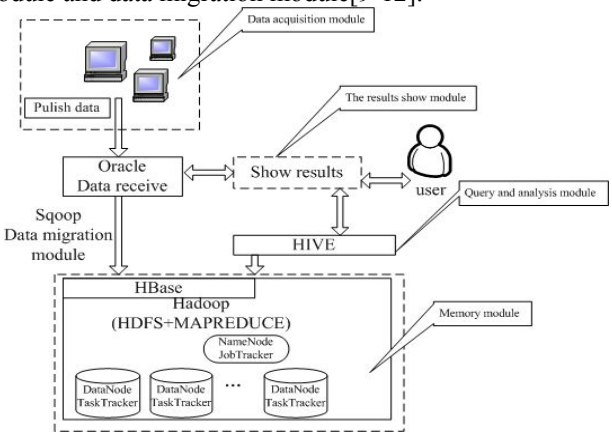


Figure 2.The overall structure of the private cloud storage platform.

Data migration module. This module used Sqoop to transfer the data in Oracle to HBase regularly and automatically.

- data acquisition module. The module supports visual data input and data input API, manual publish of meteorological data, inserting data acquisition devices and obtaining data. The received data would be stored in Oracle.
- Memory module. The module implemented by HBase and HDFS is in charge of the storage and backup of Meta data and entity data. Data stored in HDFS is not limited by data types and kept as any data type.
- Query and analysis module. HSQL is supported and convenient to query. This module provides APIs which are about outer data query and management.
- The results show module. This module shows the query results and the structure of distributed file system.
- Data migration module. This module used Sqoop to transfer the data in Oracle to HBase regularly and automatically.

Data model design

Data model includes meteorological data model and meta-data model. They are concretely meteorological tables in HBase and meta-data tables. One column family in HBase tables may be made up of one or many columns [13-15].

TABLE I.
METEOROLOGICAL HBASE DATABASE TABLE

Row key	Time stamp	Column Family				
		Column1	Column2		Column3	
Time+ Position	t	Time	Position	Attribute 1	Attribute 2	Attribute 3

Meteorological user query statistics takes time and station as query conditions and the value combination of time and position as unique value, Row key to convenient

to query. Compared the integration of Hive and HBase with query from Hive or HBase, query's response time was much longer, but it is very efficient to regard the combination as Row key.

TABLE II.
META-DATABASE TABLE

Field Name	Type	Not Null	Column Family
Time	Int	Yes	Column1
Position	Int	Yes	Column1
Attribute1	String	Yes	Column2

Attribute of meteorological table in HBase and Oralce was stored in meta-data table. The operation on meta-data tables would implement the management of mapping to corresponding meteorological database. The metadata table's fields include field names of meteorological attributes, null or not null and types which are corresponded in relational database. The metadata tables HBase corresponds to require field names of meteorological attributes, null or not null, and the fields of column clusters they belong to.

Ⅳ. THE EXPERIMENTAL RESULTS AND DATA ANALYSIS

Establish data sets

Meteorological data table in relational database

The materials the experiment utilizes are ground surface observational data. They include eight attributes: position, max_temperature,time, avg_temperature, min_temperature, avg_humidity, atmospheric_pressure, max_temperature, avg_water_vapor_pressure.

TABLE III.
METEOROLOGICAL DATA TABLE

Field Name	Type	Primary	Not Null
attribute	varchar	Yes	Yes
avg_temperature	varchar	No	Yes
avg_humidity	varchar	No	Yes
avg_water_vapor_pressure	varchar	No	Yes
atmospheric pressure	varchar	No	Yes
max_temperature	varchar	No	Yes
min_temperature	varchar	No	Yes

Time and position are combined into attribute, whose value is unique, so it is defined as primary key.

Table in Hbase

TABLE IV.
HBASE DATABASE TABLE

Row key	Time stamp	Column Family						
		temperature			pressure		humidity	
ATTRIBUTE	T	AT	MINT	MAXT	AVP	AP	AH	AH

As shown in table Ⅳ, AT is avg_temperature; MAX is max_temperature; MINT is min_temperature; AVP is avg_water_vapor_pressure; AP is atmospheric_pressure; AH is avg_humidity.

Attribute is regarded as Row key of the table. Timestamp is a 64-bit integer and is automatically assigned while written into HBase. Temperature, pressure, humidity are three column clusters. Each column cluster consists of several columns. Temperature includes three columns: AT, MAXT, MINT. Pressure includes two columns: AVP, AP. Humidity only includes AH.

Analysis of experiment results

In order to test the performance of this platform, we built an experimental environment using 4 PCs. The system configuration parameters are shown in table Ⅴ-Ⅵ. Machines are connected by switchboards through 100MB Ethernet network. Master with the name NameNode and Hmaster are running on PC1; DataNode1, DataNode2, HRegionServer1 and HRegionServer2 are running on PC2 and PC3; Cloud storage system, including Relational Database and servers are running on PC4.

TABLE V.
HARDWARE PARAMETER TABLE

Name	Quantity	Specification
PC	4	CPU(Dual-core)1.8G,Memory 2G

TABLE VI.
SOFTWARE PARAMETER TABLE

Software	Version
Operation System	Ubuntn 12.0
JDK	1.6.0_30
Hadoop	Hadoop 1.0.1
Hive	0.9.0
HBase	0.92.0
Relational Database	Oracle 10g

The data used in the experiment are meteorological data from 194 surface stations from 1951 to 2011. Seven different datasets are cut out from the experiments, 1.3GB, 2.5GB, 5GB, 8GB, 10GB, 13GB, and 25GB.

The evaluation target is system response time, the range, whose unit is second, from when the system sent out to query or download dataset command from the client ends to storage models, storage models operate data to when the result are fed back to the clients. We have designed four experimental schemes: (1) query a record, (2) query one million records, (3) query and count one million records and (4) query and download one million records.

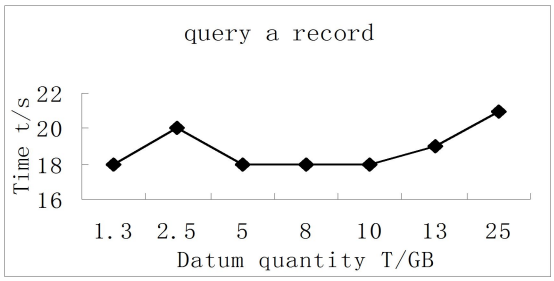


Figure 3.query a record

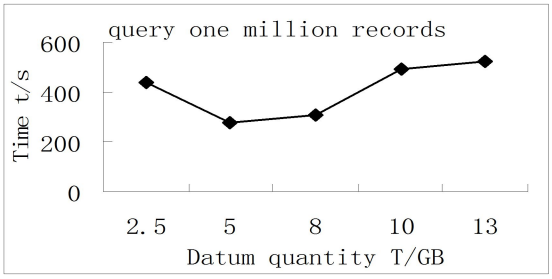


Figure 4.query one million record

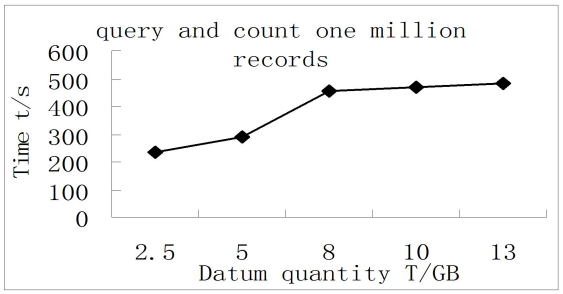


Figure 5.query and count one million record

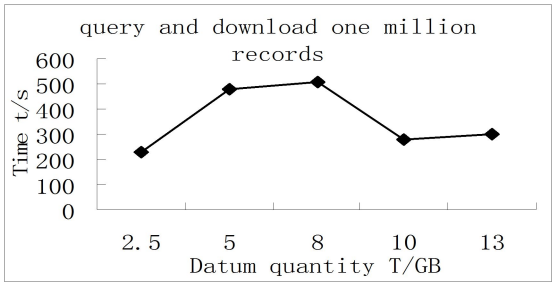


Figure 6.query and download one million record

Fig.3 and Fig.4 show that the system response time decreases significantly when data are queried and the datasets are greater than 5G. Fig.5 shows that the system response time's increasing rate obviously decreases when data are queried and counted and the datasets are greater than 8G. Fig.6 shows that the system response time decreases significantly when data are queried and downloaded and the datasets are greater than 8G. Therefore, we can make a conclusion that this system has obvious advantages in dealing with large datasets.

Ⅴ Conclusion And Future Work

This paper describes how to design and develop a meteorological private cloud storage system. Based on Hadoop distributed architecture and HDFS distributed file system, we combined distributed database Hbase, the data warehouse management tool Hive, and the data migration tool Sqoop to build, develop the private cloud storage system and implement mass meteorological data storage, rapid data insertion, highly efficient query and download, and probed meteorological attribute management etc. Experiments indicate that the platform has extendibility, easy maintainability and high efficient management of mass meteorological data.

At present, this platform basically realized the expected target, but it is not perfect enough and continues to be improved on the basis of the present. The stored meteorological data on the platform currently do not contain nonstructural data such as weather satellite images, Because meteorological data are relatively precious and their security is significant. So it is the system that still needs to be researched on the respect of security.

References

[1] Dou yiwen,Lu li,Liu xulin etc. "Meteorological Data Storage and Management System." *Computer Systems & Applications*, 7rd ed. vol. 20, 2011, pp. 116-120.

[2] Peter Mell, Timothy Grance. "The NIST Definition of Cloud Computing." In *NIST Special Publication,*2011, pp. 800-145.

[3] Yu si,Gui xiaolin,Huang ruwei etc. "Improving the Storage Efficiency of Small Files in Cloud Storage." J*OURNAL OF XI'AN JIAOTONG UNIVERSITY*, 6rd ed. vol. 45, 2011, pp. 59-63.

[4] Dong jiguang, Chen weiwei, Tian langjun, Wu haijia. "Replia placement study in large-scale cloud storage system." *Journal of Computer Applications*, 3rd ed. vol. 32, 2012, pp. 620-624

[5] Nicolae B, Moise D, Antoniu G BlobSeer: "Bringing high throughput under heavy concurrency to Hadoop Map-Reduce applications," *IEEE International Symposium on Parallel & Distributed Processing (IPDPS)*, pp 1-11,2011

[6] Cui jie, Li taoshen, Lan hongxing. "Design and Development of the Mass Data Storage Platform Based on Hadoop." *Computer Research and Development*, 2012(S1), pp. 12-18.

[7] Zhou ke, Wang hua, Li chunhua. "Cloud Storage Technology and Its Application." *ZTE Communications*, 4rd ed. vol. 16, 2010, pp. 24-27.

[8] Frank Doelitzscher, Anthony Sulistio, Christoph Reich, Hendrik Kuijs, David Wolf.Private. "cloud for collaboration and e-Learning services:from IaaS to SaaS."

Springer-Verlay New York, Inc.New York, pp 23-42, January 2011.

[9] Chang F,Dean J, Ghemawat S, Hsieh WC, Wallach DA. etc. “Bigtable :A distributed storage system for structured data. ” Proc.of the 7th USENIX Symp. on Operating Systems Design and Implementation. pp 205-218, 2006.

[10] Dean, J. and Ghemawat, S.” MapReduce: Simplified data processing on large cluters.” Proceedings of the 6th Symposium on Operating Systems Design and Implementation, San Francisco, 2004.

[11] Zhang chunming,Rui jianwu,He tingting. “An Approach For Storing And Accessing Small Files On Hadoop.” *Computer Applications and Software.* 11rd ed. vol. 29, 2012, pp 95-100.

[12] Zhang jianxun, Gu zhiming, Zheng chao. ”Survey of research progress on cloud computing.” *.Application Research of Computers*. 2rd ed. vol. 27, 2010, pp 429-433.

[13] Sanjay Ghemawat, Howard Gobioff, Shun-Tak “Leung.The Google File System.” Proc.of the 19th ACM Symp. on Operating Systems Principles.pp 29-43, 2003.

[14] Yang zhihao, Zhao taiyin, Yao xingmiao, etc.”Private cloud computing system realization method adaptable to data and computing intensive tasks.” *Application Research of Computers*. pp 29-43,2003.

Journal of Simulation
(JSM, ISSN: 2310-4791)
CALL FOR PAPERS

Journal of Simulation (JSM, ISSN: 2310-4791) is a scholarly peer-reviewed international scientific journal published bimonthly for researchers, developers, technical managers, and educators in the simulation field. It provide a high profile, leading edge forum for academic researchers, industrial professionals,engineers, consultants, managers, educators and policy makers working in the field to contribute and disseminate innovative new work on all the areas of simulation.

JSM invites original, previously unpublished, research, survey and tutorial papers, plus case studies and short research notes, on both applied and theoretical aspects of simulation.

Area of Interests

- Simulation
- Economics Simulation
- Management Simulation
- Model
- Operating Systems
- Software Systems
- Communication Protocols
- Interconnection Networks
- Design
- Prototyping
- Testing Methods and Tools
- Artificial Intelligence
- Algorithms
- Computational Science
- New and Important Applications and Trends

Manuscript Submission

- JSM invites original, previously unpublished, research papers, review, survey and tutorial papers, application papers, plus case studies, short research notes and letters, on both applied and theoretical aspects.
- Submission implies that the manuscript has not been published previously, and is not currently submitted for publication elsewhere. Submission also implies that the corresponding author has consent of all authors.
- Upon acceptance for publication transfer of copyright will be made to Academic Publishing House, author will be requested to pay a mandatory publication charge. Article submission implies author agreement with this policy.
- Whenever applicable, submissions must include the following elements: title, authors, affiliations, contacts, abstract, index terms, introduction, main text, conclusions, appendixes, acknowledgement, references, and biographies.

Publication Charges and Orders

All the journals published by Academic Publishing House are 'Open Access' journals, which means that the publication cost should be covered by the author's institution or research funds. The author's company or institution will be requested to pay a flat publication fee of EUR 200 for an accepted manuscript regardless of the length of the paper. The publication charges are mandatory. Authors can order hard copies of the issue with a price of EUR 20 per copy.