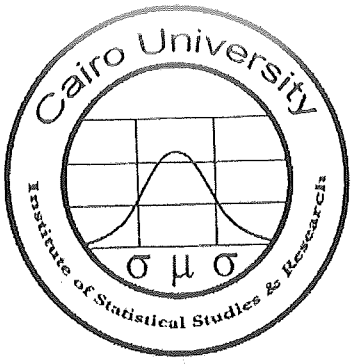


Cairo University
Institute of Statistical Studies and Research

The 49th Annual Conference on Statistics Computer Sciences and Operation Research

Volume
Computer Sciences

22-25, Dec . 2014

Index
Computer Sciences

Multimodal Biometric ATM

[1]Ghada Abdelhady, [2]Hussam Elbehiery

Abstract

Nowadays using debit cards has become a way of life for many Egyptians, whether you're withdrawing cash from an ATM, or buying anything from any stores, no doubt that your debit card is a simple way to access your money. ATM card and password theft is a common problem these days and most of banks depend on passwords to guarantee the security of the debit cards. Unfortunately, this is not enough to guarantee the security. So there is an urgent need to an integrated cryptosystem to save our information in the bank account and all required transactions. Biometrics is a technology that strives towards identifying or authenticating the identity of a living person based on biological key that make the process definitive and effortless. Biometric ATM's are the latest inventions to help us avoid fraud and duplication. This paper will present a multimodal biometric cryptosystem that has been applied on ATM simulator. This research uses TDES and DES-EC to accomplish the encryption process. It uses three biometric features to change the security levels.

Keywords — *AES; ATM; Biometrics; DES; DES-EC; Elliptic Curves; TDES; Symmetric encryption.*

I. INTRODUCTION

Biometrics is a technology that strives towards identifying or authenticating the identity of a living person based on biological key that make the process definitive and effortless. A person's live biometric data is being matched against a bio print in the database, such as a smart card. If it matches, it means that the person is who he or she claims to be and access is granted. This process is called verification. Using the password in one to many systems is checking against a full database of passwords. This process is called identification. The system only cares if the password is valid one, not if the person using the password is authorized to use it.

Biometric Encryption is a best method to secure our data and information. It guarantees the identification and authentication of the encryption process. To implement this technology, a powerful mathematical is used which known as Ciphering [4].

[1] Assistant Professor: Information Systems dept., Higher Institute of Computer science and Information Systems, Culture and Science City, Egypt. ghada.abdelhady10@gmail.com

[2] Assistant Professor: Information Systems dept., Higher Institute of Computer science and Information Systems, Culture and Science City, Egypt. hussam.elbehiery@gmail.com

Ciphering algorithms can be divided into two categories: private-key (symmetric) and public-key (asymmetric). Private-key systems use a common private key shared between the communicating parties, while public-key do not require any key exchange [4]. The implementation of ciphering systems presents several requirements and challenges. First, the performance of the algorithms is often crucial. One needs encryption algorithms to run at the transmission rates of the communication links. Slow running cryptographic algorithms translate into consumer dissatisfaction and inconvenience. On the other hand, fast running encryption might mean high product costs since traditionally, higher speeds were achieved through custom hardware devices. In addition to performance requirements, guaranteeing security is a formidable challenge.

Our study will present a new algorithm that has been added to the known symmetric encryption algorithms [1]. This system is called New DES based on Elliptic curve. It has the benefits of Data Encryption Standard, DES and also the benefits of Elliptic Curve Cryptosystem (ECC) as a public key algorithm. It has the simplicity of DES algorithm. Also, it has the strength of ECC. Except that the key in ECC is a public key but the key in our new algorithm is still private and generated by a new technique mentioned in details in [8]. Also we will use different keys as we are going to use three biometric scanners, Fingerprint, retina and iris scanners.

Biometric Encryption is the most suitable method to secure our data and information, in which the security is forced to check the incoming objects by scanning of our body and the scanned objects are stored in database as an image. Then, the stored images are compared with the original image that already stored in database [12].

The coming sections explain History and state of the art and how DES is modified and transformed into DES-EC to be secured. Also, we explain the flowchart of the proposed system and the suggested prototype pictures. The last section demonstrates industry and market analysis including the comparison between Multimodal Biometric ATM using EC and Biometric ATM using Triple Data Encryption Standard, TDES which is currently used over world.

II. BIOMETRIC CRYPTOSYSTEM OVERVIEW

Let us go to one of the symmetric-key algorithms as DES, TDES, AES and recently, DES-EC (New DES based on Elliptic Curve) used in our study [1]. Most of them are easy and fast in hardware implementation and also they are so flexible in software. Thus, software-based systems would seem to be a better fit because of their flexibility [3]. Fortunately, many embedded processors combine the flexibility of software on general purpose computers with the near-hardware speed and better physical security than general purpose computers. Embedded processors are already an integral part of many communications devices and their importance will continue to increase. If we combine this with their flexibility to be programmed and their ability to perform arithmetic operations at moderate speeds, it is easy to see that they are a very promising platform to implement cryptographic algorithms [5] and [14].

The theory of operation of DES-EC was appeared for the first time in [1]. DES-EC tried to bring back DES to life using elliptic curves which are considered the simplest possible curves after lines and conics. Elliptic curves over finite fields provide an inexhaustible supply of finite abelian groups. Such curves involve elementary arithmetic operations that make it easy to implement (in either hardware or software) [6]. They are generally more secure than others are. Elliptic curves could be applied to DES to improve its performance and make it hard to be broken. This paper presents a proposed cryptosystem using three symmetric algorithms. The first algorithm is TDES that is commonly used in ATM, the second is DES-EC using the Fiestel function and S-boxes like DES but each stage in this algorithm will be based on elliptic curve (EC) [1] as explained in the following section.

The word biometric refers to technologies that utilize specific human characteristics (specifically, biological features that are unique to each individual) to establish the identity of a person, or gain access to secure areas. There are many types or devices for biometrics including: Fingerprint Scanner, Retinal Scanner, Facial Recognition, Voice Recognition, and Keystroke Recognition [12].

Security based on biometrics is commonly used to limit access to only certain individuals. This is useful for safeguarding information in government and corporate environments. Biometrics can also be used to identify people who do not wish to be detected. Biometric recognition or authentication refers to the automated method of verifying a match between two human biometric like fingerprints.

III. DES-EC: NEW DES BASED ON ELLIPTIC CURVE

As we mentioned that the new system "DES-EC" uses the Fiestel function and S-boxes like DES but each stage in this algorithm will be based on elliptic curve (EC), so before we present the new algorithm, we will explain the regular DES with its stages in the following section.

The security requirements, such as authentication, confidentiality and integrity, always make computationally intensive processes and can easily become the bottleneck of the related applications. The research paper [14] has explained an upgraded implementation of DES on the recent ATMEL, Microcontroller Corporation; ATMEGA32. Elliptic curve (EC) has proved to be a good choice for building asymmetric encryption system. So, the research paper [14] showed the implementation of another new symmetric system: "New DES based on EC" on an embedded system. The new system will be based on EC properties [14].

A. DES Encryption Stages

DES is a block cipher, meaning it operates on plaintext blocks of a given size (64-bits) and returns cipher-text blocks of the same size. Each block of 64 bits is divided into two blocks of 32 bits each, a left half block L and a right half R. The operation of the DES can be described in the stages of DES encryption algorithm shown in Fig. 1 [1].

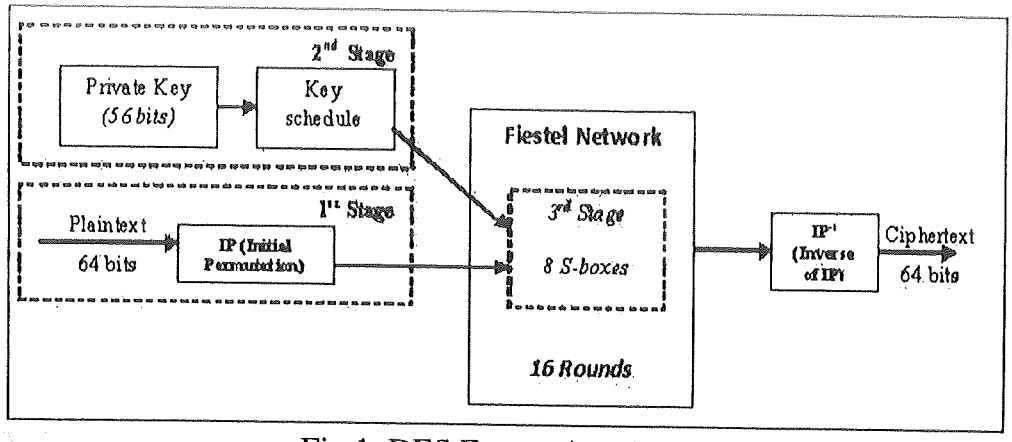


Fig.1. DES Encryption Stages.

B. Elliptic Curves (EC)

All the EC's used in this study defined on a prime finite field Fp that is popular for EC used in cryptography [5]. The simple form of EC equation defined in F_p is shown in equation (1).

$$(y^2) \bmod p = (x^3 + ax + b) \bmod p \qquad (1)$$

where *a, b* are constants and p is the order of the prime finite field F_p .Note that the right-hand side is a special cubic polynomial. The usage of elliptic curve in cryptography also requires that the curve be non-singular. Geometrically, this means that the graph has no cusps or self-intersections. Algebraically, this involves calculating the discriminant shown in equation (2) in order to prevent repeated roots on the right-hand side [6]. Equation (2) states that the curve is non-singular if and only if the discriminant is not equal to zero.

$$(4a^3 + 27b^2) \bmod p \neq 0 \qquad (2)$$

Despite their simple form, elliptic curves have been studied for many years and have many significant applications in mathematics. Maybe one of the most interesting results related to the application of elliptic curves is that they are used to prove Fermat's Last Theorem [3] and [6]. Beside that they are the corner stone of Elliptic Curve Cryptography, ECC. ECC is particularly beneficial for application where:

- Computational power is limited (wireless devices, PC cards)
- Integrated circuit space is limited (wireless devices, PC cards)
- High speed is required [3].
- The properties of elliptic curves, defined on Fp, to generate points used in encryption have been presented in [11].

C. DES-EC stages and Performance Evaluation

DES-EC depends on three stages: the plaintext masking stage, new key schedule stage and New S-boxes, as shown in Fig. 2.

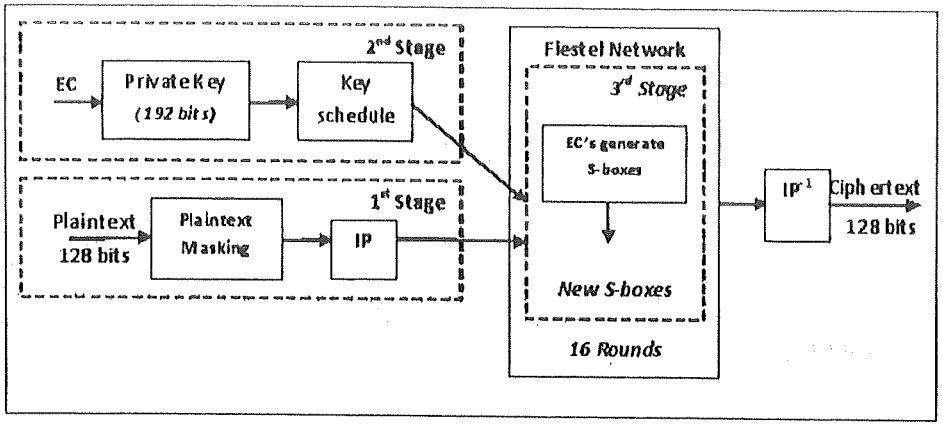


Fig.2. DES-EC Encryption Stages.

We notice in Fig. 2 that each stage will be based on EC or groups of EC's. Since the new algorithm is symmetric like regular DES, the structure presented in Fig. 2 is shared by the encryption and the decryption algorithms but in opposite directions. Meaning that the third stage represents the first stage after applying the inverse permutation and the first stage represents the last stage that includes the initial permutation and the inverse of the plaintext masking. The details of each stage are explained in [7], [8] and [14].

D. TDES

Using TDES is the same operation as DES but with 3 different keys. For example if we have key1, key2, key3. First the TDES algorithm encrypts the data with key1, then decrypts the output cipher with key2, but since key2 is not equal to key1, then it will scramble the cipher further. After that, it will take the output cipher of the second step and encrypts it with key3. This will be the final output cipher of the TDES.

Deciphering will be the reverse of the encryption algorithm which is first decrypting with key3, then encrypting with key2, after that decrypting with key. All these processes are shown in TDES Flowchart in Fig. 3 [2] and [3].

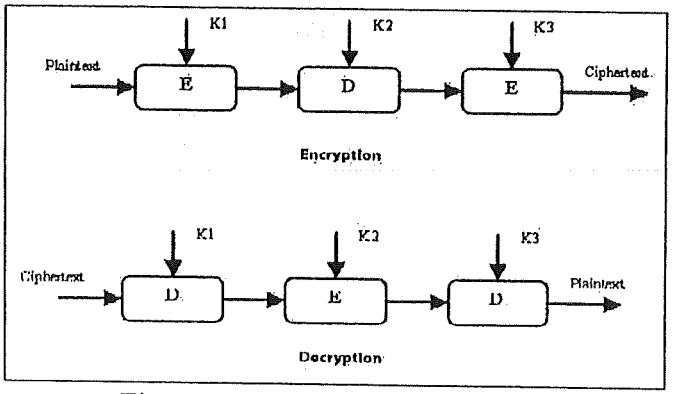


Fig.3. DES-EC Encryption Stages.

The suggested algorithm uses the elliptic curves into the encryption and decryption stages of DES and TDES as well to get secured system comparing to the current algorithms and also easy in hardware implementation as we proved in [14].

E. Performance Evaluation

As AES is the most widely used symmetric cipher today, so we present a theoretical comparison between the proposed algorithm and AES.

The measurements that have been used for the comparison are the key size, time and space complexity and execution time of encryption/decryption schemes. The encryption time is considered the time that an encryption algorithm takes to produce a cipher text from a plaintext.

Encryption time is used to calculate the throughput of an encryption scheme. It indicates the speed of encryption. The throughput of the encryption scheme is calculated as the total plaintext in byte encrypted divided by the encryption time. Table 1 shows the comparison between AES and the proposed algorithm according to the mentioned measurements.

TABLE 1. AES VS. NEW ALGORITHM

	AES	Proposed Algorithm
Key size (bits)	192	192
Block size (bits)	*128*	*128*
[a] Time complexity	*O (n log n)*	*O (n log n)*
[a] Space Complexity	*O (log n)*	*O (log n)*
No. of rounds	12	16
Strength	secured	Secured
Execution time sec.	Encryption: 0.05	Encryption: 0.04
	Decryption: 0.03	Decryption: 0.02
Speed (byte/sec)	Encryption: 2560	Encryption: 3200
	Decryption: 4266.67	Decryption: 6400

[a] *n* is the size of the inputs in bits

As we noticed from table 1 that the speed of the suggested algorithm is higher than the speed of AES although AES has the minimum number of rounds and our algorithm is secured like AES.

Average time required for exhaustive key search for DES, AES and the proposed algorithm, is shown in Table 2 [1].

TABLE 1. AVERAGE TIME REQUIRED FOR EXHAUSTIVE KEY SEARCH

Key size (bits)	No. of Alternative keys	Time required at 1 decryption/μs	Time required at 10^6 decryption/μs
56 (DES)	$2^{56}=7.2*10^{16}$	2^{55}μs=1142 years	10.01 hours
128 (AES)	$2^{128}=3.4*10^{38}$	2^{127}μs=$5.4*10^{24}$ years	$5.4*10^{18}$ years
192 (New DES)	$2^{192}=6.3*10^{57}$	2^{191}μs=10^{44} years	10^{38} years

IV. BIOMETRICS SCANNERS

There is not one biometric modality that is best for all implementations. Many factors must be taken into account when implementing a biometric device including location, security risks, task (identification or verification), expected number of users, user circumstances, existing data, etc [12].

V. MULTIMODAL BIOMETRIC ATM

TDES is the ciphering algorithm that is applied in ATM over world to encrypt the customers' accounts. The block diagram of Biometric ATM which is applied for one scanner (fingerprint) that is commonly used is shown in Fig. 4. The block diagram is considered as the proposed block schematic of crypto biometric authentication system using fingerprint scanner. It will be repeated for the other scanners (Iris and Retina) used in the suggested work.

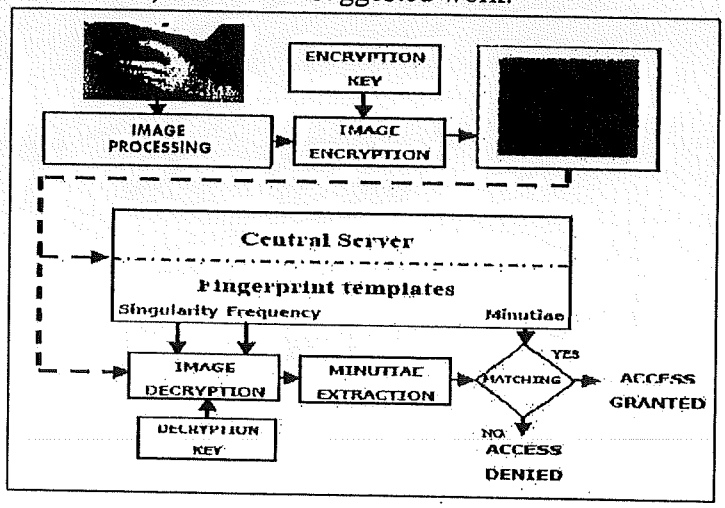


Fig.4. Schematic of embedded crypto biometric authentication system

Using biometric measures will enhance the security of ATM banking. As PIN-based ATM, machines are more secure than Biometric-based machines from the point of view of resistance to coercion crimes, Biometrics are used side by side

with PINs and the system is able to detect and block emulation attacks. In our study, the encryption key will be used as a pass code for the first stage of the authentication as shown in the previous figure.

In our country, it is urgent to use more secured ATM using the unique features as the biometrics that it is not easy to be attacked. In addition, the proposed system classifies the security levels according to the required amount and its percentage from the daily withdrawal limit of each customer. The details of the proposed system are explained in the following Section.

A. Multimodal Biometric ATM Prototype

The main target in our study is the achievement of the security. Also get a prototype for a multimodal biometric ATM with multi-level of security. Security is not accomplished by encryption alone. While the choice of encryption algorithms and key management routines are critical, they are usually not the weak link in a secure application, Methodology and Execution Plan.

By the end of this section, we will have a prototype simulated Multi level Biometric ATM using integrated secured cryptosystem and three biometric scanners. We can conclude our prototype details in the following block diagram shown in Fig. 5

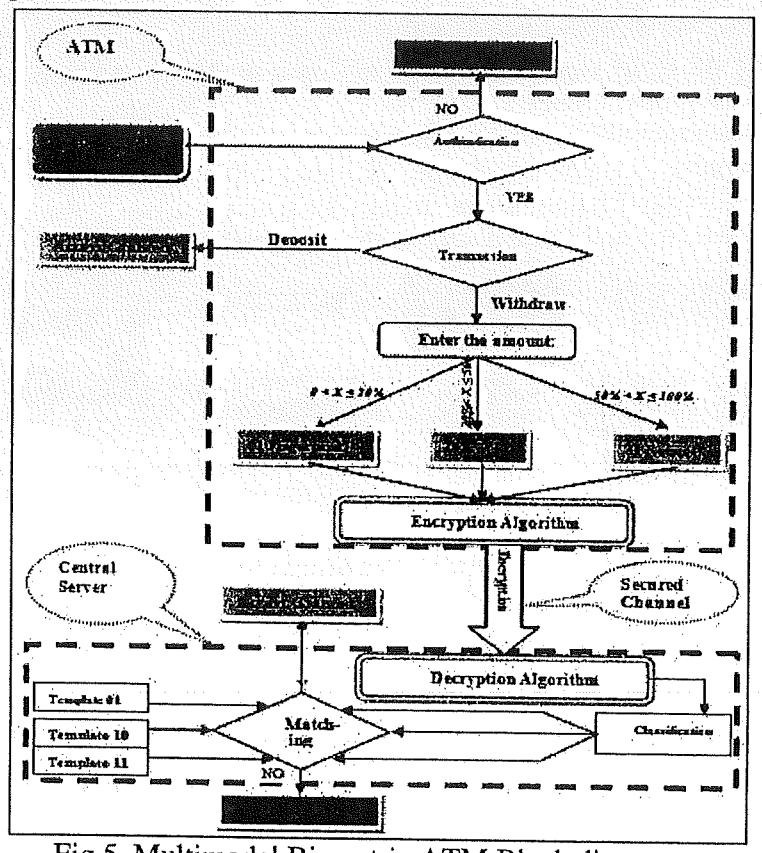


Fig.5. Multimodal Biometric ATM Block diagram

In this system, we should have two expected states (Deposit and withdraw) to complete the transaction process. The user will Scan three biometric scanners according to the required amount in withdraw state, Fig. 6 and 7 present a simple graphical user interface to simulate the biometric ATM and the central server bank respectively.

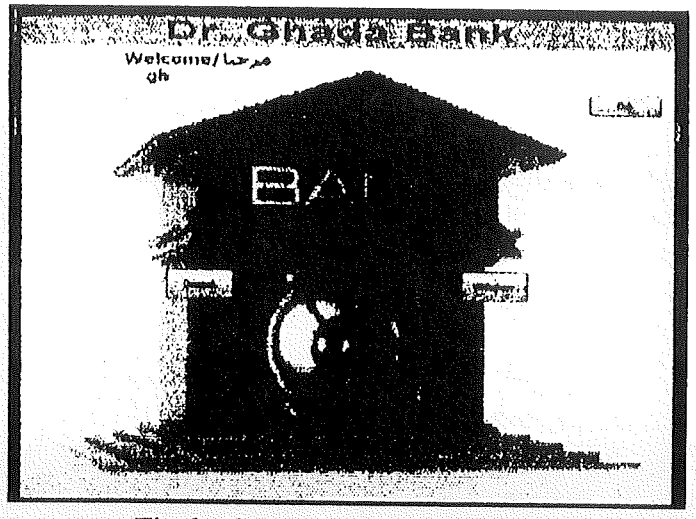


Fig.6. Biometric ATM Simulation

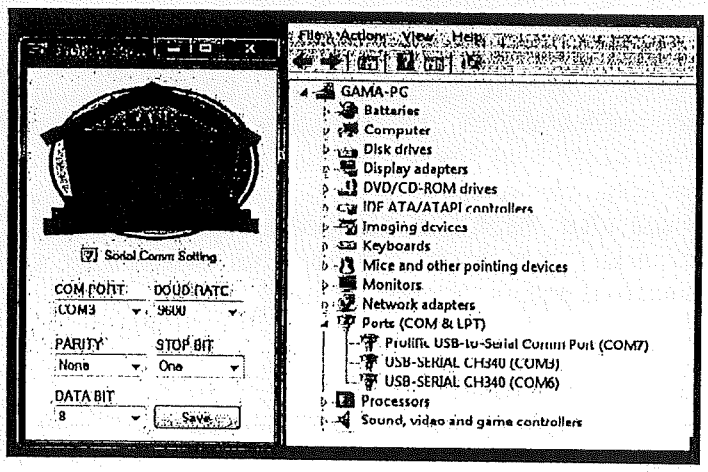


Fig.7.Central bank server simulation

Fig. 6 includes the simulation of the ATM which is basically a simple graphical user interface and a simple database using text files. It is basically having an excel sheet stored in the source location named as the server. The source location could represent the bank database, which in reality is implemented in a more complex way.

The process includes first creating the user account from the administrative mode which saves the customer's data including his/her name, key, biometrics, credit and telephone number. All these data are stored in the excel sheet in the central bank which is considered the server shown in Fig. 7.

After that, in order for a customer to access the data and make a transaction, customer has to enter her/ his passcode and then the algorithm will encrypt his data and send it across secured channel to the server.

The encrypted data will be decrypted inside the server bank to be compared and if it is matched with the stored data, the customer will get the access to make the required transaction. Customer can then start a transaction in which he will type the amount of money he wants and then the algorithm will check the required percentage to ask him for his biometrics according to the required percentage explained in the flowchart Fig. 5. Also the algorithm will encrypt the biometrics and again it will be sent via secured channel for matching process and if the access granted the algorithm will modify the stored cipher file with the new credit balance.

VI. Industry and Market Analysis

Most of ATM over world uses TDES to encrypt the information of banking accounts. Although AES is more secured than TDES, TDES is used in ATM Banking, that is because the real time elapsed in TDES to encrypt or decrypt data is less than the time elapsed by AES.

In the proposed system, we impose new secured system to encrypt the banking data and the transactions according to the required level of security. Also in this study, we have three scanners for identifications rather than the one biometric scanner that is almost using the fingerprint. Our system uses three Biometric scanners applied during the transaction processes according to the required percentage of the total amount for the customer. Beside that the proposed system includes acknowledgment process to send a message to the customer cell phone before and after the transaction process to inform his/her the details of the transaction to guarantee the security from different aspects. Table 3 shows the market analysis and its strategy:

Table 2. Market analysis

	ATM in Market	**Suggested Prototype**
Used Encryption Algorithm	TDES	TDES and DES-EC
Security level	Almost one security level	Three security levels
Acknowledgement message	Only after the transaction has been processed	Before and after the transaction process
Biometric scanners used	Fingerprint is commonly used	Fingerprint, iris and retina

The multimodal biometrics ATM presented in our proposal will be the first over world according to table 4 and our research.

VII. Conclusion

Debit card fraud happens when a thief "skims" or swipes the information off the magnetic stripe on the back of your card to create a duplicate copy of your card. They also have to capture your PIN to access your account and withdraw money or make purchases. Debit card fraud can also happen if your card is lost or stolen and you haven't taken steps to protect your PIN.

This research introduced a multimodal biometric cryptosystem based on elliptic curve to get a prototype of biometric ATM. The proposed system combines three different biometrics, fingerprints, retinas and irises, which are considered as the most three accurate biometrics, with using PIN that is already used. By using different biometrics, the administrator can decide the required security level for each transaction, deposit/draft process for customers, Building database of the three different biometrics templates used (Fingerprint, Iris and Retina).

Using biometric features as a form of verification and security is the perfect solution for the debit card fraud problem. Since biometric features are impossible to steal.

A simple database simulation carrying the cipher text of the customer's data, which is secured since in order to be broken, requires the key which is the biometric image, as well as the complex algorithm implemented. All cases and options that can be happened during any transaction have been studied and simulated via the biometric ATM simulator and the central bank server simulator

Finally, we present a simple graphical user interface simulating the transaction process and all its requirements related to the customer account via a secured channel in Biometric ATM using integrated biometric cryptosystem.

Acknowledgment

I would like to acknowledge the technical support of this research at Higher Institute of Computer Science and Information Systems.

References

[1] A. Ghada, "New DES based on elliptic curve: DES-EC ", LAP LAMBERT Academic Publishing GmbH & Co. KG, ISBN (978-3-845-40245-1), July 2011.

[2] A. Ghada and Hussam Elbehiery " Multimodal biometrics cryptosystem using elliptic curve", LAP LAMBERT Academic Publishing GmbH & Co. KG, ISBN (978-3-659-60797-4), 6th Oct., 2014.

[3] S. William, "Cryptography and network security principles and practices", Fourth Edition, Prentice Hall, November 16, 2005.

[4] L. Knudsen, "Block ciphers- analysis, design and applications", Ph.D. thesis, DAIMI PB-485, Aarhus University, Denamark, 1994.

[5] W. Mao "Modern Cryptography theory and practice", Prentice Hall PTR, July 25, 2003.

[6] P. Christof and J. Pelzl, "Understanding of cryptography ", A Textbook for Students and Practitioners, Springer-Verlag Berlin Heidelberg 2010.

[7] N. Torri and K. Yokoyama, 'Elliptic curve cryptosystems', Fujtsu Sci. Tech. J, December 2002, pp. 140-146.

[8] A. Ghada, F. Helail and A. Elkouny, "DES enhancement using elliptic curve equation", CNIR journal, May 2010.

[9] W. Thomas, J. Guajardo, and C. Paar, "Cryptography in embedded systems: an overview", Department of Electrical Engineering and Information Sciences Communication Security Group (COSY), Ruhr-Universit Äat Bochum, Bochum, Germany, Proceedings of the Embedded World 2003 Exhibition and Conference, Design & Elektronik, Nuernberg, Germany, February 18-20, 2003, pp. 735-744.

[10] B. Hagai, "Security implications of hardware vs. software cryptographic modules", White Paper, Discretix Technologies Ltd., Information Security Analyst, October 2002.

[11] B. Sandro, P. Bennati, R. Giorgi and E. Martinelli, "Elliptic curve cryptography support for ARM based embedded systems", Dept. Ingegneria de ll' Informazione, University of Siena, Via Roma 56, 53100 Siena, Italy.

[12] L. Avanindra and S. Dutta, "ECC based biometric encryption for network security", Journal of Computing, volume 3, issue 6, June 2011, issn 2151-9617.

[13] A. Illiasaak, N. Thamrin and M. Hani, "Crypto embedded system for electronic document", Regional Postgraduate Conference on Engineering and Science (RPCES 2006), Johore, 26-27 July.

[14] A. Ghada and Hussam Elbehiery, "Implementation of new symmetric ciphering on ATmega 32, 6th International Computer Engineering Conference, ICENCO 2010, Cairo, December 27-28, 2010.

The Proposed Emergency Reactive Controller (ERC) Architecture for TAntNet Traffic Routing of Road Network

Ayman M. Ghazy[1], Hesham A. Hefny[1]

Abstract. Dynamic traffic routing algorithms plays an important role in avoiding congestion and directing vehicles to better routes. Threshold based AntNet (TAntNet) algorithm utilize the pre known information of road network to effectively directing vehicles to good route. TAntNet routing system does not take in account the emergencies situations. In some cases special locations of network suffer from emergencies situations and need to be closed and prevent vehicle from crossing this locations, this can happens when an accident occurs or any other emergency situation happen. The existing routing system for road network still needs more modification to work efficiency at different circumstances and emergencies situations. This paper presents a modified architecture of routing system of road network, which include an embedded emergencies controller. Emergencies controller gives the administrator (at the control center) the ability to control the routing system, such that it avoids routing through the emergencies locations.

Keywords: Emergency, Control, Reactive, Swarm Intelligence, Road networks, Dynamic traffic routing, AntNet, TAntNet, Forward ant, Backward ant, Check ant, ERC, Emergencies Reactive Controller.

1 Introduction

Traffic congestion is a serious problem in most modern cities. Nowadays traffic jams and congestion on roads become one of the most difficult problems that face every one in every day. Traffic congestion wastes fuel and hours of work time. Construction of new roads may be very costly or impossible in many cases. Therefore, the importance of finding methods for efficient utilization of the existing infrastructure appears. Researches devote a lot of effort to minimize traffic congestions and to improve road utilization and safety.

Vehicle routing is one of the most important services that provided by a modern Intelligent Transportation Systems. The traffic on road network is dynamic, in other words the traffic flow on roads are changed over time. However there exist many alternative routes to reach destination but most drivers using the route based on their previous experiences. But the route state is changed over time, so the good route may be one bad after some time and vice versa. Hence it is clear that the depending on the derivers experience is not the optimal way to follow a specific route. The previous discussion lead us to the needing to use an intelligent route guides, that are take into account the real time road state, i.e. the dynamic data of the network such as congestions. The navigation systems, can direct drivers to good route and consequently transfer the overload to other routes. This make road network could operate efficiency at peak volume levels. Figure 1, describe the dynamic vehicle routing system [1]:

[1] Institute of Statistical Studies & Research, Cairo University, Computer Sciences & information Department

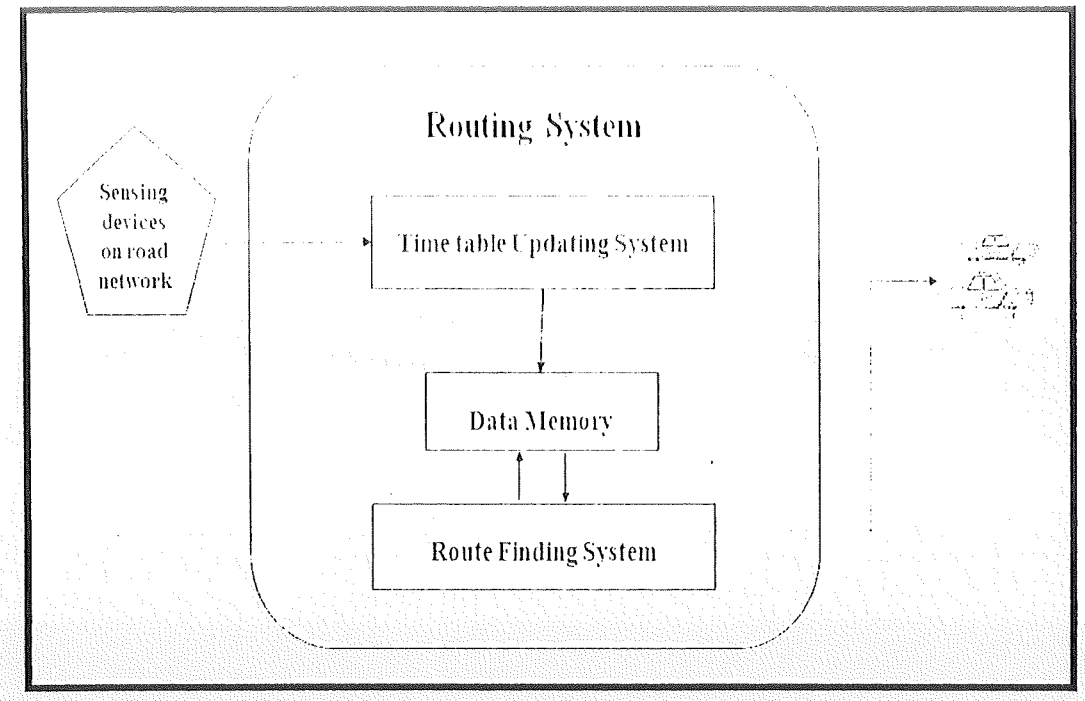


Fig. 1: Dynamic vehicle routing system

In the last decade, researchers have directed their attentions to swarm intelligence (SI) based techniques for finding good routes applying to dynamic data. Ant routing algorithms is one of the most promising SI techniques that are capable of finding near optimal solutions at low computational cost. Ant routing algorithms have been studied in many researches [2,3,4,5,6,7]

AntNet is a distributed agent based routing algorithm inspired by the behavior of natural ants [8]. It has been shown that under varying traffic loads, AntNet can adapt to changes and perform better than shortest path routing algorithm [9]. In [10], AntNet algorithm was modified to TAntNet to be more efficient when routing for road network.

Dynamic vehicle routing system is designed to work over all the day, so it must be able to deal with different circumstances and work efficiency at emergency situation. The routing algorithm must be flexible enough and under control in case of emergency. In this paper, Emergencies Reactive Controller (ERC) for TAntNet is proposed for routing of road network, The ERC enables controls in TAntNet algorithm in case of emergency. The emergencies controller can be used by the responsible person (at the control center) to make the routing algorithm avoid routing through the locations of emergency, without needs to restart the routing algorithm, instead of stopping the routing system and update the structure of network.

For the purpose of this paper, the standard TAntnet algorithm is introduced in Section 2. The proposed Emergencies Reactive Controller "ERC" is described in Section 3. Section 4 concludes the paper.

2 Threshold based AntNet algorithm (TAntNet)

Dynamic routing algorithms play an important role in the road traffic routing to avoid congestion and to direct cars to best routes. An Ant Based Control (ABC) algorithm has been applied in [3] for routing of road traffic through a city. In [11], a version of the AntNet algorithm has been applied to improve traveling time over a road traffic network with the ability to divert traffic from congested routes. In [12] a city based parking routing system (CBPRS) that used Ant based routing has been proposed. In [4] a modification to the Ant Based Control (ABC) and AntNet has been presented for routing vehicle drivers using historically-based traffic information.

In AntNet algorithm, when ants have already found a good route between a source and a destination node, they continue keep searching for another better route among all paths over all nodes of the network. This may be logical and understood for the case of high speed data communication networks, where best routes among different hosts may change in few milliseconds. However, this is not the case for dynamic traffic routing of road networks. Once we have found a good route between a source and a destination point, it is likely to remain a good route for a considerable period of time. Within such a period of time, there is no meaning and no need to keep launching ants from the source node to that destination searching for what has been already discovered and already in hand. After such a period of time, the route needs to be checked again to be sure if it is still good or there is a need to search for another good route. Realizing such a fact, for the case of road networks saves much of the unneeded computations and reduces the computational cost of the AntNet algorithm. This is the idea behind the threshold based AntNet algorithm, which proposed at [10].

The main idea of TAntNet is exploitation of the pre-known good travel times as threshold values achieves the following benefits:

- Removing of unneeded computations.
- Conserving on the discovered good path.
- Fast re-direct to the discovered good path.

TAntNet used three types of ants, forward, backward and check ants. The routing table used in TAntNet appears in table1.

Table 1: The routing table (T_k) after adding two new columns

P_{X1}	P_{Y1}	P_{Z1}	T_Good_1	G_1
P_{X2}	P_{Y2}	P_{Z2}	T_Good_2	G_2
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
P_{XN}	P_{YN}	P_{ZN}	T_Good_N	G_N

The FAntNet algorithm can be described as follows:

```
/* Main loop */
FOR each (Node s)                                   /*Concurrent activity*/
      t: current time
      WHILE t ≤ T              /* T is the total experiment time */
            Set d : = Select destination node;
            Set T_sd = 0                    /* T_sd travel time from s to d */
            IF (G_d = yes)
                Launch Check Ant (s, d);     /* From s to d*/
            ELSE
                Launch Forward Ant (s, d); /* From s to d*/
                IF (T_sd <= T_Good_sd)
                    Set G_d = yes
                END IF
            END IF
      END WHILE
   END FOR
CHECK ANT (source node: s, destination node: d)
      T_sd = 0
      WHILE (current_node ≠ destination_node)
            Select next node using routing table (node with highest probability)
            Set travel_time= travel time from current node to next_node
            Set T_sd = T_sd + travel_time;
            Set current_node = next_node;
      END WHILE
      IF (T_sd > T_Good_sd)
          Set G_d = No
      END IF
END
Forward Ant (source node: s, destination node: d)
      WHILE (current_node ≠ destination_node)
            Select next node using routing table
            Push on stack (next_node, travel_time);
            Set current_node = next_node;
      END WHILE
      Launch backward ant
      Die
END
```

Backward Ant (source node: s, destination node: d)

WHILE (current node ≠ source node) do

Choose next node by popping the stack

Update the traffic model

Update the routing table as follows:

IF ($T_{sd} \leq T\,Good_d$)

$P_{hd'} \leftarrow 1$

$P_{nd'} \leftarrow 0$, $\forall n \neq h, n \in N_k$

/* *where: h is the node "come from", k is the current node, N_k is the set of neighbor' s nodes, d' is the destination or sub path destination* */

ELSE

$P_{hd'} \leftarrow P_{hd'} + r(1 - P_{hd'})$

$P_{nd'} \leftarrow P_{nd'} - rP_{nd'}$, $\forall n \neq h, n \in N_k$

/* *where r is the reinforcement value* */

END IF

END WHILE

END

3 The proposed Emergencies Reactive Controller (ERC) for TAntNet

Reactive behavior usually appears in computer network, when one start of a data session or a failure is happening for an existing node [13]. While reactive behavior appears on road network, when emergency situation occurs on road or on intersection, and this emergency make road or intersection becomes disabled. Failure of road link or intersection can be return to many causes such as car crashes, technical failures, emergency mentainance bridge collapses or incidents induced by road works. Other events are caused by nature such as floods, landslides, heavy snowfall, storms, wildfires, earthquakes and other kinds of natural hazards [14].

When emergency case occurs in special area, the directed vehicles to this area increase the congestion problem and over time the area of traffic congestion will be enlarged, because the vehicles that enter the emergency area cannot pass and over time, roads will be closed each after other.

This paper presents an ERC that will used to immediately drop the road or node that has emergency, from the searching space of the routing algorithm, and consequently prevents the flow of vehicles to the emergency area. ERC will also make some update in the data structure of the TAntNet algorithm so it can work efficiency under the new situation. The propoased ERC consists of several procedure that makes the responsible person able to control in algorithm and can close special road or node. The ERC do the following updates in the routing tables of the TAntNet algorithm:

1- Blocking the columns of emergency nodes at all routing tables.
2- Distribute the value of the columns of emergency nodes to its neighbors; by percentage depend on the quality of neighbors pheromone values.

Figure 2 show the dynamic traffic routing system after enhanced with ERC.

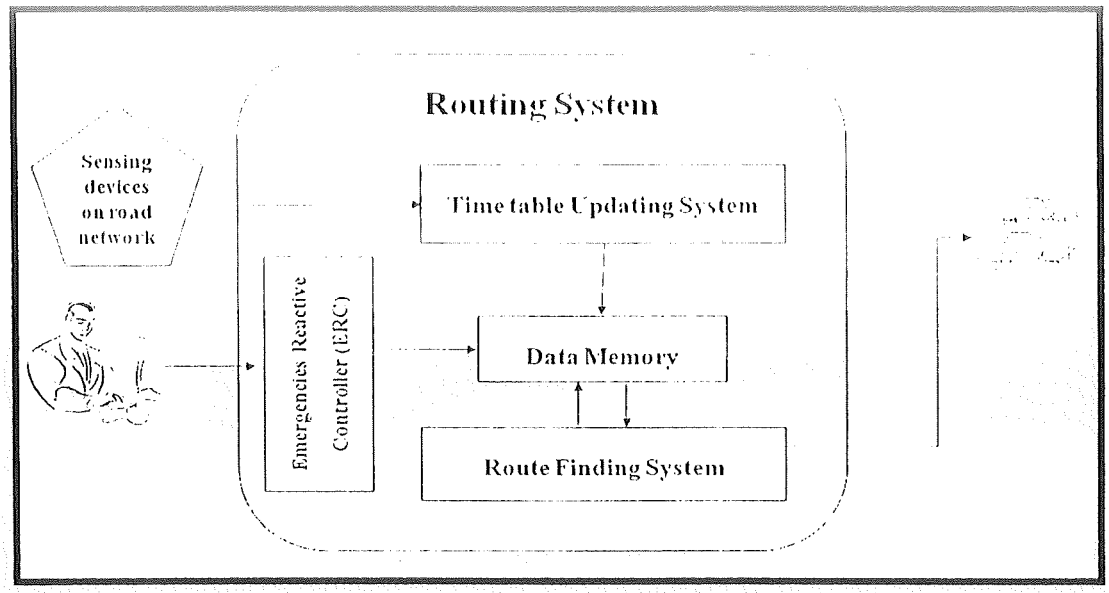


Fig. 2: The Proposed Dynamic Vehicle Routing System with ERC

Figure 3 describe the detailed Architecture of ERC in case of close road or node.

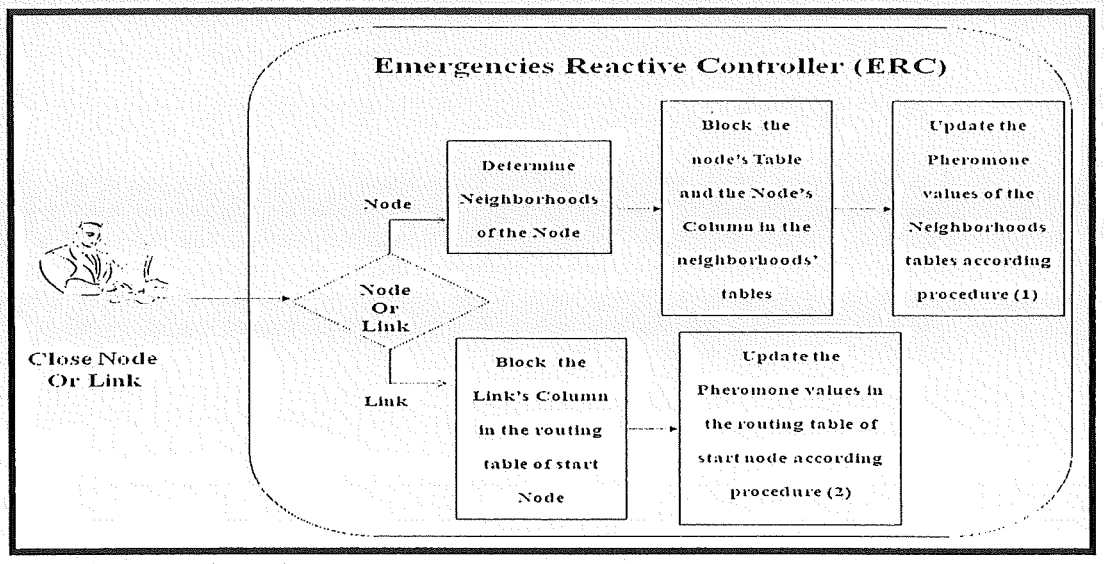


Fig. 3: ERC Architecture for close Node or Link

When the node *A* is closed, the ERC does the following steps (see Fiq. 3).

1- Determines the neighbors of the node *A*.

2- Block the routing Table of node *A* and block the column *A* in the routing Tables of the neighbors of node A, Figure 4 shows an example of blocking node.

3- The pheromone values of routing tables of the neighbors of node A is updated as show in procedure (1):

PROCEDURE (1)

FOR EACH Routing Table $K \in N_i$

FOR EACH Destination $d \in N - \{k\}$

FOR EACH Neighbor $i \in N_k - \{A\}$

$$p_{id} = p_{id} + p_{id}\left[\frac{p_{Ad}}{\sum_{l \in N_k}^{l \neq A} p_{ld}}\right]$$

END FOR

END FOR

END FOR

When the link that starts at node A and ends at node B is closed, the ERC do the following steps (see Figure 3):

1- Block the column B in the routing Table of node A, Figure 5 shows an example of blocking link.

2- The pheromone values of routing table A is updated as show in procedure (2):

PROCEDURE (2)

FOR EACH Destination $d \in N - \{A\}$

FOR EACH Neighbor $i \in N_A - \{B\}$

$$p_{id} = p_{id} + p_{id}\left[\frac{p_{Bd}}{\sum_{l \in N_A}^{l \neq B} p_{ld}}\right]$$

END FOR

END FOR

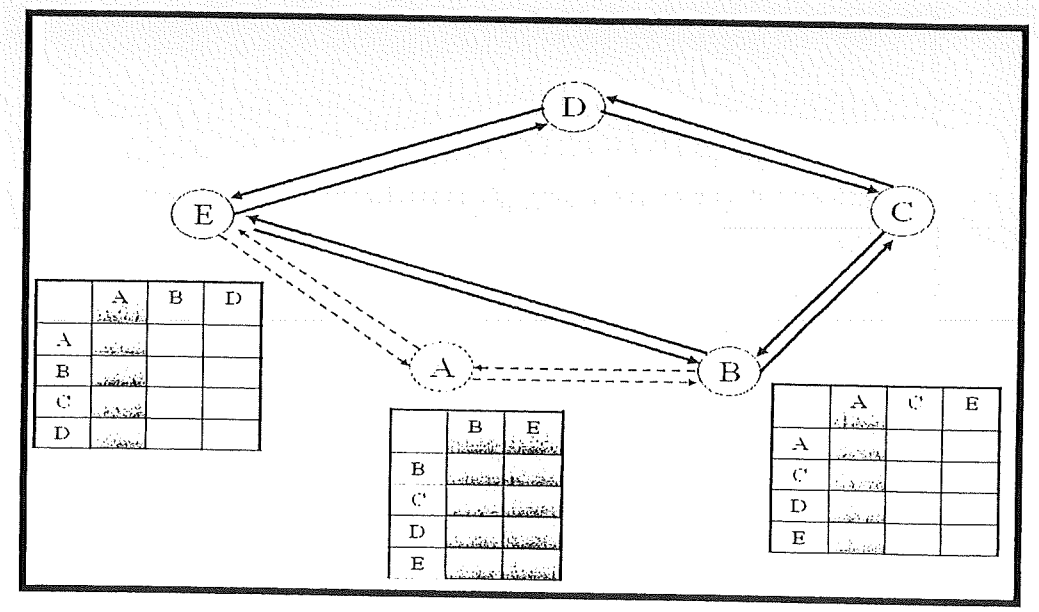


Fig.4: Example on Blocking Node

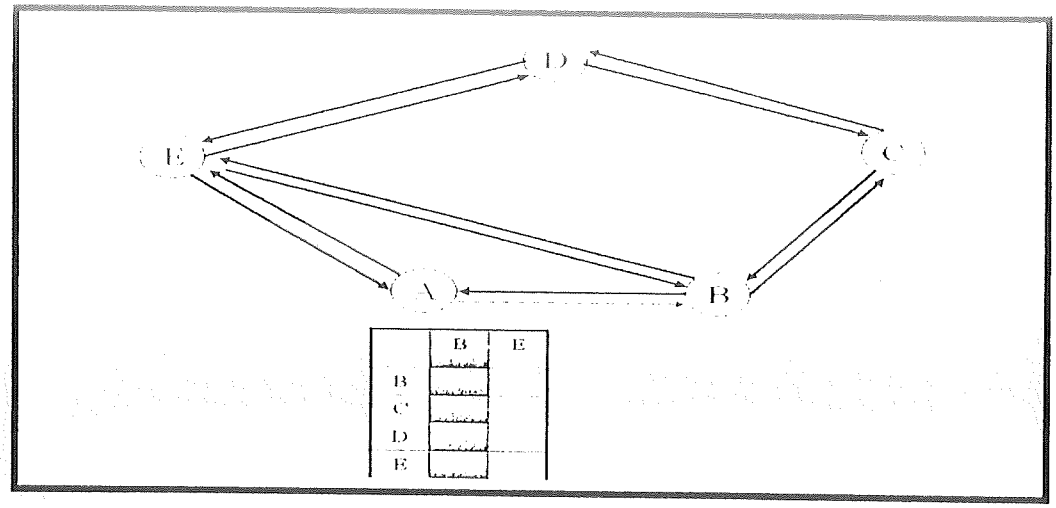


Fig.5: Example on Blocking Link

After the emergency situation is disappear the vehicles is allowed to cross this area and also it is important to renter the removed node or links to the search space of the routing algorithm again. Figure 6 shows the detailed Architecture of ERC in case of close road or node.

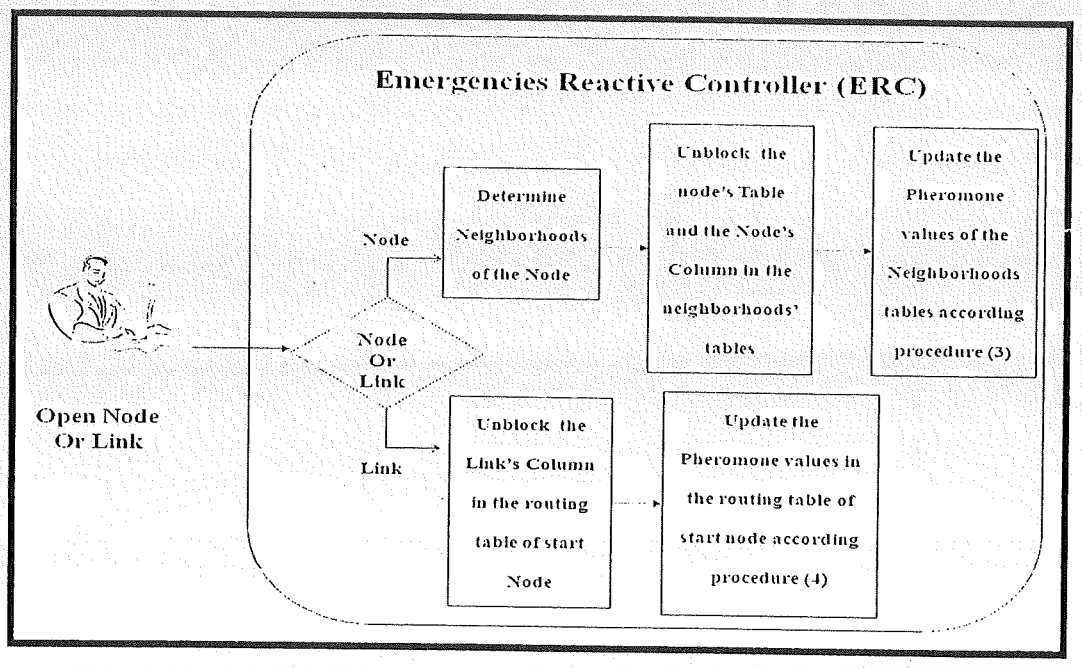


Fig. 6: ERC Architecture for open Node or Link

When the node *A* is opened, the ERC does the following steps (see Fiq. 6).

1- Determines the neighbors of the node *A*.

2- Unblock the routing Table of node *A* and unblock the column *A* in the routing Tables of the neighbors of node A.

3- The pheromone values of routing tables of the neighbors of node A is updated as show in procedure (3):

PROCEDURE (3)

FOR EACH Routing Table $K \in N_i$

FOR EACH Destination $d \in N - \{k\}$

$p_{Ad} = \frac{1}{n}$

FOR EACH Neighbor $i \in N_i - \{A\}$

$p_{id} = p_{id} - \left[\frac{p_{id}}{n}\right]$

END FOR

END FOR

END FOR

When the link that starts at node A and ends at node B is opened, the ERC do the following steps (see Fiq.3):

1- Unblock the column B in the routing Table of node A.

2- The pheromone values of routing table A is updated as show in procedure (4):

PROCEDURE (4)

FOR EACH Destination $d \in N-\{A\}$

$p_{Ad} = \frac{1}{n}$

FOR EACH Neighbor $i \in N_i - \{B\}$

$p_{id} = p_{id} - \left[\frac{p_{id}}{n}\right]$

END FOR

END FOR

4 The Efficiency of the proposed architecture

First it is clear that the proposed ERC will achieve the most important benefit, which is preventing the flow of cars to the area of emergency. But we need to show that the new update will not affect the performance of TAnNet algorithm. In this section we will proof that the using of ERC will not decrease the efficiency of the algorithm but on contrary it will increase its efficiency.

We will compare between the processing time of the TAntNet algorithm without using the ERC and the preprocessing time after applying ERC.

The TAntNet processing time without using ERC:

The time needing to launch forward and backward agents between each source and destination can be computed as shown in formula (1):

$$\sum_{i,j=1,..,n}^{i \neq j} (F_{ij} + B_{ij}) \qquad (1)$$

Where

F_{ij} : the processing time to launch an forward agent from i to j.

B_{ij} : the processing time to launch an backward agent from i to j.

The TAntNet processing time after using ERC:

Suppose that we used the ERC and stop node k.

The time needing to launch forward and backward ant between each source and destination after applying ERC and stop node k can be computed as shown in formula (2):

$$\sum_{i,j=1,\dots,n}^{i\neq j;\, i,j\neq k}(F_{ij}+B_{ij}) \qquad (2)$$

Formula (1) can be written as follow:

$$\sum_{i,j=1,\dots,n}^{i\neq j}(F_{ij}+B_{ij}) = \sum_{i,j=1,\dots,n}^{i\neq j;\, i,j\neq k}(F_{ij}+B_{ij}) + \sum_{j=1,\dots,n}^{j\neq k}(F_{kj}+B_{kj}) + \sum_{i=1,\dots,n}^{i\neq k}(F_{ik}+B_{ik}) \qquad (3)$$

Since processing time is greater than zero, then

$$\sum_{i,j=1,\dots,n}^{i\neq j;\, i,j\neq k}(F_{ij}+B_{ij}) + \sum_{j=1,\dots,n}^{j\neq k}(F_{kj}+B_{kj}) + \sum_{i=1,\dots,n}^{i\neq k}(F_{ik}+B_{ik}) > \sum_{i,j=1,\dots,n}^{i\neq j;\, i,j\neq k}(F_{ij}+B_{ij}) \qquad (4)$$

From equation (3) the L.H.S in formula (4) can be replaced by the L.H.S in equation (3), then

$$\sum_{i,j=1,\dots,n}^{i\neq j}(F_{ij}+B_{ij}) > \sum_{i,j=1,\dots,n}^{i\neq j;\, i,j\neq k}(F_{ij}+B_{ij}) \qquad (5)$$

Formula (5) shows that the processing time will decrease by applying ERC, and consequently this time will be used to launches more ants, which will lead to fast convergence to good route and consequently the performance of the algorithm will be enhanced. In addition the applying of ERC will avoid vehicles from entering the emergency area, and hence prevent congestion caused by emergency situation, which will enhance also the situation of the traffic on the road network.

5 Conclusion

This paper presents an enhancement to the dynamic traffic routing system using TAntNet algorithm by embedding an Emergencies Reactive Controller "ERC" that can be used in the emergencies cases to enable control in the algorithm and prevents using of special node or special link. The ERC used to prevent the flow of vehicles to the emergency area. It is easy to notice that at the normal case, when no emergency occurred, there is no closed of any nodes or links, and consequently the algorithm will acts with no difference in its performance than that of the normal TAntNet algorithm. At emergency case, some links or nodes will be closed, and the searching space of paths will be less than that of the normal algorithm. This consequently will enhance the performance of algorithm. In addition to gain the most important objective, which is making the routing algorithm avoiding directing vehicles to the emergency area which will decrease the congestion on this area and make the problem of congestion does not transfer to surrounding areas.

References

1. Kroon, R.: Dynamic vehicle routing using Ant Based Control. Master's thesis, Delft University of Technology (2002)
2. Gupta, A. K., Sadawarti, H., Verma, A. K.: Computation of Pheromone Values in AntNet Algorithm. International Journal of Computer Network and Information Security. 9, 47-54 (2012)
3. Shah, S., Bhaya, A., Kothari, R., Chandra, S.: Ants find the shortest path: a mathematical proof. Swarm Intelligence 7(1): 43-62 (2013)
4. Kroon, R., Rothkrantz, L.J.M.: Dynamic vehicle routing using an ABC-algorithm. Transportation and Telecommunication in the 3rd Millennium, Prague, 26-27 (2003)
5. Suson A. Dynamic routing using ant-based control. Master thesis, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology (2010)
6. Yousefi P, Zamani R.: The Optimal Routing of Cars in the Car Navigation System by Taking the Combination of Divide and Conquer Method and Ant Colony Algorithm into Consideration. International Journal of Machine Learning and Computing. 3, 44-48 (2013)
7. Jabbarpour M. R., Malakooti H., Noor R. M., Anuar, N. B., Khamis, N.: Ant colony optimisation for vehicle traffic systems: applications and challenges. International Journal of Bio-Inspired Computation. 6(1), 32-56 (2014)
8. Di Caro, G., Dorigo, M.: AntNet: Distributed stigmergetic control for communications networks. *Journal of Articial Intelligence Research (JAIR,) 9,* 317-365 (1998)
9. Dhillon, S.S., Van Mieghem, P.: Performance analysis of the AntNet algorithm. *Computer Networks. 51,* 2104 –2125 (2007)
10. Ghazy A. M., El-Licy F., Hefny H. A.: Threshold based AntNet algorithm for dynamic traffic routing of road networks. Egyptian Informatics Journal, Vol. 13, Issue 2, Pp. 111-121 (2012)
11. Tatomir, B., Rothkrantz, L.J.M.: Dynamic traffic routing using Ant based control. In: *IEEE International Conference on Systems, Man and Cybernetics (SMC 2004)*, Impacts of Emerging Cybernetics and Human-Machine Systems, pp. 3970-3975 (2004)
12. Boehlé, J. L., Rothkrantz, L. J. M., Wezel, M. V.: CBPRS: a city based parking and routing system. Technical report ERS-2008-029-LIS, Erasmus Research Institute of Management, ERIM, University Rotterdam (2008)
13. Ducatelle, F., Di Caro, G., Gambardella, L.: Using ant agents to combine reactive and proactive strategies for routing in mobile ad hoc networks. Int. Journal of Computational Intelligence and Applications (IJCIA), Special Issue on Nature Inspired Approaches to Networks and Telecommunications . 5(2), 169–184 (2005)
14. Jenelius, E. and Mattsson, L. G.: Road network vulnerability analysis of area-covering disruptions: A grid-based approach with case study. Transportation Research Part A: policy and practice. 46(5), 746 – 760 (2012)

Fuzzy Approach for Recommending Suitable Crops with its Plantation Dates for Suez Canal Zone

Assem H. Mohammed[1,2], Ahmed M. Gadallah[1,3], Hesham A. Hefny[1,4]

Abstract Suez Canal zone is one of the most promising zones in the whole world and not only in Egypt. If invested well planned. The agriculture is considered of the most important investments in such zone. Because Availability of water and suitable climate and fertile soil. This paper provides a fuzzy query approach for discovering the more suitable crops to be planted with its plantation date in the Suez Canal zone. The proposed approach consists of six steps. The first step is the selection of the climate data of the under study year of the Suez Canal zone step. In the second step the proposed approach selects the crops from Crops suitable climatic requirements database step. The generation of suitability fuzzy membership functions in the third step. Consequently, the evaluation of the suitability degree for each day in the fourth step. Then the period's evaluation in step five. Finally, in step six the proposed approach selects the suitable periods for the plantation of each crop. The proposed approach suggests a set of crops to be planted with is plantation date in Suez Canal zone. The proposed approach applied on the potatoes and wheat crops with climate data of temperature and humidity of Ismailia.

Keywords: Fuzzy query, Relational database, SQL, Spatial Agro-Climatic Database

1 Introduction

A Spatial database is a collection of data concerning objects located in some reference space, which attempts to model some enterprise in the real world. The real world abounds in uncertainty, and any attempt to model aspects of the world should include some mechanism for incorporating uncertainty. There may be uncertainty in the understanding of the enterprise or in the quality or meaning of the data[1],[12].So that there has been a strong demand to provide approaches that deal with inaccuracy and uncertainty in geographical information systems (GIS) and their underlying spatial databases. Spatial agro-climatic database is a spatial database that contain the data of the climate variables for some places during a specific time period.

SQL [1]is a standard query language for relational databases. The SQL was initially presented by Chamberlin and Boyce [2]. Since then, it has been used for data selection in many relational databases and information systems. SQL is one of the key factors for accepting relational databases and for their use in storing and retrieving vast amount of data. Its advantages among others are: optimized work with relational database management systems and understandable interpretation for users.

The SQL uses two-valued logic (crisp logic) in querying process. It means that a small error in data values or in cases when user cannot unambiguously define the criteria by crisp values, may involve some inadequately selected data. This limitation of the SQL can be avoided by fuzzy logic [14]. The fuzzy set theory was initiated by Zadeh [2].

[1] Cairo University, Institute of Statistical Studies and Research, Cairo, Egypt
[2] eng_asemhm@yahoo.com
[3] ahmgad10@yahoo.com
[4] hehefny@cu.edu.eg

Since then, many researches and applications have been published. Fuzzy queries have appeared in the last 30 years to cope with the necessity to soften the Boolean logic in database queries. A fuzzy query system is an interface to users to get information from database using natural language sentences[3].This area is still interesting as there are possibilities for the improvement of existing approaches and for creating new ones[13] .

The goal of this paper is to present a fuzzy approach for querying a spatial agro-climatic database depending on human like Language. Such query approach will help in determining the more suitable crops to be planted in the Suez Canal zone.

The rest of this paper is divided into four section. The second section presents the related works. The proposed approach is addressed in the third section. The fourth section illustrates a case study. The conclusion is presented in the last section.

2 Related Works

Many approaches have worked on the problem of climate changes and agriculture .some approaches show the impact of climate change on crop production like [15] and [16]. it proved that agricultural productivity had affected by been climate change .it also advise to make some adaptation on planting dates or by planting crops that not more sensitive to climate changes to get over such effects. Also climate changes affect the water resources in regions in which agriculture based on rain. As water resources is the most important parameter in plantation process a new approach has been developed to make adaptation to crop planting date with climate changes like [4] and [5]. Also a crop yield forecasting model to map the relations between climate data and crop yield was developed like in [6]. This technique based on time series data of 27 years for yield data and weather data. Other approaches developed Fuzzy-based Decision Support Systems for evaluating land suitability and selecting crops to be planted is provided like in [7] and [8]. A fuzzy rule was developed for evaluating land suitability and selecting the appropriate crops to be planted considering the decision maker's requirements in crops selection with the efficient use of the powerful reasoning and explanation capabilities of DSS.

Unfortunately, all of the above algorithms do not provide weight or matching measures for period's suitability of plantation after making adaptation on climate changes. Also in all of them no clustering for the climate data is made. This mean that the algorithm calculate the suitability of the period day by day every time of searching suitability for planting any crop so that it take long time. Some algorithms depend on the average values of climate data like in [5], [6] and [7] and this is not true as most of crops affected by the minimum and maximum values of climate variables.

This paper provides a fuzzy query approach for discovering the more suitable crops to be planted with its plantation date in the Suez Canal zone. The proposed approach

consists of six steps. The first step is the selection of the climate data of the under study year of the Suez Canal zone step. In the second step the proposed approach selects the crops from Crops suitable climatic requirements database step. The Generation of suitability fuzzy membership functions in the third step. Consequently, the evaluation of the suitability degree for each day in the fourth step. Then the period's evaluation in step five. Finally, in step six the proposed approach selects the suitable periods for the plantation of each crop.

3 The Proposed Approach

The proposed approach as shown in figure 1 depends on the agro climatic database of Egypt from [18]. The agro climatic database contain the historical data of Egypt and the predicted data for the next year. The proposed approach applied on the data of Ismailia governorate. Also the proposed approach gets the climatic requirements of each crop from the Crops suitable climatic requirements database. Finally, the proposed approach assumes that the water and other variables that affect the planation process of the selected crops already available. In other words, the proposed approach checks the suitability of the climate only.

The proposed approach consists of six steps:

1- The selection of the climate data of the under study year of the Suez Canal zone step
2- The Selection of the crops from Crops suitable climatic requirements database step
3- The Generation of suitability fuzzy membership functions step
4- Evaluation of the suitability degree for each day step
5- period's evaluation step
6- Selection of the suitable periods step

3.1 The selection of the climate data of the under study year of the Suez Canal zone step:

In this step the proposed approach selects the data of Suez Canal zone from the agro-climatic database. Then it checks the accuracy and completeness of data.

3.2 The Selection of the crops from Crops suitable climatic requirements database step:

In this step the proposed approach selected the crop that is wanted to check its suitability.

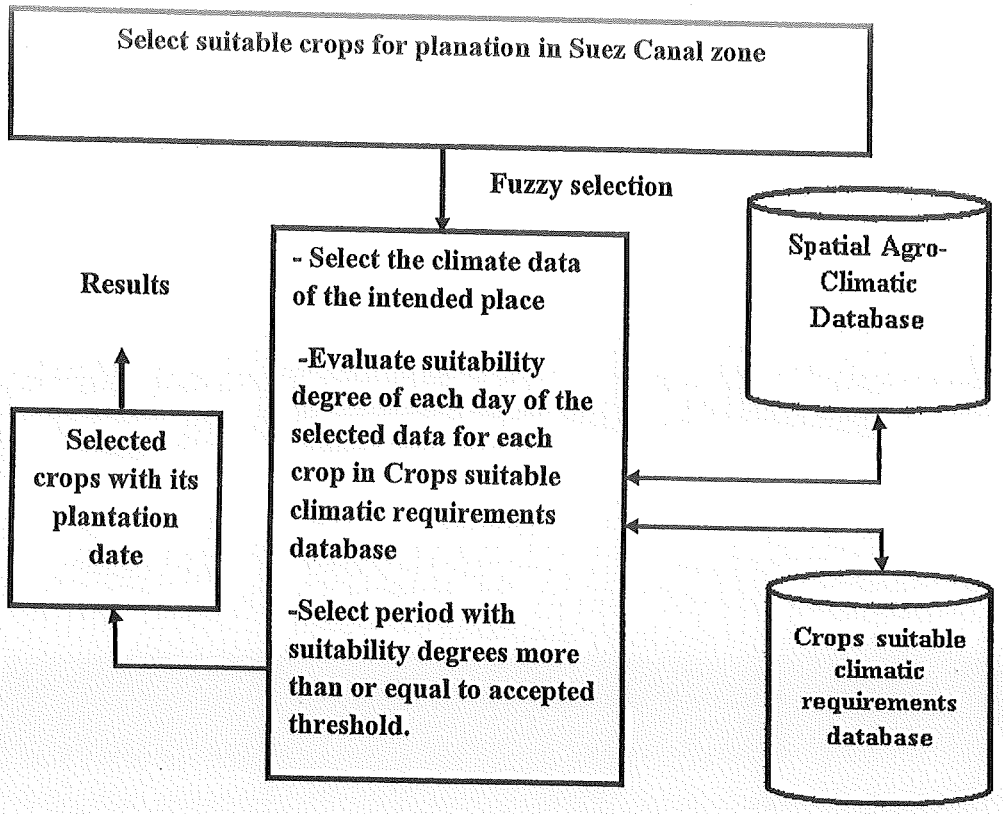


Fig.1.The proposed approach architecture

3.3 The Generation of suitability fuzzy membership functions step:

This step generates the suitability fuzzy membership functions based on the crops climatic requirements gathered in the previous module. This step allows defining a set of fuzzy membership functions describing the crop suitable values of a specific climate variable. For example, based on the values from [9] and [10] for potatoes which gathered in the previous module, the suitability membership function of temperature is defined as depicted in equations (1) and fig. 2 for first stage of potatoes life time.

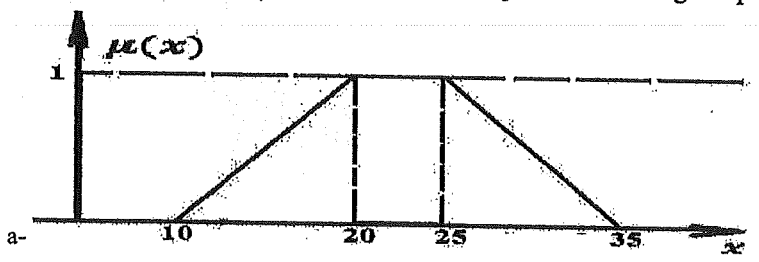


Fig. 1. The fuzzy membership function of the potatoes first phase temperature

$$\mu_{first\_temp\_s}(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \le x < b \\ 1 & b \le x < c \\ \frac{d-x}{d-c} & c \le x < d \\ 0 & x \ge d \end{cases} \quad (1)$$

Where a=10, b=20, c=25 and d=35

3.4 Evaluation of the suitability degree for each day step:

In this step the prosed approach uses the suitability fuzzy membership functions to evaluate the suitability degree of each day in the select data of the year.

3.5 period's evaluation step:

In this step the proposed approach select continuous period with length equal to the age of the under study crop. Then it evaluate the suitability of the whole period.

3.6 Selection of the suitable periods step:

In this step the proposed approach selects the period with suitability degree more than or equal to the accepted threshold. Finally, the propped approach adds the accepted periods with the selected crop to the suitable list

4 Illustrative case study

The proposed approach is applied to selects the plantation periods of potatoes and wheat in Ismailia. The agro climatic spatial database for Ismailia is used to predict, for the incoming year, both of temperature and humidity as some of most important climate change variables. This approach have been tested on the data of potatoes and wheat as shown in the next two section.

4.1 potatoes

The potatoes crop growth divided into two stages (periods). The first one is the roots and stems growth period and the second one is the tubers growth period.
The climate requirements for potatoes plantation, in the roots and stems growth period, are: [9], [10]

a- The average age of the plant is 60 days.
b- Maximum temperature less than 35
c- Minimum temperature more than 15
d- Best temperature from 20 to 25
e- Best humidity from 65% to 70%

The climate requirements for potatoes plantation in the tubers growth period are: [9], [10]

a- The average age of the plant is 60 days.
b- Maximum temperature less than 28
c- Minimum temperature more than 5

d- Best temperature from 15 to 18
e- Best humidity from 65% to 70%

The figures 2,3,4,5,6,7,8,9,10,11,12 and 13 show the start and end of each buttonhole for potatoes plantation dates each buttonhole consists of roots and stems(R&S) sub period and tubers growth sub period. Also it show the suitability degree of each sub period:

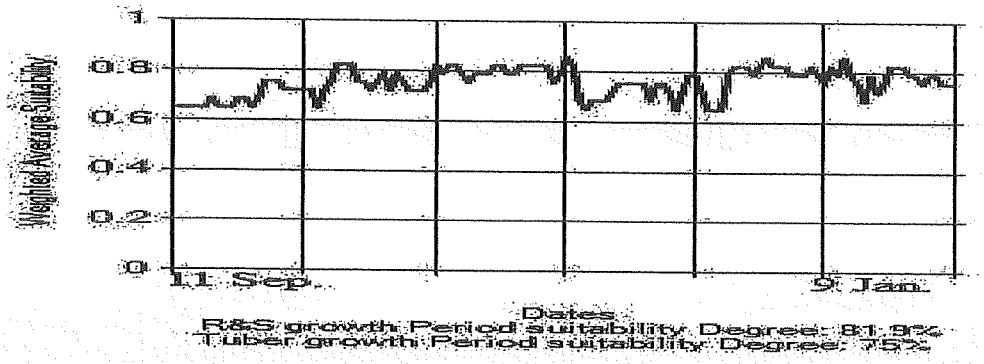


Fig. 3 The start of suitable period for the plantation of the potatoes

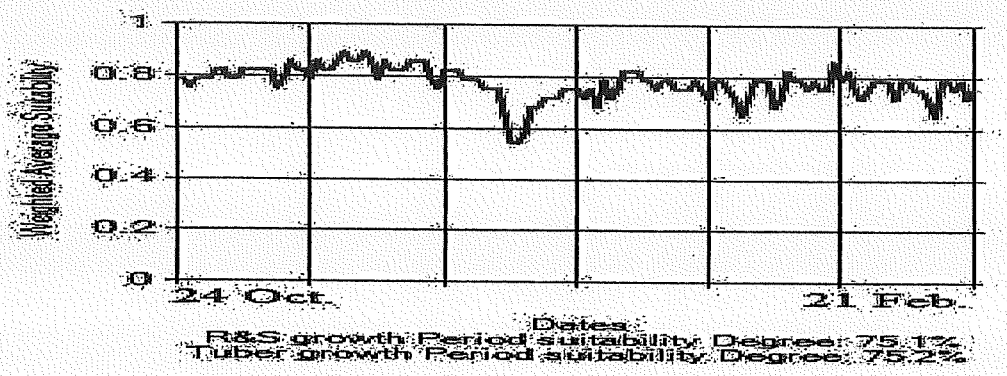


Fig. 4 The end of suitable period for the plantation of the potatoes

Table .1 the suitable period for the plantation of the potatoes recommended by the proposed approach

Potatoes Plantation dates a	Roots and stem growth period suitability	Tubers growth period suitability
From 11 Sep.	81.9%	75%
To 24 Oct.	75.1%	75.2%

4.2 wheat

Like in the potatoes crop, the proposed approach is applied to select the plantation periods of the wheat crop in Ismailia. According to the data from [11], and [17] the growth age is divided into five stages. Table 3 the results of the proposed approach.

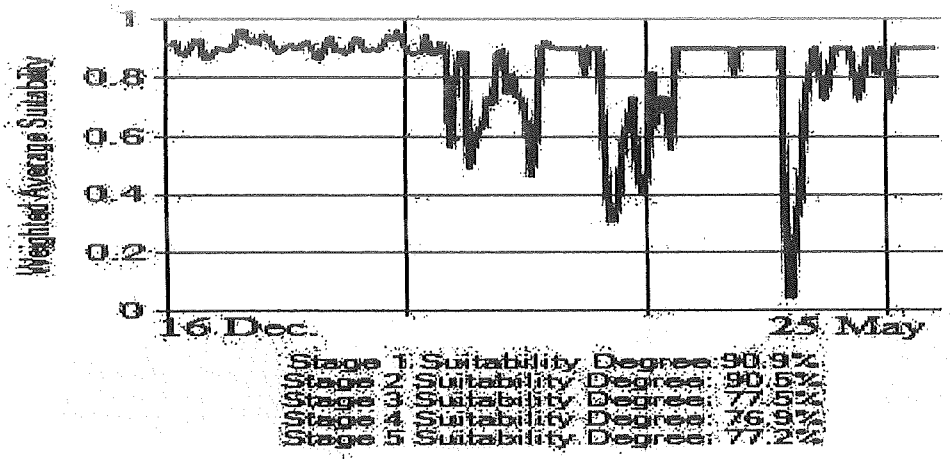


Fig. 5 The start of suitable period for the plantation of the wheat

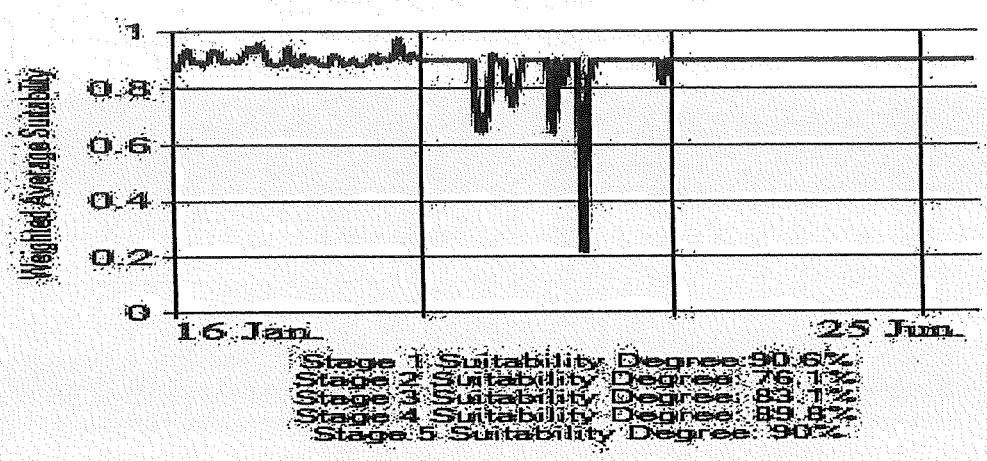


Fig. 6 The end of suitable period for the plantation of the wheat

Table .2 the suitable period for the plantation of the wheat recommended by the proposed approach

Optimized Plantation dates	Stage1 (Germination) suitability	Stage2 (Tillering) suitability	Stage 3 (Stem longation) suitability	Stage4 (Flowering) suitability	Stage 5 (maturity) suitability
From 16 Dec.	90.9%	90.5%	77.5%	76.9%	77.2%
To 16 Jan.	90.6%	76.1%	83.1%	89.8%	90%

Considering table 1 and table 2 the proposed approach proved that at the potatoes and the wheat crops could be planted in the Suez Canal zone. Table 1 presented the suitable dates for the plantation of the potatoes. While table 2 Table 1 presented the suitable dates for the plantation of the wheat.

Table .3 Comparative analysis for the proposed approach against other fuzzy query based approaches

Approach	Inputs	Processing	Outputs
A Crop Model and Fuzzy Rule Based Approach for Optimizing Maize Planting Dates in Burkina Faso, West Africa[4]	The average value of climate variables	1- Calculate the average value for each climate variable. 2- Determining of the plantation date of a certain crop.	A single period for the plantation date for the selected crop within two weeks.
Crop Yield Forecasting by Adaptive Neuro Fuzzy Inference System[6]	The average weekly values of climate variables	3- Divide the year into weeks. 4- Calculate the average value of the weak for each climate variable. 5- Set the calculated average value to each day of the weak. 6- Determining of the plantation date of a certain crop.	A single plantation date for the selected crop.
Consideration of Climate Conditions in Reservoir Operation Using Fuzzy Inference System (FIS)[16]	The climate variables that affect the availability of water	1- The expected amount of water at a certain time is calculated based on the climate variables that affect it. 2- Determining the crops to be planted.	Suggests a single crop plantation date based on the availability of the amount of water.
The proposed approach	The values of climate variables for each day of the year. It also takes into account the max and min value for climate variables that have max and min like temperature.	1- Clustering the year to periods that more similar to each other in the climate condition with threshold not more than the accepted threshold. 2- Defining a set of fuzzy membership functions describing the relationship between a crop data and the suitable values of a specific climate variable. 3- Evaluating the resulted clustered periods from first phase respecting the crop suitable temperature and humidity data. Finally, selecting the more suitable plantation periods with suitability grade greater than or equal to a predefined threshold.	The propped approach divides the planation period into sub periods according to the crop growth phases and it calculates the suitability of the whole sub periods

Considering table3 the proposed approach has many advantages more than the similar database fuzzy query based approaches in table3. The first advantages, it depends on the max and min value of temperature as in real life a lot of crops get damaged when the min temperature less than some threshold, where the average value of the temperature of the same period is accepted. The second advantage is that the proposed approach divide the year into more similar period according to climate variables values at the first time of the approach execution. Such clustering make the selection of the most suitable period faster as the approach will select from a small number of cluster not from the all days of the year. Thirdly, the proposed approach provides more than one plantation date with a suitability degree for each one. Finally, all the previous work treats the age of any crop as one stage, while in reality a lot of crop need different climatic requirements in each stage of its age.

5 Conclusion

This paper proposes a fuzzy based approach for suitable crops with its plantation dates for Suez Canal zone. Based on the available historical spatial agro-climatic database for Ismailia governorate in Egypt, the proposed approach suggested a set of crop with its planation dates. The proposed approach was applied on the wheat and potatoes crops. Accordingly, such approach greatly helps in agricultural planning to enhance the plantation plan of any crop in a specific location over a set of incoming years. Also, it directly increases the profit of a crop plantation process. As a future work, the proposed approached is planned to be more generic to be applied to any crop at any location in Egypt given such location agro climatic historical database.

References:

1. J. Zhang and M. Goodchild, "Uncertainty in Geographical Information", Taylor & Francis, London (2002)
2. L. Zadeh "Fuzzy sets", Information and Control 8, 338-353 (1965)
3. Branco, A., Evsukoff, A., Ebecken, N., "Generating fuzzy queries from weighted fuzzy classifier rules", ICDM work-shop on Computational Intelligence in Data Mining) 21-28 (2005)
4. W. Moussa, L. Patrick, B. SEYDOU, T., S. Moussa, H., "A Crop Model and Fuzzy Rule Based Approach for Optimiz-ing Maize Planting Dates in Burkina Faso, West Africa," Journal of Applied Meteorology and Climatology Volume 53(2014)
5. S.A. Mohaddes and M. Ghazali, "Application of the Fuzzy Approach for Agricultural Production Planning in a Watershed, a Case Study of the Atrak Watershed, Iran," American-Eurasian J. Agric. & Environ. Sci., 3 (4): 636-648 (2008)
6. K. Pankaj, "Crop Yield Forecasting by Adaptive Neuro Fuzzy Inference System," Mathematical Theory and Model-ing 1, 3 (2011)
7. S. Imas and H. Sri, "A Fuzzy Based Decision Support System for Evaluating Land Suitability and Selecting Crops," Journal of Computer Science 6 (4), 417-424 (2010)
8. G. Rajeshwar, B. Parag, S. Khmaitkar, "Predicting Suita-bility of Crop by Developing Fuzzy Decision Support Sys-tem," IJETAE Volume 3, Special Issue 2(2013)
9. kenaaonline.com/files/0009/9805/%D8%B2%D8%B1%D8%A7%D8%B9%D8%A9%20%D9%88%D8%A7%D9%86%D8%AA%D8%A7%D8%AC%20%D8%A7%D9%84%D8%A8%D8%B7%D8%A7%D8%B7%D8%B3.pdf(Accessed date: 8-15-2014)
10. caae-eg.com/new/index.php/2012-12-25-10-49-19/2010-09-18-17-00-51/2011-01-15-19-27-42/160-2010-11-05-15-53-50.html (Accessed date: 8-15-2014
11. kenanaonline.net/page/1966 (Accessed date: 8-15-2014)
12. B. Theresa and E. Frederick, "Fuzzy and Rough Set Approaches for Uncertainty in Spatial Data," Springer- erlag Berlin Heidelberg 103–129(2010)
13. H. Miroslav, "Fuzzy Improvement of the SQL," Yugo-slav Journal of Operations Research 12 2, 239-51(2011).
14. G. José, U. Angélica, and P. Mario" Fuzzy Databases, Modeling, Design and Implementation," Idea Group Publishing, London, (2006)
15. N. Defang, I. Manu, M. Bime, O. Tabi, H. Defang, "Impact of climate change on crop production and development of Muyuka subdivision – Cameroon," International Journal of Agriculture, Forestry and Fisheries. 2(2): 40-45(2014)
16. R. Hamid, A. Mohammad, H. Mohammad, "Consideration of Climate Conditions in Reservoir Operation Using Fuzzy Inference System (FIS)," British Journal of Environment & Climate Change, 3(3): 444-463(2013)
17. El-Boraie, Gaber and Abd El-Raman "optimizing water use efficiency of wheat under South Sinai condition (Wadi Sudr): 1- yield responses at different growth stages", Journal of agricultural science Mansoura university, 34 (12): 11477 – 11488, (2009)
18. worldweatheronline.com/ (Accessed date: 10-20-2013)

Formal Model: Epistemic Hybrid Automata

M. Rana[1], M. Ammar[2], H. Hesham[3]

Abstract

In multi-agent systems reasoning about the behavior of each agent is influenced by its acquired knowledge at each point of time. Each gent has two types of behavior; a dynamic behavior, representing its continuous evolution in the environment, and a discrete behavior, representing how it moves from step to another. Most of the recent research didn't present a framework that allows agents to reason about both knowledge and the continuous dynamic behavior of the agent. This paper presenting a novel framework that allows to represent the continuous behavior of agent and on the other hand, it allows agent to reason about knowledge. To achieve this goal, the paper proposes a new logic called Epistemic Hybrid Tree Logic (EHTL) that allows us to specify and hence verify properties of agents expressed in terms of epistemic and hybrid logic. Such framework would help in agent's decision making especially when agent sited in a dynamic environment.

Keywords: *Modal Logic, Temporal Logic, Epistemic Logic, Hybrid Automata, Epistemic Hybrid Automaton (EPH), Epistemic Hybrid Tree Logic (EHTL).*

1 Introduction

Every real time system (RTS) has behavior that represents how system evolve over environment, RTS can be represented as an agent or a set of multi-agent; where each agent is represented in a logical framework, logic that used to represent the logical framework is modal logic [1]. The behavior of MASs changes over time and makes system gain knowledge at every point of time, the logic used to represent the knowledge of agent and how system evolve over time is epistemic logic [16,18,20] and temporal logic [8,17], where epistemic logic and temporal logic are subclasses of modal logic. Epistemic logic used to reason about agent knowledge or other agent's knowledge, where temporal logic (TL) used to represent the flow of events over time, TL has different types one of them is real time logic [2] it is used to represent quantitative properties by representing the time constraints, it is usually interpreted over timed automata [2,7]. The problem with timed automata is that it cannot capture the continuous and discrete behavior of the agent at the same time, so Henzinger in [4,6] presented a mathematical model called hybrid automata which is a generalization of timed automata. Hybrid automaton captures the continuous behavior of a system using differential

1. Masters Student Department of Computer Science, Institute of Statical studies and Research, Cairo University, Egypt
2,3. Department of Computer Science, Institute of Statical studies and Research, Cairo University, Egypt

equations and captures the discrete behavior of the systems using finite state automaton.

Some researchers presented hybrid automata as a verification model of real time logic [10,11,12] and others tried to present a logical framework that allows agent to reason about knowledge in the same framework of time by combining epistemic logic with temporal logic [9], epistemic with alternating temporal logic [18] and epistemic with real time logic[15,21], but the problem with these works is that they are not able to represent knowledge in the same framework with the dynamic and discrete behavior of the agent. To our Knowledge, no one tried to do so.

To this end, the main contribution of this paper is to extend the work presented in [9,10,12] and present a new framework called epistemic hybrid tree logic that allows us to specify and hence verify formulas expressed in terms of knowledge and continuous dynamics. To do so, the paper present a new interpretation model called epistemic hybrid automata.

The rest of paper is organized as follows. In section 2 we present the related work, in section 3 we give abstract view about hybrid automata and epistemic logic, in section 4 we give some definition to understand syntax and semantic of epistemic hybrid automata (EPH). In section 5 we present a novel framework that interpreted over EPH and illustrate how our model will work using parallel composition, in section 6 we clarify our quantitative and knowledge properties. Finally we conclude in section 7 by presenting conclusion.

2 Related Work

A lot of papers had tried to combine epistemic logic with temporal logic, as [21] the author combined epistemic logic and timed temporal logic with deontic logic, he uses verification model timed automata, this work only represents the discrete steps without the dynamic behavior of the system and the type of knowledge that each agent reasons about is the behavior of each other agent, where we can get this type of information using from hybrid automata, also [9] combined temporal logic with epistemic logic and interpreted over finite automata this model lacks of representing the time constraints of the system, it represents only the flow of events also not able to represent all kinds of knowledge, it only gives an integer counter to represent that agent knows a new knowledge. In [18] presented type of temporal logic named alternating logic (AL) which used to introduce a strategy that allows agents to cooperate to perform a specific task, authors combined (AT) with temporal logic. The logic interpreted over epistemic alternating automata, this model also lacks of presenting the quantitative properties and reason only about

each automaton behavior knowledge. Other researchers used hybrid as verification tool as in [10,11,12], that used to verify Region Computation Tree Logic, which used to represent the quantitative properties and presenting both dynamic and discrete behavior of the system. This paper combines the advantages of [9,10,11,12,21] to present our framework.

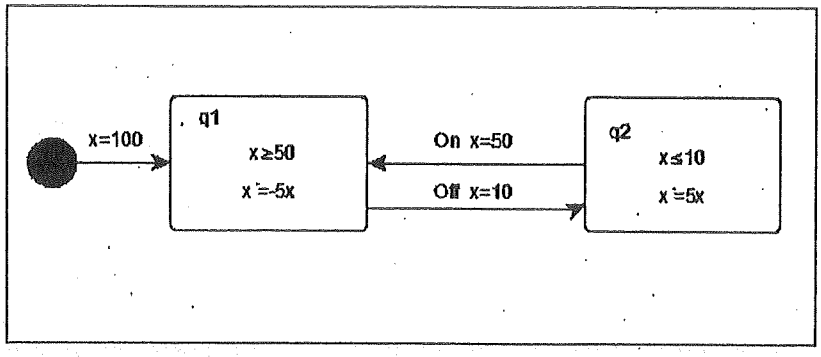


Fig. 1. Simple hybrid automaton

3 Hybrid Automata and Epistemic logic: Preliminaries

Before introducing our framework we need first to give an overview about hybrid automata [4,6] and epistemic logic [16,1].

3.1 Hybrid Automata

Hybrid automaton is a generalization model of timed automaton [2,7] which is used to represent the continuous and discrete behavior of an agent. Formally, each agent can be expressed as an automaton HA where:

$$HA = (Var, Q, Flow, Inv, E, Jump, Assgn, Event),$$

where Q is a finite set of locations representing the mod of agents and inside each location there is some real values Var. An agent stays in a certain location according to dynamical constraints $Flow$ that is usually represented as differential equation. The agent then moves from one location to another according to linear constraints Inv, the connection between locations is represented by transition E, each transition has a $Jump$ condition and an $Event$ that fired on jumping, when agent moves from one location to another some values are reset before going to the other location using $Assgn$ function. An example of hybrid automaton is shown in **Fig. 1.**

In the following we present epistemic logic.

3.2 Epistemic Logic

If a human interacting with other humans, then reasoning about their knowledge is necessary to understand their behavior. Also the gain of knowledge distributed over other humans is important to reason about the state of knowledge and to be able to take a decision and to make agent able reason like humans, so we need formalism that support knowledge representation. Such formalism called epistemic logic [16]. Epistemic logic used to express the knowing property for agent. The syntax is expressed in the following:

$$\varphi: = \neg \varphi \mid (\varphi \wedge \varphi) \mid C_m \varphi \mid K_i \varphi \mid D_m \varphi$$

Where $K_i \varphi$ means that agent i knows φ, $C_m \varphi$ means that the group of agents m have the common knowledge of φ and $D_m \varphi$ means that the knowledge φ is distributed over the group of agents m. the semantic used for epistemic logic was first presented by kripke at 1972 named "possible world structure" where he assumed that to make agent knows a specific knowledge φ, the value of φ must be true at every world that agent consider possible and there is an accessibility epistemic relation between each two worlds that makes the worlds indistinguishable for the agent. This type of semantic faces a problem named omniscience problems [16], it makes researchers present alternative semantics to avoid the omniscience problem.

Example: Alice and Bob are sitting in their dark office wondering whether or not it is raining outside. Now, none of them knows, but Alice knows something about her friend Carol, namely that Carol wears her red coat only if it is raining. Bob does not know this, but he just saw Carol, and noticed that she was wearing her red coat. Even though none of them knows whether or not it is raining, it is distributed knowledge amongst them that it is raining. If either one of them tells the other what they know, it will be clear to the other that it is raining. It can be represented as:

$$(K_b \alpha \wedge K_a (\alpha \rightarrow \beta)) \rightarrow D_{a,b} \beta)$$

4 Proposed Epistemic Hybrid Automata

In this section, we present a a new formal model that allows us to express in the same framework both continuous dynamic behaviour together with epistemic behaviour of system. Our formal model named epistemic hybrid automaton (EPH). In the following we present our model's syntax and semantic.

4.1 Syntax

Before presenting the syntax and semantic of epistemic hybrid automata we need to define what is meant by agent, knowledge of agent, epistemic relation, and type of constraint that may appear inside our epistemic hybrid automaton.

Definition-1 (Agent): Let $\mathbb{A}$ be set of all agents in the system and an agent represents any hybrid automaton behaviour participates in the system, where each agent has his own knowledge.

Definition-2 (Knowledge of Agent): Let $k_m^{\,i}$ means that the knowledge of agent m at state i, where $m \in \mathbb{A}$ is an agent, $1 \leq i \leq n$, n is an integer number.

Agent gain knowledge while moving from state to another and for the knowledge value to be true at each state there must be an epistemic relation. In the following we define the epistemic relation.

Definition-3 (Epistemic Relation): Let $\Gamma \subseteq \mathbb{A}$ is a subset of agents and $m \in \mathbb{A}$ is an agent, the symbol $\sim_m$ means that there is epistemic relation between states that makes agent m knows $\emptyset$. the symbol $\sim_{\Gamma^C}$ means there is epistemic relation between states that makes all agents in Γ have common knowledge $\emptyset$, the symbol C stands for "Common knowledge" in epistemic logic. The epistemic relation here is supposed to be equivalence relation.
In the following we use the same definitions for constraints that may appear as guard on states as in [12].

Definition-4 (Linear Constraints): Let $\mathbb{X}$ be a set of n real variables and $W = \sum_{i=1}^{n} s_i . x_i$ with $x_i \in \mathbb{X}$ be a linear combination of variables from $\mathbb{X}$, where $1 \leq i \leq n$, $s_i \in \mathbb{R}$, a set of linear constraints $\phi(\mathbb{X})$ over $\mathbb{X}$ with the typical element ϕ is defined as follows:

$$\phi ::= w \sim b \mid \phi_1 \wedge \phi_2 \mid \text{true}$$

Where $b \in \mathbb{R}$, $\sim \in \{ \leq, \geq, =, >, < \}$ and $\phi_1, \phi_2 \in \phi(\mathbb{X})$.
To represent the continuous behaviors of MASs we use differential equations whose solutions described as continuous functions in time. In the following, we define the dynamic constraints.

Definition-5 (Dynamical Constraints): Let $\mathbb{X}$ be a set of n real, with a typical elements $x_i \in \mathbb{X}$, $\dot{\mathbb{X}}$ be a set of first derivation of the variables of $\mathbb{X}$ with a special element $\dot{x}_i \in \mathbb{X}$. A set of $\mathbb{D}(\mathbb{X} \cup \dot{\mathbb{X}})$ with typical element d is defined as:

$$d ::= \dot{x} \sim c \mid \dot{x} \sim a . x \mid d_1 \wedge d_2 \mid \text{true}$$

Where $a \neq 0, c \in \mathbb{X}$, $\sim \in \{ \leq, \geq, = \}$ and $d_1, d_2 \in \mathbb{D}(\mathbb{X} \cup \dot{\mathbb{X}})$.
In real live agents can be in a situation that forces them to take a decision, so we allow agent to take decision by adding some preferences we use the same definition as in [8].

Definition-6 (Preferences): Let $q \in Q$ be a location, whose preference locations $\{q_1, q_2 \ldots q_n\}$ with respective utilities $\{\mu_1, \mu_2 \ldots \mu_n\}$, we say that q_1 is the

preference location for q if $\mu_1 = MAX\{\mu_1, \mu_2 \dots \mu_n\}$, note that preference can be not used when agent doesn't need to take a decision.

Now we are going to present our EPH basic components.

4.1.1 Epistemic Hybrid Automaton Components (EPH)

Epistemic hybrid automaton is a tuple of:
$EPH = (Var, Q, Flow, Inv, E, Jump, Assgn, Event, \Upsilon, \sim_a, N_a)$, consists of the following components:

- Var: $\mathbb{X} \cup \mathbb{A}$ a finite set of variables, $\mathbb{X} = \{x_1, x_2 \dots x_n\}$of real values that represent the continuous dynamics, and A is the auxiliary variable that used to measure the performance to take a decision.
- Q: A finite set of locations.
- $Flow$: Is a labeling function that assign to each control location $q \in Q$ dynamical constraint $Flow(q)$ whose free variables are from $\mathbb{X} \cup \dot{\mathbb{X}}$.
- Inv: $Q \to \phi(\mathbb{X})$ Is a function that assigns a linear constraint $Inv(q)$ to each location $q \in Q$.
- $E \subseteq Q \times Q$ *is a finite set of discrete transition, each transition* $(q_1, q_2) \in E$ has a source location q_1 and target location q_2.
- $Jump$: $Q \to \phi(\mathbb{X})$ a function assign to each transition $e \in E$ a constraint $Jump(e)$, which must hold on fire.
- $Assgn$: $\mathbb{X} \cup \mathbb{A} \to \mathbb{X}$ is a mapping function which assigns a real value to each variable , which reset the variables before the control of EPH goes from location q_1 to location q_2.
- $Event$: is a function that assign an event $Event(e)$ to each transition.
- Υ: $E \to R$ is the cost function that capture the preference of agent over e.
- $\sim_a \subseteq Q \times Q$ is an epistemic accessibility relation $\sim_a$ for each agent $a \in \mathbb{A}$.
- N_a: is the agent knowledge updates function, where $a \in \mathbb{A}$.

Note: the auxiliary variable can be omitted from the EPH when the agent doesn't need to take a decision, and so the cost function.

4.2 Semantic

Epistemic hybrid automaton consists of a set of locations, and to make agent move from a location to another we need to know how to evaluate the constraints that guard each one and what is meant by state. We will use the same evaluation of linear and continuous constraint in [12].

4.2.1 Evaluation of linear constraints

Let $\varphi \in \phi(\mathbb{X})$ be constraint and $v \in \mathbb{X}^n$ be the valuation of variable $\mathbb{X}$, then we write $v \vDash \varphi$, if v satisfies the constraint φ, which is defined inductively as:

$$\varphi = true \quad \&\& \quad \varphi = \sum_{i=1}^{n} s_i.x_i \sim c \ iff \sum_{i=1}^{n} s_i.v_i \sim c$$

$$\varphi_1 \wedge \varphi_2 iff \ \varphi_1 \vDash v_1 \ and \ \varphi_2 \vDash v_2$$

4.2.2 Evaluation of dynamical constraints

Let $\mathbb{D}(\mathbb{X} \cup \dot{\mathbb{X}})$ be a dynamical constraints and $f: \mathbb{R}^{\geq 0} \rightarrow \mathbb{R}^n$differentiable function and then we write $f \vDash_* d$ if f satisfies the dynamical constraint d, which is defined inductively as:

$d = true \quad \&\& \quad d = \dot{x} \sim c \ iff \ f'(t) \sim c$ holds

$d = \dot{x} + s.x \sim c \ iff \ f'(t) + s.f(x) \sim c$ holds

$d_1 \wedge d_2 iff \ f \vDash_* d_1 \ and \ f \vDash_* d_2$

where $f'(t)$ is the differentiation of the function f for $t \in \mathbb{R}^{\geq 0}$

4.2.3 Evaluation of knowledge of agent

Let $\mathbb{A}$ be the set of all agents and ${k_a}^i$ is the knowledge of a at state i, where $a \in \mathbb{A}, 1 \leq i \leq n$ and then we write $N_a \vDash {k_a}^i$ which is defined inductively as:

${k_a}^i = \omega$ iff $N_a \vDash \omega$.

4.2.4 State

The state of agent $m \in \mathbb{A}$ is given by ${\sigma_m}^i = \langle q_i, v_i, {k_m}^i, t \rangle$, where $q_i \in Q$, v_i is the valuation of real values, ${k_m}^i$ is the knowledge of agent m at state i and t is the time. A state ${\sigma_m}^i = \langle q_i, v_i, {k_m}^i, t \rangle$ is admissible if $v \vDash inv(q_i)$ and between each two states an epistemic relation.

The semantics of epistemic hybrid automaton is defined as hybrid automaton in terms of a labeled transition system between states. Transitions between states have two kinds: continuous transitions, and discrete transitions. Both of them are defined as follows:

4.2.5 Operational Semantics

A transition rule between two admissible states state ${\sigma_m}^1 = \langle q_1, v_1, {k_m}^1, t_1 \rangle$ and ${\sigma_m}^2 = \langle q_2, v_2, {k_m}^2, t_2 \rangle$:

4.2.6 Discrete transition

If $e = (q_1, q_2) \in E$, $t_1 = t_2$; $v_1 \vDash Jump(e)$, $N_m \vDash {k_m}^2$, such that ${k_m}^2$, is the new knowledge that added to ${k_m}^1$ and $v_2 \vDash Inv(q_2)$, such that v_2 is the valuation coming from $Assgn(e, \mathbb{X})$, in this case $Event(e)$ occurs.

4.2.7 Continuous transition

if $q_1 = q_2$, $(t_1 - t_2) > 0$ is the time elapsed at location q_1, there is a differentiable function f with f $\vDash_*$Flow(q_1), $N_m(t_1) \vDash {k_m}^1$, $N_m(t_2) \vDash {k_m}^2$, m ; $f(t_1) \vDash$ v1;

$f(t_2) \models v_2$ and for all $t \in [t_1, t_2]$; $f(t) \models Inv(q_1)$. An execution of epistemic hybrid automaton is a sequence of transitions from one state to another. For this purpose, we use the same definition presented in [12], where the execution path defined as follows:

4.3 Execution Path

A path of EPH is a finite or infinite sequence of states, $\rho = {\sigma_m}^0, {\sigma_m}^1, {\sigma_m}^2 \ldots$, where the transition between ${\sigma_m}^i$ and ${\sigma_m}^{i+1}$ for $i \geq 1$ is related either by a discrete or a continuous transition. The set of all possible paths of EPH is denoted as $\Pi(EPH)$, the execution path of EPH is a path ρ starting with the initial state ${\sigma_m}^0$.

The continuous change of states in a path ρ generates an infinite number of reachable states. Therefore, we use mathematical intervals to represent these states; we call this symbolic representation mathematical path interval, which is defined as follows:

4.3.1 Path Interval

Given a path $\rho \in \Pi(EPH)$, a sub sequence of admissible states $\Omega = ({\sigma_m}^{i+1}, \ldots, {\sigma_m}^{i+j}) \subseteq \rho$. Path interval $\Omega = (q, V, K, T))$, where K is the tuple of knowledge at the interval, V is the tuple of interval valuation of the variables during the time interval T and $t_{1+1} \leq T \leq t_{1+i}$. Now we will see the run of EPH where the change from one interval to another is fired by using discrete steps, we also use the same definition presented in [12].

4.3.2 Run

A run of EPH is $\rho EPH = \Omega_0, a_1, \Omega_1, a_2, \ldots$, a where a transition from interval $\Omega_i \xrightarrow[t_{i+1}]{a_{i+1}} \Omega_{i+1}$ where $a_{i+1} \in Event$ is the generated event before the control goes to Ω_{i+1}.

Interval is said to be reached if it reached from the initial state, we define reachability as in [12] as follows:

4.3.3 Reachability

Interval Ω_i in a path ρEPH is said to be reachable if $\Omega_i \in \rho EPH$ and also a state ${\sigma_m}^i$ is said to be reachable if ${\sigma_m}^i \in \Omega_i$.

4.4 Parallel Composition

Our model is composed of a set of agents each agent has his own behavior, they communicate by means of parallel composition [4], by parallel composition we mean agents interact in terms of action (event) compositions, if we have two

epistemic hybrid automatons: EPH_1, EPH_2 where the first interval $\Omega_1=\langle q_1, v_1, {k_{EPH1}}^j, T_1 \rangle$ came from EPH_1 and $\Omega_1=\langle q_2, v_2, {k_{EPH2}}^j, T_2 \rangle$ came from EPH_2 we represent that we have two compound intervals by, $\Lambda = (({q_1, v_1, {k_{EPH1}}^j}), (q_2, v_2, {k_{EPH2}}^j), T)$ which simply represents the reached intervals at both location q_1 and q_2 The execution path between two compound intervals is represented using transition from the first compound interval to the second compound interval $\Lambda_1 = (\Omega_1, \psi_1, T_1), \Lambda_2 = (\Omega_2, \psi_2, T_2)$, the interaction between the two automatons have two types:

- Synchronized Event: synchronized event define that two or more agents fire the same event simultaneous. l is a joint event $\Omega_1 \frac{l}{t} \Omega_2, \psi_1 \frac{l}{t} \psi_2$ we say that Ω_1 is synchronized with ψ_1.
- Interleaved Event: that type is allowing some agents to fire events while others are idle.

In our model the agents update their knowledge from the other agents when there are synchronized events occurring between two or more agents.
Now we need to present our new logic epistemic hybrid tree logic (EHTL), which is interpreted over epistemic hybrid automaton.

5 Epistemic Hybrid Tree Logic (EHTL)

Now we are going to present a new logic called epistemic hybrid tree logic that composes of real time logic and epistemic logic, EHTL is interpreted over all possible generated intervals from the epistemic hybrid automata, where agent moves from location to another according to the timed constraints on each state, the fired event that causes the discrete steps and the gained knowledge, first we are going to define the meaning of timed variables which taken from [10], then the syntax and semantics of EHTL as follows:

5.1 Timed Variables

Let $\mathbb{T}$ be a set of non-negative real variables called timed variables and $\phi(\mathbb{T})$ be a set of linear constraints over T. The valuation V of timed variables $\mathbb{T}$ is a function $\xi:\mathbb{T} \rightarrow \mathbb{R}^{\geq 0}$, given that $\pi \in \phi(\mathbb{T})$ we write $\xi \vDash \pi$ if π satisfies ξ.

5.2 EHTL Syntax

Let $\mathbb{X}$ be a set of real variables, $\mathbb{T}$ be a set of non-negative real variables disjoint from $\mathbb{X}$, $\phi(\mathbb{X})$ and $\phi(\mathbb{T})$ be sets of linear constraints with free variables from $\mathbb{X}$ and $\mathbb{T}$. Δ be a set of atomic proposition denoting the locations, Event be a set of atomic proposition denoting disjoint from Δ and $\mathbb{A}$ be a set of all agents.

5.2.1 Formulas of EHTL

$\varphi ::= p \mid ev \mid \phi \mid \varphi_1 \wedge \varphi_2 \mid \neg\varphi \mid x.\varphi \mid \pi \mid \forall(\varphi_1 \sqcup \varphi_2) \mid \exists(\varphi_1 \sqcup \varphi_2) \mid K_m\varphi \mid C_\Gamma\varphi$

Where $p \in \Delta$, $a \in$ Event, $x \in \mathbb{T}, m \in \Gamma$ where $\Gamma \subseteq \mathbb{A}$, φ_1, φ_2 are EHTL formulas and the logical connectives $\wedge$ have their usual meaning.

5.3 Semantics of EHTL

We will interpret the formulas of as EHTL over the set of all possible intervals [10] generated from possible runs of EPH. Where the interval $\Omega = (q, V, K, T) \in \Pi_{EPH}$ and ξ is the valuation function of timed variables. The satisfaction relation $(\Pi_{EPH}, \Omega) \models_\xi^T \varphi$ means that φ is satisfied in the interval Ω within the time T, for some evaluation functions ξ is defined as:

- $(\Pi_{EPH}, \Omega) \models_\xi^T p$ iff $p = v(\Omega)$.
- $(\Pi_{EPH}, \Omega) \models_\xi^T ev$ iff there is Ω' and t'$\in T$ with $\Omega \xrightarrow[t']{ev} \Omega'$.
- $(\Pi_{EPH}, \Omega) \models_\xi^T \phi$ iff there is $\beta \subseteq \Omega$ for each ${\sigma_m}^i \in \beta, {\sigma_m}^i \models \phi$.
- $(\Pi_{EPH}, \Omega) \models_\xi^T \varphi_1 \wedge \varphi_2$ if $(\Pi_{EPH}, \Omega) \models_\xi^T \varphi_1$ and $(\Pi_{EPH}, \Omega) \models_\xi^T \varphi_2$
- $(\Pi_{EPH}, \Omega) \models_\xi^T \neg\varphi$ *iff* $(\Pi_{EPH}, \Omega) \not\models_\xi^T \varphi$.
- $(\Pi_{EPH}, \Omega) \models_\xi^T x.\varphi$ iff there is t $\in T$ such that $\xi(x) = t$, $(\Pi_{EPH}, \Omega) \models_\xi^{T=t} \varphi$.
- $(\Pi_{EPH}, \Omega) \models_\xi^T \pi$ iff $\xi \models \pi$.
- $(\Pi_{EPH}, \Omega) \models_\xi^T \forall(\varphi_1 \sqcup \varphi_2)$ iff for every run $\Pi \in \Pi_{EPH}$, $\Pi = \Omega_0, \Omega_1, \ldots, \Omega$, for $j \geq 0$ we have $(\Pi_{EPH}, \Omega_j) \models_\xi^T \varphi_2$ and $(\Pi_{EPH}, \Omega_k) \models_\xi^T \varphi_1$ where $0 \leq k < j$.
- $(\Pi_{EPH}, \Omega) \models_\xi^T \exists(\varphi_1 \sqcup \varphi_2)$ iff there exist run $\Pi \in \Pi_{EPH}$, $\Pi = \Omega_0, \Omega_1, \ldots, \Omega$, for $j \geq 0$ we have $(\Pi_{EPH}, \Omega_j) \models_\xi^T \varphi_2$ and $(\Pi_{EPH}, \Omega_k) \models_\xi^T \varphi_1$ where $0 \leq k < j$.
- $(\Pi_{EPH}, \Omega) \models_\xi^T K_m\varphi$ iff for all Ω' and t'$\in T$ such that $\Omega \sim_m \Omega'$, ${K_m}^\Omega \models \varphi$, then $(\Pi_{EPH}, \Omega') \models_\xi^{T=t'} K_m\varphi$.
- $(\Pi_{EPH}, \Omega) \models_\xi^T C_\Gamma\varphi$ iff for all Ω' such that $\Omega \sim_\Gamma^C \Omega'$, *then* $(\Pi_H, \Omega') \models_\varepsilon^{T'} \varphi$.

The quantifiers here $\exists$, $\forall$ called path quantifiers $(\Pi_{EPH}, \Omega) \models_\xi^T x.\varphi$ means that φ holds at time T, $p, \varphi_1, \varphi_2 \in \Delta$ and $(\Pi_{EPH}, \Omega) \models_\xi^T K_m\varphi$ means that agent m knows φ at time T.

Such model will help in real time systems that need decision making, because such systems sited in a dynamic environment that may causes errors appearance or changing of the system objective because of the new collected knowledge. In the following we present the properties we need to take into account when building our model.

6 Quantitative and Knowledge Requirements

As presented in [10] safety property is essential to guarantee that system will never reach a bad state, and liveness property state that something good will always happened, but these properties will always be falsified since you can't guarantee what time the agent will keep in a single state. We also use the **bounded interval** to indicate that a specific event occurs within a specific period of time, for-example at provider automaton $to-target \rightarrow \forall\Box(help.t_1 \rightarrow \forall \Diamond (rescue.t_2 \wedge t_2 \leq (t_1 + 2h))$. This property can be used to represent safety property but at some times we don't need to use event interval so we use **bounded invariance** interval it state that once an event occurred a specific formula will happened for a specific period of time as stating that $to-target \rightarrow \forall\Box(help.t_1 \rightarrow \forall\Box(rescue.t_2 \wedge t_2 \leq (t_1 + 2h))$, this property also can be used to represent liveness property. Knowledge affects in agent decision making, knowledge is supposed to be veridical ‹veridicality means what is known is true. Which means that agents don't know the falsities, the sentences in the knowledge base are supposed to be true, to guarantee the truth of knowledge we use $S5_m$ which is often chosen as the standard logic of *knowledge,* since it composed from KT45 axioms ($K_i\,\varphi \wedge K_i(\varphi \rightarrow \beta) \rightarrow K_i\,\beta$, $K_i\varphi \rightarrow \varphi$, $K_i\varphi \rightarrow K_i\,K_i\varphi$, $\neg K_i\varphi \rightarrow K_i\,\neg K_i\varphi$) it combines the relation constraints of reflexivity, transitivity, and Euclidicty. These combinations of constraints lead to the equivalence relation. System $S5_m$ extends its power from that epistemic logic axiom systems viewed as measures of infallibility and replies to skepticism, where the positive introspection axiom which mean knowing of your knowledge is a way of blocking skepticism, and negative introspection which mean knowledge of your ignorance is a sort of closed world assumption which leads skepticism nowhere to go.

7 Conclusion

Table. 1. Comparison between related work and presented work.

	Representing knowledge	Representing time constraints	Representing dynamic and discrete behaviour	Decision Making
9 [Mohalik. S and Ramanujam. R]	✓			
10 [Mohammed. A and Furbach. U]		✓	✓	✓
11, 12 [Mohammed. A and Furbach. U]		✓	✓	
18 [Van.H and Wooldridge. M]	✓			
21 [Wozna. B and Lomuscio.A]	✓	✓		
Our Work	✓	✓	✓	✓

This paper presented epistemic hybrid tree logic, which extends real time logic with epistemic logic to allow agent to reason about knowledge at the same framework of time and to be able to take decision according to the knowledge agent acquired, we use a verification model epistemic hybrid automata that allows us to represent the continuous and discrete behavior of the agent using event approach and state based approach. Our system is composed of a set of epistemic hybrid automaton each automaton updates its knowledge when there is synchronized event between some automatons. The execution of the system is interpreted over a set of intervals and we propose a comparison to clarify the difference between our work and related work shown in **Table. 1**. As a future work we can add dynamic logic as in [3,5] to allow to represent agent's actions or add probabilities so we can take more decisions.

References

[1] Alexander. C and Michael. Z., *Modal Logic.* Oxford University Press. 1997.

[2] Alur. R and Henzinger. T. A *Logics and models of real time: A survey In Real-Time*: Theory in Practice, pages 74106. Springer, 1992.

[3] Aucher, Guillaume., *Dynamic epistemic logic as a substructural logic.,* Logic and Interactive Rationality., (3), 2014.

[4] Bornot. S and Sifakis. J *On the composition of hybrid systems. In Hybrid systems: computation and control,* pages 49-63. Springer, 1998. University of Chicago Press, 1988.

[5] Bryan. R, Joshuaand. S, Audrey. Y, *Dynamic epistemic temporal logic.,* Logic, Rationality, and Interaction., 263-277,Springer, 2009.

[6] Henzinger T. A., *The theory of hybrid automata.* Proceedings of the 11th Annual Symposium on Logic in Computer Science, New Brunswick, NJ, pp. 278–292. IEEE Computer Society Press (1996).

[7] Hutzler. G, Klaudel. H, Wang. D., *Towards Timed Automata and Multi-agent Systems*. In: Hinchey, M.G., Rash, J.L., Truszkowski, W.F., Rouff, C.A. (eds.) FAABS 2004. LNCS (LNAI), vol. 3228, pp. 161–172. Springer, (2004)

[8] Mordechai. B, Amir. P and Zohar. M. *The temporal logic of branching time,* Acta informatica, 20(3), 207-226,Springer, 1983.

[9] Mohalik. S and Ramanujam. R., *Automata for epistemic temporal logic with synchronous communication.* Journal of Logic, Language and Information, 19(4):451-484, 2010.

[10] Mohammed. A and Furbach. U., *From reactive to deliberative multi-agent planning.* In MSVVEIS, pages 67-75,Springer, 2009.

[11] Mohammed.A and Furbach. U., *Multi-agent systems: Modeling and verication using hybrid automata.* Springer, 2010.

[12] Mohammed.A and Furbach. U., Mas: *qualitative and quantitative reasoning.* In Programming Multi-Agent Systems, pages 114-132. Springer,2012.
[13] Pnueli. A., *The temporal logic of programs.* In Foundations of Computer Science, 1977., 18th Annual Symposium on, pages 46-57. IEEE, 1977.
[14] Pnueli. A., *Applications of temporal logic to the specification and verification of reactive systems*: a survey of current trends. Springer, 1986.
[15] Ruan. J., *Reasoning about time, action and knowledge in multi-agent systems.* KI-K unstliche Intelligenz, 25(1) : 75-76, 2011.
[16] Sim. K. M., *Epistemic logic and logical omniscience: A survey.* International Journal of Intelligent Systems, 12(1):57-81, 1997.
[17] Van. B, Johan, Pacuit, Eric. *Temporal Logics of Agency.* Journal of Logic, Language and Information. 9(4).389-393.2010
[18] Van.H and Wooldridge. M., *Cooperation, knowledge, and time*: Alternating-time temporal epistemic logic and its applications. Studia Logica, 75(1):125-157, 2003.
[19] Van. H and Wiebe, *Epistemic logic for AI and computer science,* Cambridge University,41(7),2004
[20] Verbrugge, H.van., *Epistemic logic: A survey,* Game theory and applications, 8(45),8-53, 2002, Nova Publishers.
[21] Wozna. B and Lomuscio.A., *A logic for knowledge, correctness, and real time.* In Computational Logic in Multi-Agent Systems, pages 1-15.Springer, 2005.

Hybrid Model for Web Pages Navigation Prediction

No'aman M. Abo Al-Yazeed *, Ahmed M. Gadallah **, Hesham A. Hefny **

Abstract:

Nowadays, users rely on the web for information. Accordingly, web usage mining has become one important subject of research. Such research area covers Web-based personalized services, prediction of user near future intentions, adaptive Web sites, and customer profiling. Web page Prediction is strongly limited by the nature of web logs, the intrinsic complexity of the problem and the tight efficiency requirements. This paper proposes a web usage mining approach to predict next page candidate to be accessed based on a hybrid page ranking model. The proposed approach combines between two page ranking approaches. The first approach computes frequency ratio indicating the number of occurrences for each page in the search result. On the other hand, second approach computes the coverage ratio from similar behavior patterns. As a result of the proposed approach a set of candidate pages are ranked and the page with highest rate is recommended. The proposed approach has been tested on real data extracted from the web server log file of a university website.

Keywords: Web Mining, Web Usage Mining, Web-based recommendation systems, Navigation pattern mining, Web Log, Web Personalization, Adaptive Web Sites.

I. INTRODUCTION

The volume of information available on the internet is increasing rapidly with the explosive growth of the World Wide Web and the advent of e-Commerce. While users are provided with more information and service options, it has become more difficult for them to find the "right" or "interesting" information, the problem commonly known as information overload [1]. Commonly, modeling of human behavior is one of the major challenges of computer science in the new century. Also, web mining becomes essential not only to better understand human behavior, but also to allow completely dynamic restructuring of contents and adaptive navigation of the Internet. The explosive

*Computer Science Dept.,ISSR, Cairo Univ.ahmgad10@yahoo.com, hehefny@hotmail.com
** Information Systems Dept., ISSR, Cairo Univ. no3man87_2010@yahoo.com

growth of web users data, in the form of log files, has produced an increasing interest of researchers in the last decade on web usage mining [2].

Recommender systems [3] are alternative, user-central, approaches to handle the problem of information overload by adapting the content and structure of websites respecting the users' needs by taking advantage of the knowledge acquired from the analysis of the users access behaviors stored in the web server log files. Generally, recommender systems guide users toward interesting or useful objects in a large space of possible options [4].

The rest of this paper is organized as follows: section 2 introduces some of related works. A brief overview about Markov Model is given in section 3. Section 4 presents the proposed approach. An illustrative case study is outlined in section 5. Finally, section 6 presents the conclusions and the future work.

II. Related Works

In recent years, there has been a growing interest in applying web usage mining techniques to build web recommender systems [5-7]. Traditional techniques, which mainly depends on user ratings on different items or other explicit feedbacks provided by the user and then recommend a set of items considered to be of interest to the user [8, 9]. Web server access logs record the users browsing history which contains abundance of hidden information regarding users and their navigations. Web usage recommender systems take web server access logs as input, and make use of data mining techniques such as association rules and clustering to extract implicit, and potentially useful navigational patterns, which are then used to provide recommendations. They could, therefore, be a good alternative to the explicit user rating or feedback in deriving user models. Such techniques discover user preferences from their implicit feedbacks, namely the web pages they have visited. Clustering and collaborative filtering approaches are ready to incorporate both binary and non-binary weights of pages, although binary weights are usually used for computing efficiency [10].

On the other hand, association rule (AR) mining can lead to higher recommendation precision and are easy to scale to large datasets [11]. Yet, incorporating page weight into AR models has not been explored in previous studies. The importance of Web usage mining has led to a number of research papers in the area. However, most of these papers were hindered by some kind of limitations[1, 12]. Different combinations of mining techniques were already

suggested for web access recommendation. Commonly, many approaches were introduced to achieve such recommendation based on Markov model which is the most common one used to model user's web navigation sessions. Markov model was extended to Lower-order Markov model and High-order Markov model. Lower-order Markov model provides high coverage, but with low accuracy. On the other hand, Higher-order Markov model give low coverage but high accuracy with more time complexity.

R. Bhushan and R. Nath,[13], introduced a new model based on learning from web logs and recommends user a list of pages which are relevant to him respecting the user's historical pattern. Then, search result list is optimized by re-ranking the resulted pages. Also, D. Dhyani et al. [14], introduced another model based on Markov process for web access prediction. Yet, such model has the drawback of its high complexity due to considering all access sequences throughout the prediction process. Also, K. Han-Gyu et al. [15], introduced an approach of recommending contents across different categories by considering both of the semantic information of contents and user interests. Author's uses Linked Data as the source to find the appropriate semantics of the contents extracted from users' viewing history. The semantic concepts retrieved for the contents are then grouped together into semantic clusters based on their similarity and relevance. Finally they recommends potentially interesting contents to general users based on the content-consumption trends monitored from leading user groups who most proactively and frequently consume contents. Also, S. chimphlee et al.[16], introduced web access prediction model by integrating rough sets clustering with Markov model. It has major drawback that lack of prediction accuracy due to approximation while forming clusters. The possibility of an object for belonging to a cluster can reduce the cluster tightness, which in turn affects prediction accuracy. The sequential mining suggested in that work is all k-th order Markov model. F. Khalil et al.[17], In this paper author's proposed a method to integrate Markov model and association rules for predicting Web page accesses. The integration is based on a low order Markov model. Sets of subsequence association rules are used to complement the Markov model for resolving ambiguous predictions by using long history data. Also, A. Maratea and A. Petrosino, [18], authors designed a heuristic majority intelligence technique, which effortlessly adjusts to changing navigational patterns; with the low cost explicitly individuate them ahead of navigation. The proposed technique imitates human behavior in an unidentified environment in occurrence of several individuals working in parallel and it has the ability to predict with better accuracy and in real time the next page group visited by a user.

V. M. Rao and V. V. Kumari,[19], authors of this paper introduce a new approach to predict users browsing behavior at two levels to meet the nature of the navigation. One is category stage and the other is web page stage. In stage one is to predict category. The unnecessary categories can be excluded. The scope of calculation is massively reduced. Next, using pruned Markov models using higher order in the level two to predict the users browsing page is more effectively and high operational performance. On the other hand, A. Anitha, [20], introduced a new approach for next page access prediction. It uses a combined approach of integrating Markov model and clustering-based pair wise nearest neighbor technique. The resultant patterns are highly relevant, and the size data set that is utilized for sequential mining process is highly reduced. The drawback of this work is, loosely connected access sequences are not considered for mining process.

Additionally, M. Jalali et al. developed a recommendation system called WebPUM, [21]. It is an online prediction system using Web usage mining system and proposes a novel approach for classifying user navigation patterns to predict users' future intentions. The approach is based on the new graph partitioning algorithm to model user navigation patterns for the navigation patterns mining phase. Furthermore, longest common subsequence algorithm is used for classifying current user activities to predict user next movement.

III. Markov Model

Generally, Markov model represent one of the most common sequential pattern mining techniques used for next page access prediction [22]. Such model in general can achieve good prediction accuracy as they consider consecutive and sequential page accesses. The number of history pages considered for prediction process determines the order of Markov model. Markov model was extended to Lower-order Markov model and High-order Markov model. Lower-order Markov model provides high coverage, but with low accuracy. On the other hand, Higher-order Markov model give low coverage but high accuracy with more time complexity. In first order Markov model, each state represents a single web page and each pair of viewed pages corresponds to a state transition. Two artificial state, start and final, are incorporated in the model. In second-order Markov model each state corresponds to sequence of two viewed web pages and so on. In Markov model, state space increases exponentially with the order of the model. Let the "m" is number of web pages and "n" is order of model then the number of states will be "m^n".

IV. The Proposed Hybrid Approach for Web Navigation Prediction

The purpose of the proposed method is to give an efficient prediction algorithm that may be used for online recommendations to users of a website. The proposed approach is divided into two main phases: *web mining phase*, and *prediction phase*.

A. Web Access Log Mining Phase

Generally speaking, web access log mining process starts with parsing the web access log file and import its interesting data into a relational database. Consequently, a preparation process takes place on the imported data as shown in Fig.1. Such preparation phase consists of the following steps:

- Data Cleaning and Filtering for irrelevant data which is not appropriate for the mining process.
- Identify each distinct user by considering each IP address represents one user. For more logs, if the IP address is the same, but the log shows a change in user agent, an IP address represents a different user. Using the access log in conjunction with the referrer logs and site topology to construct browsing paths for each user; if requested page is not directly reachable by a hyperlink from any of the pages visited by the user, there is another user with the same IP address.
- Creating user sessions which is considered one of the major steps in data preparation techniques, in this step the page accesses of each user is divided into individual sessions according to a time threshold *(in our experiment 30 minutes)*, If the time between page requests exceeds a certain limit, it is assumed that the user is starting a new session.
- Path Completion was done to fill in these missing references page that are not recorded in the access log by using the site topology and link structure in site.
- Page Ranking is the next step in web mining phase, this step concerns with creating a relation that contains each web page, number of times the page was visited "Frequency" (as mentioned in equation No.1), computing percentage of it "Total Frequency Ratio" (See equation No.2), and mean stay time in each page "Mean Duration" (See equation No.3). Accordingly, such relation can be used to identify pages with high and low request and report it to web server admin to rearrange web site according to this statistics.

Finally, basket transformation method [23], is applied to web data by transforming log data to the most appropriate format for mining purposes. Yet, not all sessions and web pages are considered. That is, sessions with visit actions less than 2 pages and web pages with low frequency are excluded. Consequently, certain thresholds are set to eliminate search space and increase the recommendation accuracy.

Page Frequency: $F = Counting\ Number\ of\ page\ Request$ (1)

Total Frequency Ratio: $TF = \frac{f}{\sum f1+f2.....+fn}$ (2)

Mean Duration: $MD = \frac{\sum D1+D2.....+Df}{f}$ (3)

B. The Prediction Phase

- In this phase the proposed prediction algorithm is applied on the prepared data. The inputs of this algorithm are the prepared web access log, while the output represents a recommended web page which appropriate for current user navigation behavior.
- The algorithm starts with receiving a set of web pages constructing a query sequence, originally we look at prepared data from phase one, searching for sessions which contain one or more pages from query sequence as its visits action.
- Next, we select each distinct web page from search result and then exclude web pages included in query sequence.
- For each web pages resulted in previous step, Counting number of times each one was requested "*r*" and divide it by total number of pages request in search result "*Frequency Ratio*" (Using equation No.4).
- Counting number of times query sequence requested together in search result "*Base*" (See equation No.5), If no matched result *(Base = 0)* algorithm exclude first page in query sequence *(L.- 1)*, and repeat until *(Base > 0)*.
- For each distinct web page, count number of times current page requested after query sequence have been requested "*Support*" (See equation No.6).
- The next step is to calculate "*Coverage Ratio (CR)*" (Using equation No.7).

- Finally a recommendation relation is constructed which contain candidate pages sorted by *Coverage Ratio, Frequency Ratio*. If two or more pages have identical rank *Total Frequency Ratio* are referenced to re-rank pages; top page are selected and recommended to user.

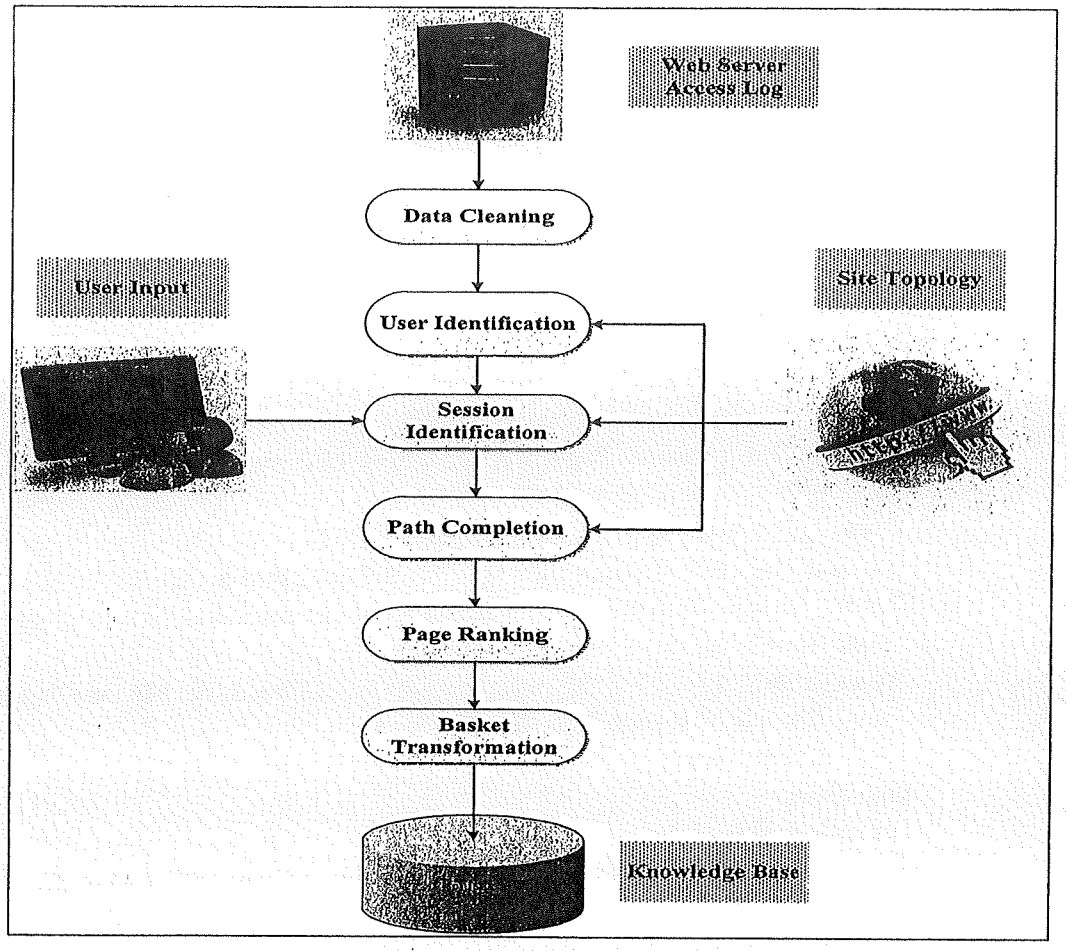


Fig.1 Web Log Preprocessing Steps.

Frequency Ratio: $FR = \frac{Number\ of\ Times\ Current\ Page\ was\ Visited}{Total\ Number\ of\ Pages\ Request}$ (4)

Base: $B = Counting\ Number\ of\ Query\ Sequence\ Request$ (5)

Support: $S = Counting\ Number\ of\ Current\ Page\ Requests\ while\ Query\ Sequence\ were\ Requested$ (6)

Coverage Ratio: $CR = \frac{S}{B}$ (7)

Algorithm 1. Candidate page prediction algorithm.

T: Transformed Log Relation.
P: Page Rank Relation.
C: Candidate Pages Relation.
S: Query Sequence "Sequence of Web Pages to Predict their Next Action".
TO: Total Number of Occurrence.
Per: Percentage of Page Occurrence in Search Result.
Begin.
Input S.
For each Page x in S.
IF (T Does not Contain Page x)
Exclude Page x From "S".
End IF.
End For.
IF (S.Length >= 1)
Select Sessions From "T" Where Session Contains any Page in "S" As Visit Action. (1)
From (1) Select Distinct Pages and Number of Occurrence "OCC" And Insert in "C". (2)
From (2) For each Page Compute: TO = TO + OCC.
For each Page in (2) Compute: Per = OCC / TO and Insert in "C".
Declare Variable Base = 0. {Number of Occurrence of Pattern "S" in (1)}
Count "Base".
While (Base < 1) {There is No Matched Pattern}
Exclude First Page From "S": S = S – 1.
Count "Base".
End While.
For each Page in (1)
Count Number of Occurrence of Current Page Given "S" IS Visited. {Support}
Compute Coverage Ratio "CR": CR = Support / Base and Insert in "C".
End For.
Sort Pages in "C" By CR, Per.
Select Top Page.
Else:
Select Top Page From "P".
End IF
Recommend Selected Page.
End.

V. An Illustrative Case Study

For our experiments, the first step is data acquisition by gathering a log file from an active Web server. The log file we used as a data source for our experiments is a log file of a university website. Commonly, Log file have the following information: User IP address, user name, date and time and offset of request, request method, requested page, protocol version, status code, bytes transferred, referred page, user agent. Gathered data had to be cleaned and filtered, Cleaning the data involved removing erroneous and invalid pages, Deleting all requests that contain a status code is not equal to "200", which represents the successful request, and also delete all requests that contain a request method is not equal to "GET". Filtering and cleaning log files also includes deleting all requests that contain graphical extensions.

In consequent, there is a step to identify each user and divide user transactions into sessions according to taking a 30-minute timeout into consideration. The frequency of each page visited by the user was calculated. Short sessions were then removed and only pages with high frequency were considered. Sample training set chosen, the algorithm is tested in such a way that, by hiding the last page in every test session, a prediction is made. As observed from test results, shown in table.1, both of the proposed model and Markov models (First and Second Order) have low prediction accuracy in small portion of data set, while heuristic model and WebPum model achieve good predictions because of its nature of mimic human behavior in unknown environment.

Table.1 Test result accuracy of Markov, Heuristic, WebPUM and the proposed models.

Data Set	Test Set	Markov Model (1st Order)	Markov Model (2nd Order)	Heuristic Model	WebPUM Model	Page Rank Model
1000	300	33.33	29.33	43.00	53.33	40.00
2000	600	33.33	30.00	41.50	58.33	40.00
3000	900	33.33	30.00	48.78	51.11	50.00
4000	1200	35.83	35.00	51.83	52.75	52.50
5000	1500	35.67	36.00	53.53	52.00	58.00
6000	1800	41.11	41.67	55.50	50.83	73.33
7000	2100	43.81	48.57	56.10	51.90	82.86
8000	2400	47.92	59.58	55.25	52.50	85.00
9000	2700	48.15	60.00	60.89	52.70	85.56
10000	3000	53.33	60.00	62.53	52.77	86.00

Yet, such accuracy is got better as data set increases. This behavior can be interpreted because of loss of background for both models (proposed model

and markov model) due to small training set. While data set growing accuracy prediction growing also, but in Markov model complexity increase exponentially because considering all states and coverage decreased. Also, in heuristic model, and WebPUM prediction accuracy increases slightly. On the other hand, the proposed model prediction accuracy increased rapidly with less complexity because of its dynamicity and coverage also stable. From the test results shown in Fig.2 and Fig.3, we can observe that when using small sets of data both model achieve low accuracy, but while data set is growing our model accuracy is growing rapidly than traditional markov model, But, From Fig.4 and Fig.5 we can show that both heuristic and WebPUM model achieve moderate predictions with slightly increasing while navigations grown. From test result graphs we can say that the proposed model accuracy percentage is higher than Markov models and heuristic model accuracy percentage.

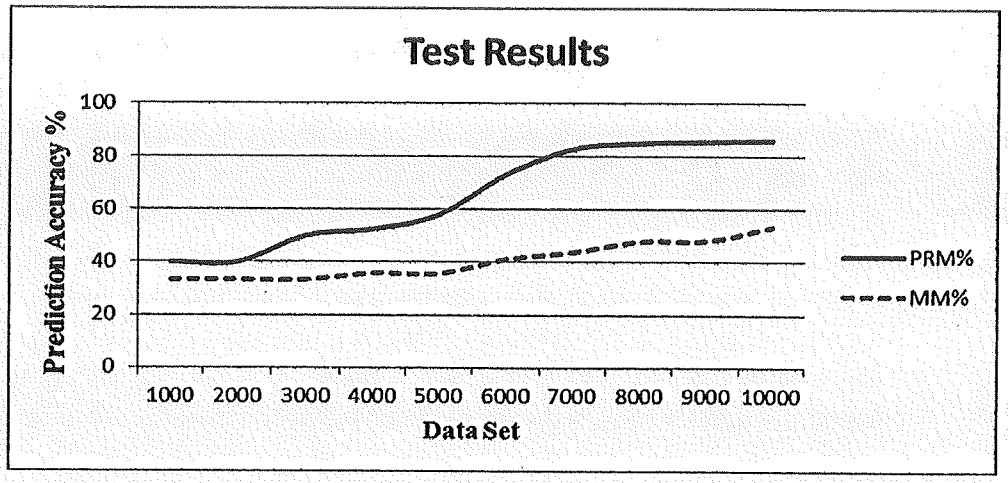


Fig.2 Prediction accuracy by the proposed model versus first order Markov model.

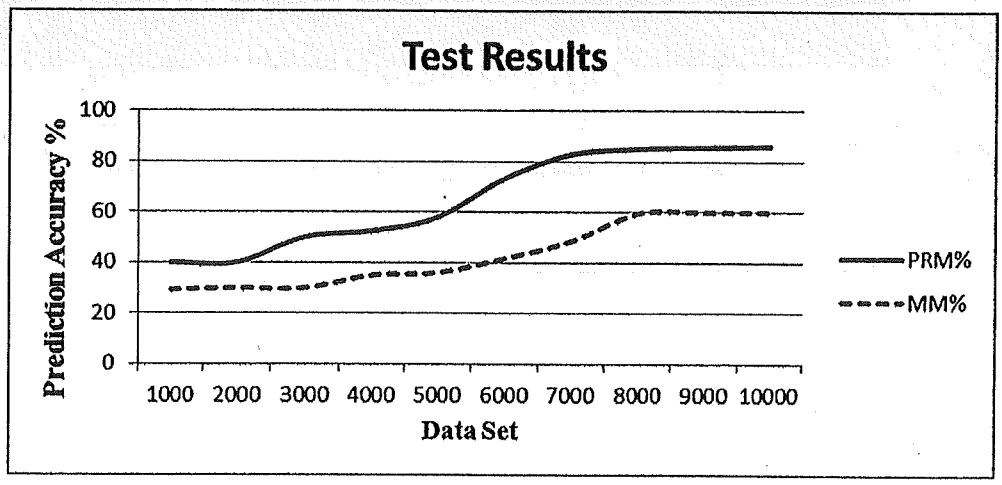


Fig.3 Prediction accuracy of the proposed model and second-order Markov model.

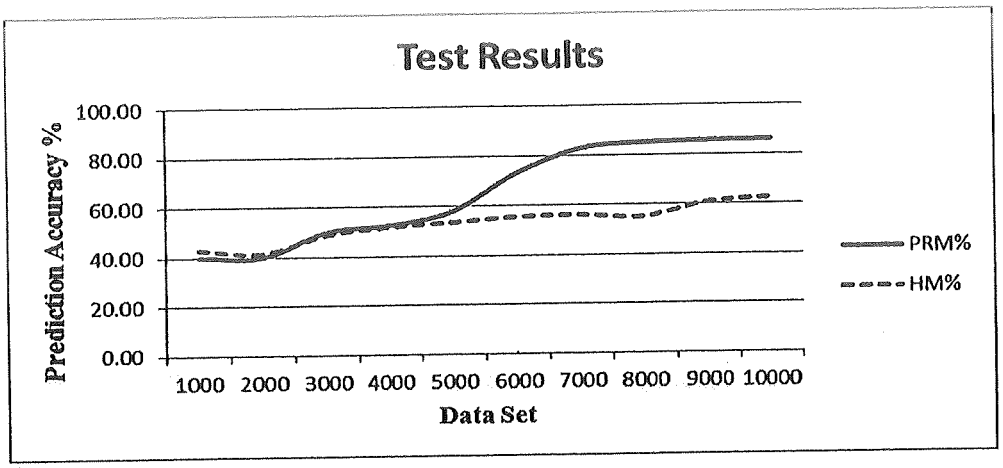


Fig.4 Prediction accuracy of the proposed model and the heuristic model.

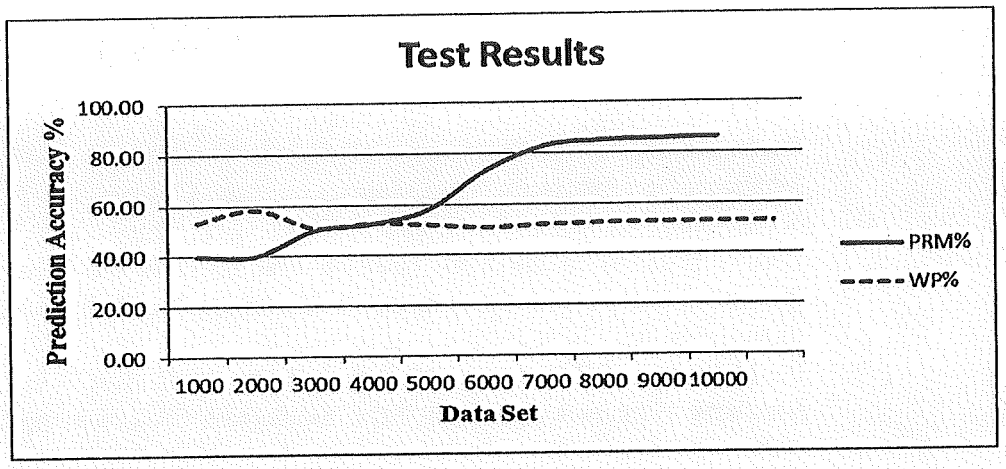


Fig.5 Prediction accuracy of the proposed model and WebPUM model.

VI. CONCLUSIONS

World Wide Web is growing rapidly with a huge number of users. Indeed, there is a need to facilitate web browsing which helps a user in his surfing session, and to engage users of a website at an early stage of surfing. Accordingly, the proposed approach for web access mining and recommendation is essential. So it is necessary to analyze the user web navigation behavior to improve the quality of web services, offered to the web user. Analysis of user web navigation behavior is achieved through modeling web navigation history.

In this paper the Hybrid model is proposed based on page ranking algorithm and it has been tested and compared with other models. As shown in the results of the illustrative case study, the proposed approach is more efficient than compared models especially with web log files including low numbers of

entries. In future, there is a plan to design an incremental algorithm for the model so that it does not start from scratch each time in web page prediction.

REFERENCE

[1] J. Vellingiri, and S. C. Pandian, "A Survey on Web Usage Mining," *Global Journal of Computer Science and Technology,* vol. 11, no. 4, pp. 66-72, USA, Mar. 2011.

[2] R. Cooley, B. Mobasher, and J. Srivastava, "Web Mining: Information and Pattern Discovery on the World Wide Web," In Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97), pp. 558-567, Nov. 1997.

[3] P. Resnick, and H. R. Varian, "Recommender Systems," *Communications of the ACM,* vol. 40, no. 3, pp. 56-58, Mar. 1997.

[4] R. Burke, "Hybrid Recommender Systems: Survey and Experiments," *User modeling and user-adapted interaction,* vol. 12, no. 4, pp. 331-370, Nov. 2002.

[5] X. Fu, J. Budzik, and K. J. Hammond, "Mining Navigation History for Recommendation," In Proceedings of the 5th International Conference on Intelligent User Interfaces, New Orleans, Louisiana, USA, pp. 106-112, 2000.

[6] W. Yi-Hung, C. Yong-Chuan, and A. L. P. Chen, "Enabling Personalized Recommendation on the Web Based on User Interests and Behaviors," In Proceedings of the 11th International Workshop on Research Issues in Data Engineering, pp. 17-24, 2001.

[7] B. Mobasher, "Web Usage Mining and Personalization," In Practical Handbook of Internet Computing, M. P. Singh (ed.), CRC Press, pp. 264-265, 2004.

[8] M. Deshpande, and G. Karypis, "Item-Based Top-N Recommendation Algorithms," *ACM Transactions on Information Systems,* vol. 22, no. 1, pp. 143-177, Jan. 2004.

[9] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl, "An Algorithmic Framework for Performing Collaborative Filtering," In Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '99), Berkeley, California, USA, pp. 230-237, 1999.

[10] M. Nakagawa, and B. Mobasher, "A Hybrid Web Personalization Model Based on Site Connectivity," In Proceedings of the 5th International WEBKDD Workshop at the ACM-SIGKDD Conference on Knowledge Discovery in Databases (KDD '03), pp. 59-70, Aug. 2003.

[11] W. Lin, S. A. Alvarez, and C. Ruiz, "Collaborative Recommendation via Adaptive Association Rule Mining," In Proceedings of the International Workshop on Web Mining for E-Commerce (WebKDD '2000), Boston, Aug. 2000.

[12] M. Géry, and H. Haddad, "Evaluation of Web Usage Mining Approaches for User's Next Request Prediction," In Proceedings of the 5th ACM International Workshop on Web Information and Data Management (WIDM '03), New Orleans, Louisiana, USA, pp. 74-81, Nov. 2003.

[13] R. Bhushan, and R. Nath, "Recommendation of Optimized Web Pages to Users Using Web Log Mining Techniques," In Proceedings of the 3rd IEEE International Advance Computing Conference (IACC), pp. 1030-1033, Feb. 2013.

[14] D. Dhyani, S. S. Bhowmick, and N. Wee-Keong, "Modelling and Predicting Web Page Accesses Using Markov Processes," In Proceedings of the14th International Workshop on Database and Expert Systems Applications, pp. 332-336, Sept. 2003.

[15] K. Han-Gyu, K. Eunae, K. In-Young, and C. Deokmoon, "Semantically-Based Recommendation by Using Semantic Clusters of Users' Viewing History," In Proceedings of the International Conference on Big Data and Smart Computing (BIGCOMP), pp. 83-87, Jan. 2014.

[16] S. Chimphlee, N. Salim, M. S. Ngadiman, W. Chimphlee, and S. Srinoy, "Rough Sets Clustering and Markov Model for Web Access Prediction," *In Proceedings of the Postgraduate Annual Research Seminar*, pp. 470-475, Malaysia, 2006.

[17] F. Khalil, J. Li, and H. Wang, "A Framework of Combining Markov Model with Association Rules for Predicting Web Page Accesses," In Proceedings of the 5th Australasian Conference on Data Mining and Analystics (AusDM '06), Sydney, Australia, pp. 177-184, Dec, 2006.

[18] A. Maratea, and A. Petrosino, "An Heuristic Approach to Page Recommendation in Web Usage Mining," In Proceedings of the 9th International Conference on Intelligent Systems Design and Applications (ISDA), pp. 1043-1048, 2009.

[19] V. M. Rao, and V. V. Kumari, "An Efficient Hybrid Successive Markov Model for Predicting Web User Usage Behavior using Web Usage Mining," *International Journal of Data Engineering,* vol. 1, no. 5, pp. 43-62, 2010.

[20] A. Anitha, "A New Web Usage Mining Approach for Next Page Access Prediction," *International Journal of Computer Applications (IJCA),* vol. 8, no. 11, pp. 7-10, Oct. 2010.

[21] M. Jalali, N. Mustapha, M. N. Sulaiman, and A. Mamat, "WebPUM: A Web-Based Recommendation System to Predict User Future Movements," *Expert Systems with Applications,* vol. 37, no. 9, pp. 6201-6212, Sept. 2010.

[22] A. Anitha, and N. Krishnan, "A Web Usage Mining Based Recommendation Model for Learning Management Systems," In Proceedings of the IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), pp. 1-4, Dec. 2010.

[23] Z. Markov, and D. T. Larose, "Preprocessing for Web Usage Mining," *Data Mining the Web: Uncovering Patterns in Web Content, Structure, and Usage,* D. T. Larose, ed., pp. 156-176: Wiley, 2007.

Human-Like Data-Sensitive fuzzy-Based Query Approach for Relational Databases

Adel A. Sabour, Ahmed M. Gadallah*, and Hesham Hefny***

Abstract

The user searching criteria may depend on the current state of the data stored in the searched database.Generally, users often have vague or imprecise criteria's when searching databases and thus may not know how to use traditional structured query language to precisely formulate queries that lead to satisfactory answers.Based on fuzzy set theory and data-sensitive fuzzy sets, this paper proposes a fuzzy query approach supporting data-sensitive fuzzy queries in which the definition of a fuzzy set can be dynamic and depend on some data in the current database state.This paper presents a more flexible fuzzy set-based approach for human-like querying of relational databases. Such approach supports queries that can use both of stationary fuzzy sets with fixed definition and non-stationary fuzzy sets or data-sensitive fuzzy sets with dynamic definitions. Also, the proposed approach allowsfuzzyjoiningofdatabase relations. Such fuzzy join makes a query statement much closed to human querying behaviors.On the other hand, many other query structures like nested, complex and correlated fuzzy queries are supported by the proposed approach.

Keywords:*Fuzzy Joins, Non-Stationary Fuzzy Set, Data-Sensitive Fuzzy Set, Information Retrieval, Fuzzy Logic, Fuzzy Sets, Fuzzy SQL.*

1 Introduction

Structured Query Language (SQL) is a very essential tool for querying relational databases. It manipulates and retrieves data which is crisp and precise by nature. In contrary, it is unable to respond to human-like queries which are uncertain, imprecise and vague in nature. The Fuzzy Queries based on stored procedures to provides many different methods of combining data sets. While care must be taken to avoid inefficient processing, this process allows you to combine data sets in ways that would be difficult or even impossible using other methods. The following SQL query illustratesdifference between the structured query languages (SQL) and Fuzzy queries:

```
SELECT  student_name, height, weight
FROM    student
WHERE   height> X and weight< Y;
```

The result of the query is shown in graphical mode shown in Figure 1. As presented in query1, the values X and Y delimit the space of interesting data. Small squares on the graph show records that satisfy and not satisfy the query criteria. For example, t_r, t_i and t_j represent unsatisfying records while t_c and t_s represent satisfying records. Unfortunately, as shown in

*Computer Science Dept.,ISSR, Cairo Univ.ahmgad10@yahoo.com, hehefny@hotmail.com
** Information Systems Dept., ISSR, Cairo Univ.adelsabour@gmail.com

figure 1, both of t_i, and t_j records are unselected although they are much closed to the specified criteria.

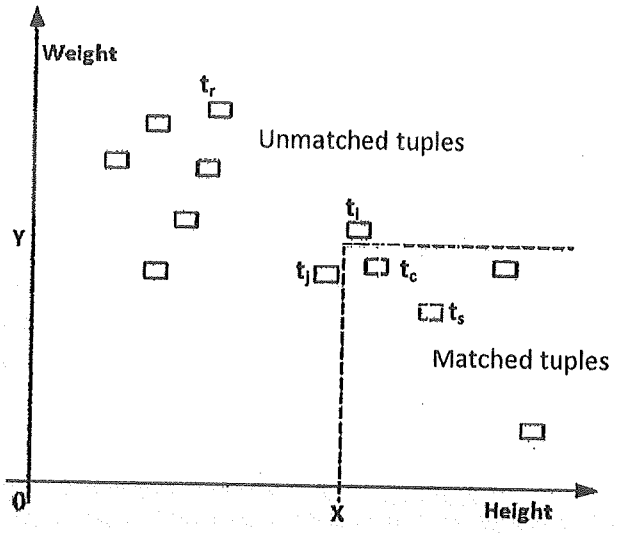


Figure 1: the eliminated records as a result of crisp queries.

In cases when the user cannot unambiguously separate data he is interested in from data he is not interested in by sharp boundaries or when user wants to obtain data that are very close to satisfy queries and to know the index of distance to full query satisfaction, it is necessary to adapt the SQL to these requirements. For example, the classification of persons into males and females is easy, but it is problematic to classify them as being tall or not tall. The set of tall people is far more difficult to define, because there is no distinct cut-off point at which tall begins. This is not a measurement problem, and measuring the height of all elements more precisely is not helpful. Such a problem is often distorted so that it can be described using the well-known existing methodology. Here, one could simply select a height, e.g. 1.80m, at which the set tall begins, see Figure 2 (a). The output of a system using this definition would not be smooth with respect to the height of a person. A person of height 1.79m would produce a different output than a person of 1.81m. In human reasoning this property is not observed and it is also undesirable for reasoning systems that are part of a control system.

Fuzzy logic was suggested by Zadeh as a method for mimicking the ability of human reasoning using a small number of rules and still producing a smooth output via a process of interpolation. It forms rules that are based upon multi-valued logic and so introduced the concept of set membership[6]. With fuzzy logic an element could partially belong to a set and this is represented by the set membership. For example, a person of height 1.79m would belong to both tall and not tall sets with a particular degree of membership. As the height of a person increases the membership grade within the tall set would increase whilst the membership grade within the not tall set would decrease, see Figure 2b. Fuzzy logic is simply the extension of conventional logic to the case where partial set membership can exist, rule

conditions can be satisfied partially and system outputs are calculated by interpolation and, therefore, have output smoothness over the equivalent binary-valued rule base. This property is particularly relevant to control systems.

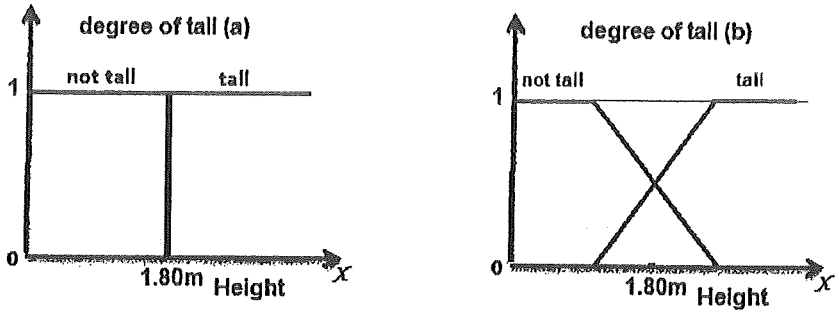


Figure 2: Definition of tall person: (a) using a crisp set and (b) using a fuzzy set [0,1]

In fuzzy queries each record match includes a matching degree is a mathematical measure of the record similarity between the input record and the referenced record that Fuzzy Queries returns from the reference table. The matching degree is represented by a decimal value between 0 and 1, where a matching degree of 1 means an exact match between the value in the input column and the value in the reference table [16].

The main objective of this work is proposing a new approach to overcome the drawbacks ofprevious work mentioned in section number 3and more expressive of uncertainty. Not only retrieve data according to the view of the user, but also according to the database current state

The rest of this paper is organized as follows: section 2 presents an overview of fuzzy query of databases. Previous work where introduced in section 3. Section 4 presents the proposed fuzzy query approach. The proposed tool and illustrative case study were given in section 5. The conclusion is given in section 6.

2 Fuzzy Query of Databases

Fuzzy SQL is a powerful extension to the Structured Query Language (SQL). Conventional SQL joins tables and screens (or selects) rows based on comparisons using Boolean Logic and ordinary arithmetic [8].Fuzzy SQL allows defines semantic concepts namely linguistic variables, linguistic values, linguistic hedges and fuzzy numbers on data elements. A linguistic variable is a variable having fuzzy (i.e. linguistic) values. For example, the linguistic variable "height" may have the linguistic values "Tall", "Medium" And "short". Each of such fuzzy values is represented by its membership function. The fuzzy vales "Tall", "Medium" and "short" are shown in figure 3. As shown in Figure 3, four different levels can be distinguished in the definition of a linguistic variable as shown in figure 3. At the top level, we have the name of the variable (e.g., height). At the level below it, we have the labels of fuzzy values (starting with an initial set of values called primary values or term set).Further down, we have membership values, and at the bottom, we have a universe of discourse.

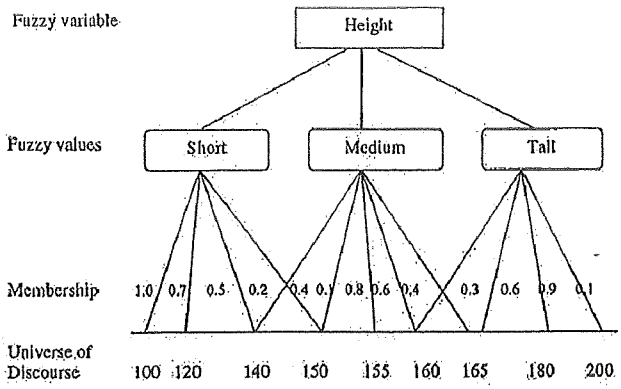


Figure 3: the linguistic variable "Height" with a set of linguistic values: "Tall", "Medium" and "Short"

It is important to observe that linguistic variables have a dual nature; at higher levels, we have a symbolic linguistic form, and at lower levels, we have a wall–defined quantitative analytical form that is, the membership function. After defining the set of fuzzy linguistic terms, it can be easily used to specify the criteria of the selection operation from a database [9]. For example, the user can generate a query statement to retrieve persons with Tall height and Medium weight.

3 Previous Work of Fuzzy Querying in Databases

This section presents several architectures and applications that have been proposed and developed for allowing more flexible and human-like queries of databases. And most of the previously developed fuzzy SQL architectures have the following drawbacks:

1. The inability for dealing with multi subjective views when interacting with multiusers, Preferences are used to help in reducing the amount of information returned in response to user queries [4].
2. The lack of a standardized format in writing statements, each tool has its own syntax.
3. Most of proposed approaches depend mainly on using a time consuming parser/translator to check and convert fuzzy queries to crisp SQL queries.
4. Most of proposed systems defined a Meta base named Fuzzy Meta Knowledge Base (FMB) formed by a set of tables to store all necessary information to describe and manipulate fuzzy attributes and terms [2]. This slows down the process of querying. That is, the generated fuzzy query must be analyzed and translated to a standard SQL statement respecting the definitions of all used fuzzy terms or operators stored in FMB. [7].
5. The problem of performance of evaluating nested fuzzy queries [3].
6. No querying of relational databases using non-stationary or data sensitive fuzzy values.

FSQL proposed by Galindo [10] has been conceived with a mild coupling architecture as shown in figure 4. FSQL is integrated through stored procedures and functions.

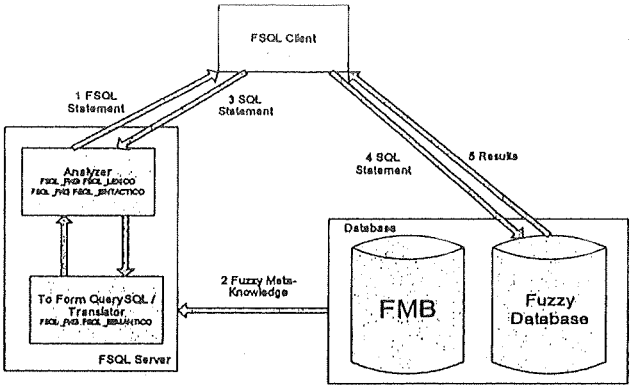


Figure. 4. FSQL implementation architecture (FRDB) by Galindo [10]

The drawbacks of such architecture numbered in drawbacks notation 3,4, and 6.

Also, another approach for fuzzy query called SQLf is proposed in [11]. As shown in its architecture in Figure 5, software layer implemented using a programming language outside the RDBMS is existed. The RDBMS receives and executes the SQL query and sends resulting data to the software layer to manipulate it respecting the query defined fuzzy terms [11]. It has the drawbacks numbered 2, 3 and 4.

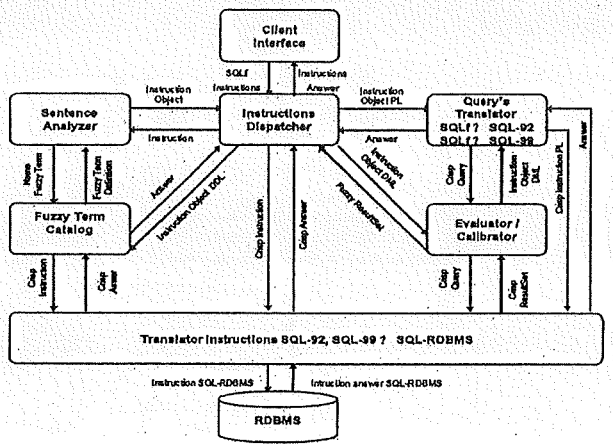


Figure5.SQLf implementation architecture by Bosc&Pivert [11].

Also, another fuzzy relational database management system (FRDBMS) is presented in [2] that aimed to allow fuzzy queries. It can model both of relational database and fuzzy relational database. The language of description and manipulation of the data is therefore fuzzy structured query language (FSQL). The core of this work is a software layer that transforms a fuzzy query statement to its equivalent SQL query statement, see Figure 6. FRDBMS has the drawbacks numbered 1, 2, 3, 5 and 6.

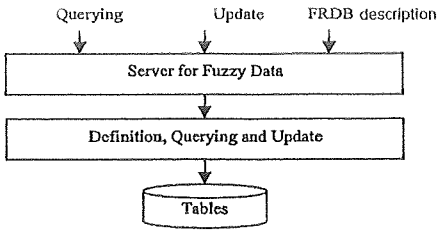


Figure6.the architecture of FRDBMS model.

Also, another Fuzzy Query Architecture has been proposed in [1]by Mishra J et al. has the drawbacks numbered 1, 2, 3, 5 and 6.On the other hand, a Fuzzy Classification Framework Architecture for fuzzy query is presented in [5] by Veryha et al. The drawbacks of this architecture numbered in drawbacks notation (4, and 6). Also, the old fuzzy queries proposed in [17] by Sabour et al. has only drawback number 6.

4 The Architecture Of The Proposed Fuzzy Query Approach

This section presents the architecture of the proposedFuzzy Query Approach as shown in Figure 7. It aims mainly to allow easily the generation and manipulation of more flexible and human-like queries.

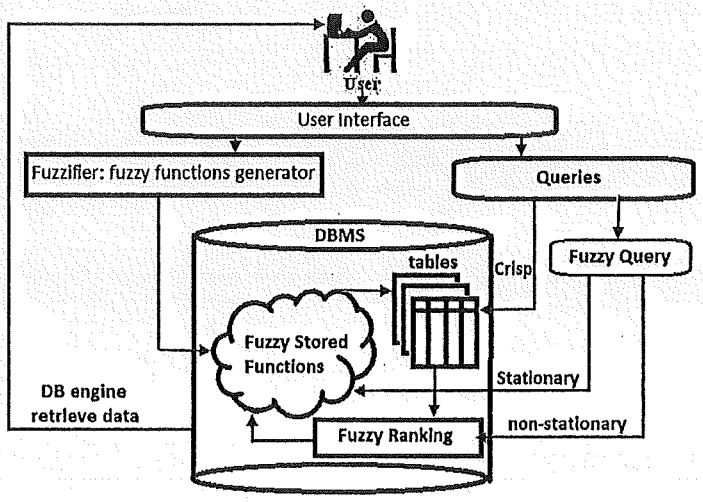


Figure 7. the new Fuzzy query architecture.

The proposed approach assumes satisfying the user subjective requirement as the most important objective. So, a user can easily define his/herown fuzzy profile including all linguistic terms that can be used in a fuzzy query statement. Such profile is stored as database objects instead of the use of fuzzy Meta knowledge base, used in many of previous works, which is time consuming.Thetool begins responding by generating PL-SQL statements that creates database stored objects (such as packages, procedures, functions, views, etc.) within the database.Such objects save all required static or dynamic datainstead

of using Fuzzy Meta Knowledge Base (FMB) when running. Accordingly,such tool acceleratestheexecution [12] and enhances the efficiency when performing fuzzy queries. In consequence, the user can write a Fuzzy Query or generated from query builder GUI, the query is identical to the standard SQL language by using stored objects within the database, the query will go directly to the Database engine to execute directly in one step as it does with regular Crisp SQL statements[13].The proposed fuzzy query approach supports **data-sensitive fuzzy queries**indicates a fuzzy set with a dynamic definition. Such definition is a data-sensitive one. In other words, the control points of a non-stationary fuzzy set change reflecting changes in the applied dataset. For example: the linguistic value "Good Student" has a data sensitive definition reflecting the difficulty or complexity of the studied courses. Accordingly, the support of "Good Student" in a simple course differs from a sophisticated course. That is, in a simple course, most students have high average grades. In contrary, in a more complex course, most students almost have less average grades.

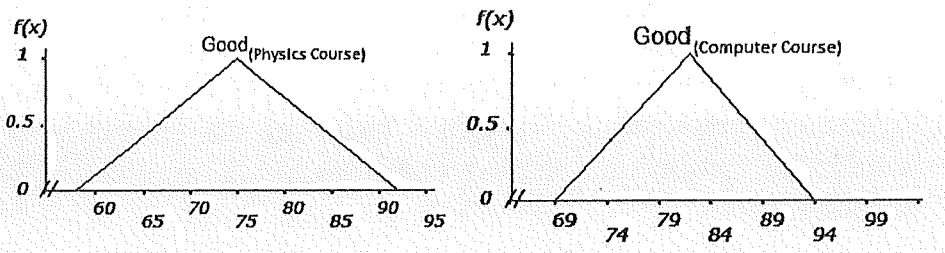


Figure 8.A. Good Students in Physics Course B. Good Students in Computer Course

Aggregated information from the searched database can be used in a flexible manner to define non-stationary fuzzy sets. The following results shows defining a set of linguistic values: LowSal, ModerateSal and HighSal over the linguistic variable employee salary based on some aggregated information like max, average and min salaries.As a result of such definitions, a salary of an employee with 3900EGP represents a HighSal Salary with 96% matching degree while another employee with same salary in another petroleum company is not considered HighSal at all. The next two figures show different definitions, on query time, of the above linguistic values reflecting the actual salaries of the searched database.

MIN	CORE	MAX	NOEMPS	CONCEPT
1783.33	2650	3316.67	26	LowSalary
3283.33	3900	4516.67	7	HighSalary
2783.33	3300	3816.67	12	ModerateSalary

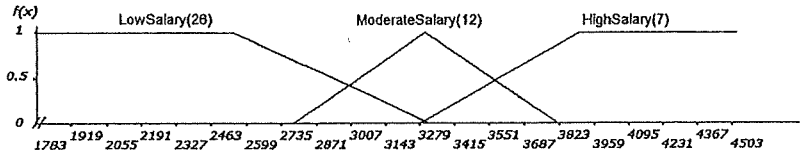


Figure 9. Lingustic values for clerk employees salaries.

MIN	CORE	MAX	NOEMPS	CONCEPT
4950	7650	10150	20	LowSalary
9850	12500	15150	6	HighSalary
8350	10000	11650	9	ModerateSalary

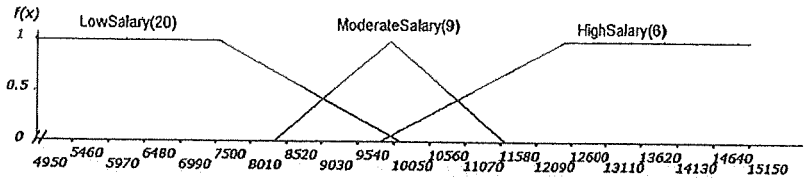


Figure 10. Lingustic values for salesman employees salaries.

5. An Application of The Proposed Approach: Fuzzy Queries Tool

Fuzzy queries tool (FQ) represents an application of the proposed approach. FQ with its friendly graphical user interface facilitates the interaction with users from profiling to constructing and executing flexible human-like queries on relational databases.

5.1 Login to FQ Service

At the first time the user aims to interact with the tool, he/she must create an account with a user name and a password. After that, the user can define his/her own linguistic terms via the tool GUI. Consequently, each created stored database object within the user session becomes part of the user fuzzy profile. Also, the user can generate a fuzzy query statement using his/her predefined linguistic terms and execute it. The main screen, displays the user available services or activities that facilitate the definition of linguistic terms, linguistic variables, values, the generation and execution of a fuzzy query statement and showing the result of the executed query statement.

5.2 Linguistic Variables and values Services

This service allows creating, modifying or deleting a linguistic variable that has a set of linguistic values that will be defined over it. The user enters just the name of the linguistic variable and a description for it. For example, a linguistic variable may be a computer grades, temperature degree, height or salary as shown in Figure 11. On the other hand, there is another service for defining a set of linguistic values over the predefined linguistic variables. A set of linguistic values can be defined easily for each predefined linguistic variable to be used when constructing a fuzzy SQL statement. For example, the user can define the linguistic values

"Tall", "Short" and "Medium" for the "Height" linguistic variable. Each linguistic value can be defined in a very flexible representation, that the user can select the more suitable membership function and modulate its control point's values to be more closed to the meaning of the linguistic value. Through a graphical user-friendly representation, the user can check the linguistic values definitions and its overlapping for a specific linguistic variable. The graphical representation will help the user to make the effective modifications in the selected membership function to satisfy the meaning of the linguistic value.The user can draw these values (Control Points) only, or add to the values set by or delete the previous drawing. Then add these values to the Linguistic Value to add other Linguistic Value for the same Linguistic Variable. And so on, until the user saves all the values as shown in Figure 11.

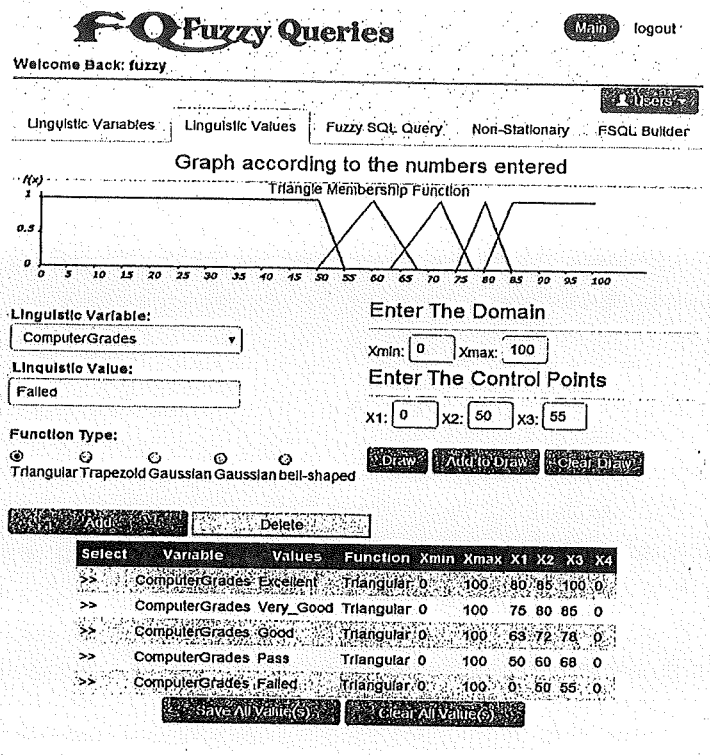


Figure 11. Defining a set of linguistic values over computer grades linguistic variable

5.3 Fuzzy SQL QueryService

This screen enables the user to enter sentences the Fuzzy Query then run to display the result as shown in Figures Next. Which shows the result and how to calling and use Linguistic Variable / Values on a numeric field.

5.3.1 Nested fuzzy query

Sub-queries are also allowed by the tool. Fig. 12 shows a sub-query statement that retrieves each employee name and salary for each employee has salary around the average salary in the company respecting a threshold value of 0.7.

```
SELECT SID, SName,height, idealHeight,Sport, MD
FROM   table(aroundjoin(student,IdealHeightSport,Height,IdealHeight,0.8))
Order By MD desc
```

Run Query

SID	SNAME	HEIGHT	IDEALHEIGHT	SPORT	MD
17	Sabour Ahmad	186	186	Cycling	1
4	Adel Sabour	200	201	Basketball	0.97
17	Sabour Ahmad	186	185	Football	0.96
19	Tasneem Ahmad	187	186	Cycling	0.96
20	Taki Ahmad	180	182	Amateur wrestling	0.93

...

Figure 17. An example of a direct fuzzy Around Join

6. Conclusion

This paper presents a more flexible fuzzy set based approach for human-like querying of relational databases. Such approach supports both of stationary fuzzy sets with fixed definition and non-stationary fuzzy sets or data-sensitive fuzzy sets with dynamic definitions.

Commonly, many approaches have been proposed allowing fuzzy querying of relational databases. Yet, such approaches presented complicated solutions which are two-fold.

Firstly, they depend, mainly, on defining a fuzzy-meta database (FMD) that is used for storing the definitions of all fuzzy terms that can be used. Accordingly, each generated fuzzy query statement needs to be analyzed to fetch each fuzzy term and brings its definition from FMD which is time consuming. If such approach depends on converting the generated fuzzy query to an equivalent traditional SQL query statement, it reads the definition and then passes it to its counterpart stored database object then process the query.

Secondly, some other approaches needs to convert any generated fuzzy query statement to classical SQL statement by eliminating all used fuzzy terms. In consequence, the definition of such fuzzy terms, stored in FMD, is used to fuzzily evaluating the resulted tuples and assigning a matching degree for each resulted tuples.

In contrary, the proposed approach depends mainly on storing all used fuzzy terms as database objects namely stored procedures and functions within packages in an organized fashion. Such stored database objects resembles the user fuzzy profile. Consequently, a fuzzy query statement can be generated directly as a traditional SQL statement using the fuzzy terms predefined as database objects. In consequent, there is no need for an analyzer or a translator to covert a generated fuzzy query into an executable query statement. Also, there is no need for a Fuzzy Meta-Knowledgebase that existed in most of previously proposed

approaches for fuzzy queries [17]. Hence, not only the complexity of allowing a fuzzy query is reduced but also the response time of executing such query. Also, because linguistic terms are represented as a set of user defined stored functions; the proposed approach allows writing any simple, nested, correlated and complex fuzzy query statements as traditional select statements in a very flexible manner. The proposed approach enables to fuzzily join to relations respecting a fuzzy condition in a human-like fashion.

7. References

1 Mishra J., "Fuzzy Query Processing", International Journal of Research and Reviews in Next Generation Networks, Vol. 1, No. 1, March 2011.

2 Grissa A. and Ben Hassine M., "New Architecture of Fuzzy Database Management Systems", The International Arab Journal of Information Technology, Vol. 6, No. 3, July 2009.

3 Qi Yang et al., "Efficient Processing of Nested Fuzzy SQL Queries in a Fuzzy Database", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 13, NO. 6, NOVEMBER/DECEMBER 2001.

4 Abbaci K. et al., "Selecting and Ranking Business Processes with Preferences: An Approach Based on Fuzzy Sets", International Conference on Cooperative Information Systems, Greece (2011)

5 Singh K. et al., "Study of Imperfect Information Representation and FSQL processing", International Journal of Scientific & Engineering Research Volume 3, Issue 5, May-2012

6 Garg A., and Rishi R., "Querying Capability Enhancement in Database Using Fuzzy Logic", Global Journal of Computer Science and Technology, Vol. 12 Issue 6 Version 1.0 March 2012.

7 Ir. Wahidin "Fuzzy Control", 2007.

8 Gadallah A. and Hefny H. "An Efficient Database Query Processing Tool Based on Fuzzy Logic", The 37th Annual Conference on Statistics and Computer Science, 2002.

9 Shawky A. et al. "FRDBM: A Tool For Building Fuzzy Relational Databases", The Egyptian Computer Journal, 2006.

10 Galindo. J et al. "Handbook of Research on Fuzzy Information Processing in Databases", Chapter XI FSQL and SQLf : Towards a Standard in Fuzzy Databases, 2008.

11 Bosc P. and Pivert O., "SQLf Query Functionality on Top of a Regular Relational Database Management," in Proceedings of Knowledge Management in Fuzzy Databases, Heidelberg, pp. 171-190, 2000.

12 Moore S." Oracle Database PL/SQL Language Reference, 12c", pp. 1, 2014.

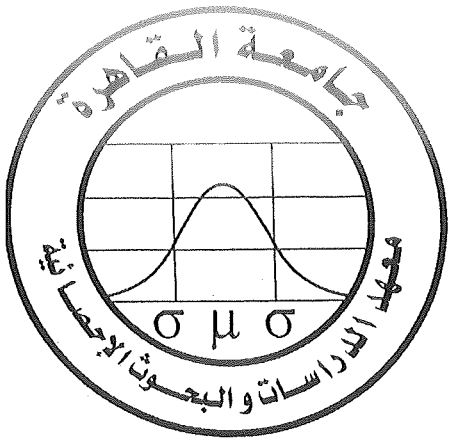
13 Elmasri and Navathe , "Fundamental of Database Systems", 2011.

14 Q. Yang, C. Liu, J. Wu, C. Yu, S. Dao, H. Nakajima, and N. Rishe. Efficient processing of nested fuzzy SQL queries in fuzzy databases. In IEEE Int'l Conf. on Data Engineering, pages 131–138, 1995.

15 Weining Zhang and Ke Wang, "A Fuzzy Equi–Join and Its Evaluation", NSERC, 2001.

16 Jedediah J. Teres,"Using SQL Joins to Perform Fuzzy Matches on Multiple Identifiers", NESUG 2011.

17 Sabour A. et al., Flexible Querying of Relational Databases: Fuzzy Set Based Approach, Springer CCIS 488, pp. 446–455, 2014.



جامعة القاهرة

معهد الدراسات والبحوث الإحصائية

المؤتمر السنوى التاسع والأربعون للإحصاء وعلوم الحاسب وبحوث العمليات

مجلد

علـــوم حاســب

٢٢-٢٥ديسمبر ٢٠١٤