Middle-East Journal of Scientific Research 24 (1): 65-72, 2016
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2016.24.01.22888

Multi-Attribute Density Estimation Based Location Selection Approach in Multi-Agent Disease Prediction Model for Decision Support System Using Diagnosis Pattern and Data Mining

[1]M. Inbavalli and [2]G. Tholkappia Arasu

[1]Assisstant Professor/MCA, Er. Perumal Manimekalai College of Engineering, Hosur, Tamil Nadu, India
[2]Principal, AVS College of Technology, Salem, Tamil Nadu, India

Abstract: The presence of decision support systems plays a vital role in many situations like business intelligence and medical solutions. There are many designs has been proposed earlier to support decision making and suffers with the problem of accuracy and time complexity. We propose a novel approach which uses Multi-Attribute Density Estimation technique (MADE) to choose the set of locations from where the data can be retrieved. The method uses various meta data which represents the availability of data in different locations of the network and based on the meta data, the method computes the MADE measure to choose most optimal locations. From identified locations, the method generates diagnosis patterns which contain various information about the medical history available in the location. The number of agent generation is performed according to the MADE factor and based on the patterns generated, the DSF (Decisive Support Factor) is computed which shows the possibility of the disease. The proposed method reduces the time complexity and improves the accuracy of decision making system.

Key words: Decisive Support System · Data Mining MADE · Diagnosis Pattern · Disease Prediction · MAS

INTRODUCTION

Data mining is the process of extracting knowledge from large data base using some extraction approach. The data mining approach can be used in any kind of multi dimensional data set and the approach makes the difference in information extraction accuracy. For example, the organization maintains various information at different location of their units and the data mining approach is necessary in identifying the required information. To identify a person affected by particular disease the query execution has to be performed with the input of particular disease called as symptom [1].

In general, there are many symptoms which are common for many diseases like cold, fever and diarrhea and vomiting. Some other are unique symptoms for many disease. So in order to identify or make any decision based on the symptoms available, efficient data mining approaches are necessary [2].

The decisive support systems are automated and intelligent tool which helps the medical practitioner in difficult situations where the symptoms resemble many diseases. The decisive support system can read set of all symptoms and identify the possibility for each of the disease may occur using some intelligent approach. The probability based approaches are one among them, which collects the data record from all the locations and computes each disease probability and produce the result to the medical practitioner. Based on the probability value, the medical practitioner can use the computed results for decision making [3].

To collect the information stored in geographic region can be done by mobile agents. The agent based technology helps generating agents and allow the movements of agents from one location or container to another. Also the agent has set of behaviors, that it can perform some specific job at the remote site. The collected results can be used as an input to the decisive support system [4].

Corresponding Author: M. Inbavalli, MCA Er. Perumal Manimekalai College of Engineering, Hosur, Tamil Nadu, India.

Diagnosis pattern is a collection of attribute values and symptoms set which is generated from the medical data set present in the different location. The generic frequent pattern mining techniques can be used to compute the support and count values. Based on the support threshold a symptom set can be selected, as the pattern for disease occurrence [5].

The problem of decisive support system is identifying the location where the required information is present to perform the specific query, this increases the time complexity of the decisive support system and the method has to choose the locations which has certain information. MADE is the one, which performs the selection based on the density of information [6].

Related Works: The researchers has been paying attention in developing intelligent agents in specific domain knowledge and transferring of agents to all the locations for data retrieval for effective prediction of disease. Few of them are

Decision support system for water distribution systems based on neural networks and graphs theory for leakage detection [6, 7], presents an efficient and effective decision support system (DSS) for operational monitoring and control of water distribution systems based on a three layer General Fuzzy Min-Max Neural Network (GFMMNN) and graph theory. The operational monitoring and control involves detection of pipe leakages. The training data for the GFMMNN is obtained through simulation of leakages in a water network for a 24h operational period. The training data generation scheme includes a simulator algorithm based on loop corrective flows equations, a Least Squares (LS) loop flows state estimator and a Confidence Limit Analysis (CLA) algorithm for uncertainty quantification entitled Error Maximization (EM) algorithm. These three numerical algorithms for modeling and simulation of water networks are based on loop corrective flows equations and graph theory. It is shown that the detection of leakages based on the training and testing of the GFMMNN with patterns of variation of nodal consumptions with or without confidence limits produces better recognition rates in comparison to the training based on patterns of nodal heads and pipe flows state estimates with or without confidence limits [8].

Decision Support System for Water Distribution Systems Based on Neural Networks and Graphs [3], resents an efficient and effective Decision Support System (DSS) for operational monitoring and control of water distribution systems based on a three layer General Fuzzy Min-Max Neural Network (GFMMNN) and graph theory. The operational monitoring and control involves detection of pipe leakages. The training data for the GFMMNN is obtained through simulation of leakages in a water network for a given operational period. The training data generation scheme includes a simulator algorithm based on loop corrective flows equations, a Least Squares (LS) loop flows state estimator and a Confidence Limit Analysis (CLA) algorithm for uncertainty quantification entitled Error Maximization (EM) algorithm. These three numerical algorithms for modeling and simulation of water networks are based on loop corrective flows equations and graph theory [9].

A Semi-Automated Distributed Decision Support System Virutal Enterprises, presents a distributed collaborative decision support framework for problem resolution in partnerships of autonomous and heterogeneous enterprises. The developed prototype for conceptual model's verification and validations is composed by three important elements: a collaborative discussion tool via argumentation with moderation, flexible decision protocol and tools for previous impact of the decision by performance evaluation methods.

A distributed clinical decision support system architecture, proposes an open and distributed clinical decision support system architecture. This technical architecture takes advantage of Electronic Health Record (EHR), data mining techniques, clinical databases, domain expert knowledge bases, available technologies and standards to provide decision-making support for healthcare professionals. The architecture will work extremely well in distributed EHR environments in which each hospital has its own local EHR and it satisfies the compatibility, interoperability and scalability objectives of an EHR. The system will also have a set of distributed knowledge bases. Each knowledge base will be specialized in a specific domain (i.e., heart disease) and the model achieves cooperation, integration and interoperability between these knowledge bases [10].

An Intelligent Decision Support System for Establishment of New Organization on Any Geographical Area Using GIS, propose an intelligent system to support making decisions for establishment of new resources on any geographical area. Using GIS we can get the whole information of any geographical area. The GIS and AI techniques are used here with spatial database [11, 12].

Water Distribution System Monitoring and Decision Support Using a Wireless Sensor Network [13, 14], comprise labyrinthine networks of pipes, often in poor states of repair, that are buried beneath our city streets and relatively inaccessible. Engineers who manage these systems need reliable data to understand and detect water losses due to leaks or burst events, anomalies in the control of water quality and the impacts of operational activities (such as pipe isolation, maintenance or repair) on water supply to customers. Water Wise is a platform that manages and analyses data from a network of wireless sensor nodes, continuously monitoring hydraulic, acoustic and water quality parameters. Water Wise supports many applications including rolling predictions of water demand and hydraulic state, online detection of events such as pipe bursts and data mining for identification of longer-term trends.

Decision support system for auditing distribution logistics information systems, is based on a multi-criteria hierarchical ranking and is built around a decision support system called SADAUDIT that allows the audit of complex systems. We illustrate the proposed methodology with the real case of three companies based in Tunisia and belonging to the same group.

All the above discussed methods have the problem of predicting the case exactly and provides false results, so that we propose a new disease prediction model to predict the disease affected.

Proposed Method: The proposed approach Multi-Attribute Density Estimation Technique (MADE) is to choose the set of locations from where the data can be retrieved. The method uses various meta data which represent the availability of data in different locations of the network and based on the metadata, the method computes the MADE measure to choose most optimal locations. From the identified locations, the method generates diagnosis patterns which contain various information about the medical records available in the location. The number of agent generation is performed according to the MADE factor and based on the patterns generated, the DSF (Decisive Support Factor) is computed, which shows the possibility of the disease. The proposed method reduces the time complexity, improves the accuracy and reduces the false prediction rate.

Functional Components: The proposed approach has the following components namely Location Selection, MADE, Preprocessing, Data Fetching, Decisive Support Factor, Disease Pattern Generation, Disease Prediction.

Location Selection: The proposed method maintains the data present in the location and also the meta data in different locations. Each node or location has variety of information and varies according to the details stored in the location. The meta data contains information like a pattern based on, for example <“temperature”, “cold”, “cough”, “128”, “2020”>, the pattern shows that the first three are the symptoms and the fourth attribute is the number of instances and the final one is the total number of records in the location. With the input symptom set given by the user, the Multi-Attribute Density Estimation is computed . Based on the computed density estimation a set of locations is selected which has more value than the density threshold.

Algorithm
Input: Meta Data Md, Symptom Set Ss.
Output: Location Set Ls, Multi attribute density set Mads.
Step 1: Start
Step 2: Collect the symptom set given by the user Sys = Σ *Symptoks* $\in$ *UserInterface*

Step 3: for each meta information Mi from Md
for each symptom Si from Sys
if $\sum_{i=1}^{size(sys)} {}_{Si \in sys(i)}$ && size(sys)==(Size(Si)-2)
//Pattern identified
Add location to the location set.
Ls = Σ *Location* (*Ls*) + *Mi*(*Li*)
end
End

Step 4: for each location Li from Ls
compute multi attribute density estimation.
if Made>MADThreshold
Mads = Σ *MADS*(*made*) + *Made*(*Li*)
end
end
Step 5: Stop.

Multi Attribute Density Estimation: The MADE factor is computed based on the meta data available and for the given meta information Mi, the number of attributes available in the Mi are identified and also the number of instances available in the location and finally the total instances Ti is computed from the meta information. The algorithm are as follows:

Algorithm
Input: Meta Information Mi
Output: MADF.

Step 1: Start
Step 2: Identify number of symptoms from Mi.
NA = Σ (*Attributes* ∈ *Mi*) – 2
Step 3: Identify number of instances the location has.
NOI = Value(Mi(NA)+1).
Step 4: Identify the total number of instances Ti.
TI = Value(Mi(NA)+2).
Step 5: Compute multi attribute density estimation MADE.
MADE = $\frac{NOI}{TI} \times NA$
Step 6: Stop.

Data Fetching: The method computes the number of agents required to perform the data collection using the MADE factor. For each location Li, from the identified set, it generates a agent and If the MADE value is less than a density threshold then the agent will be allocated with another location. Each agent will be initialized with location parameter and query parameters. The Agents are moved instantly to the remote container and asked to perform the data processing. The agents retrieve the necessary information from the data base available at the remote location and moved back to the original location or home container. The number of agent is only depending on the number of locations where the data available.

Disease Pattern Generation: First the set of disease available in the data set is identified and the records belong to each disease is separated. From each group of records, the set of records with identified symptoms are identified and compute support values for each of the disease. Finally a pattern of disease is generated which has disease name, support value and total instance of the disease records. The generated pattern will be used to perform prediction of diseases. The algorithm are as follows.

Algorithm
Input: Records set Rs, Symptom set Ss.
Output: Pattern Set Ps.
Step 1: Identify distinct disease available,
Diset = $\int_{i=1}^{i=N} \Sigma\, Ps\,(Disease) \nexists\, Diset$
Step 2: For each disease Dis from Diset
Separate records from ps according to Disease.
Ts = $\int_{j=1}^{i=k} \Sigma Ps(i) \in Dis$
Compute count value of symptom pattern.
Count = $\sum_{i=1}^{size(Ts)} Ts(i).\ symptoms == Ss(Symptoms)$
Support = $\frac{Count}{Ts}$
Add to pattern Ps = Σ *Pattern* (*Ps*) + {*Dis*, *Support*, *Ts*, *Location*}
end.
Step 3: Stop.

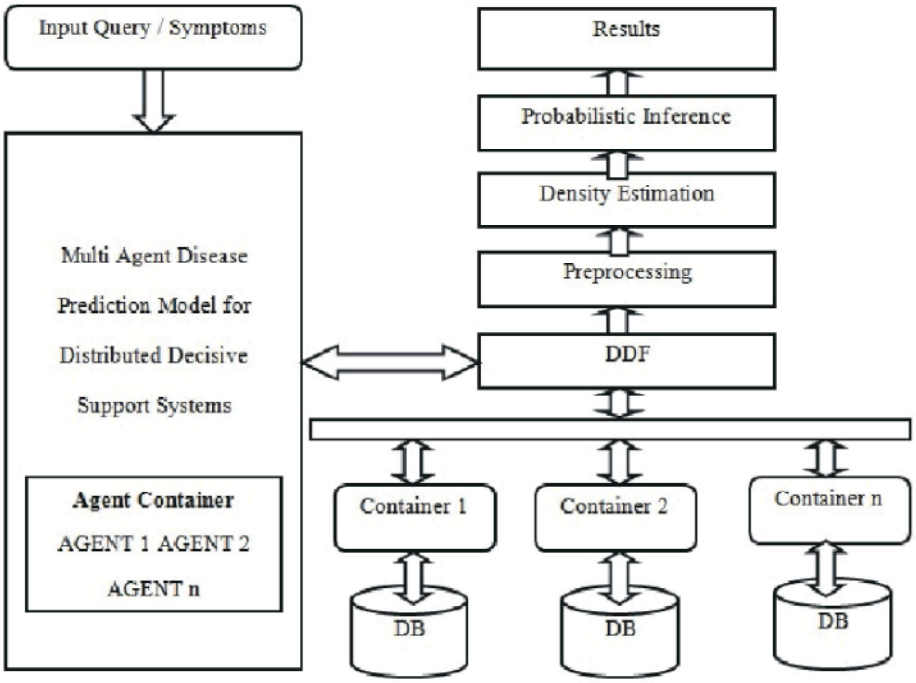


Fig. 1: Proposed System Architecture

Disease Prediction: The disease pattern generated by all the agents are collected and based on that the decisive support factor for each of the disease are computed. The decisive support factor is computed by cumulating of all the support values and averaging them based on the number of locations. Finally from the computed Decisive Support Factor value a maximum valued disease is identified and returned as result.

Algorithm

Input: Disease Pattern Dp.

Output: Identified Disease D.

Step 1: Start

Step 2: Identify distinct disease available

$$\text{Diset} = \int_{i=1}^{i=N} \Sigma\, Dp\,(Disease) \nexists\, Dp$$

Step 3: For each disease Dis from Disease Set Diset

compute disease support factor DSF = $\frac{\Sigma Support(Dis)}{\Sigma Ti(Dis)}$

end

Step 4: Select most DSF valued disease.

Step 5: Stop.

RESULTS AND DISCUSSION

The proposed system is implemented and tested in Java Agent Development Environment (JADE). It is a software Framework executed in the Java language. This section illustrates the prediction quality of the proposed algorithms. The proposed approach uses number of mobile agents according to the MADE value and if the MADE value is less than a threshold then the agent will be added with another location. The dataset used for evaluation are UCI Diabetes, Arrhythmia and Breast Cancer. In this analysis evaluation metrics are accuracy, time complexity and space complexity with varying locations and data sets. The proposed approach illustrate that the prediction quality is more accurate when the number of records increases.

The Table 1 shows the efficiency of the proposed algorithms with 1 million to 4 million input data records and the efficiency of the algorithms prediction quality.

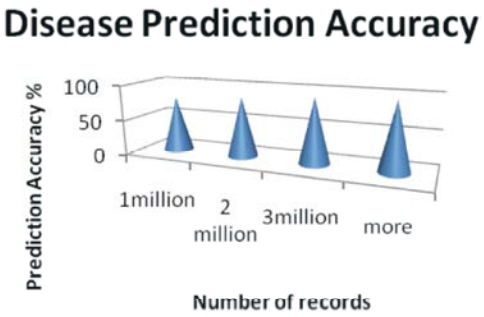
The Graph 1, shows the prediction accuracy of the proposed approach and it shows that the proposed approach has more accuracy when the number of records increases.

Table 1: Shows the efficiency of algorithms according to number of records used

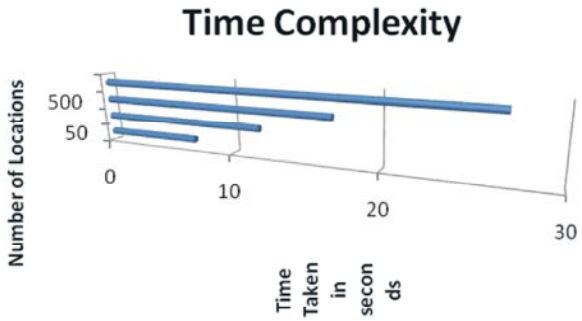
No. of Patterns/inputs	Prediction quality (%)
1 million	83
2 million	88
3 million	95
more	98.5

widal	tiredness	tr	pressure	sugar	colastrol
2	2	102	0		33
3	4	100	0	158	
4	3	101			
1	2	102	0	168	43
3	4	100	0	156	
4	1	101		320	
2	2	102	0		21
3	4	100	0	158	
4	3	101			35
1	2	102	0	168	
3	4	100	0	156	25
4	1	101		320	
2	2	102	0		45
3	4	100	0	158	

Fig. 2: Shows the snapshot of data table used to predict the disease



Graph 1: Shows the efficiency of algorithm



Graph 2: Time complexity of the proposed approach

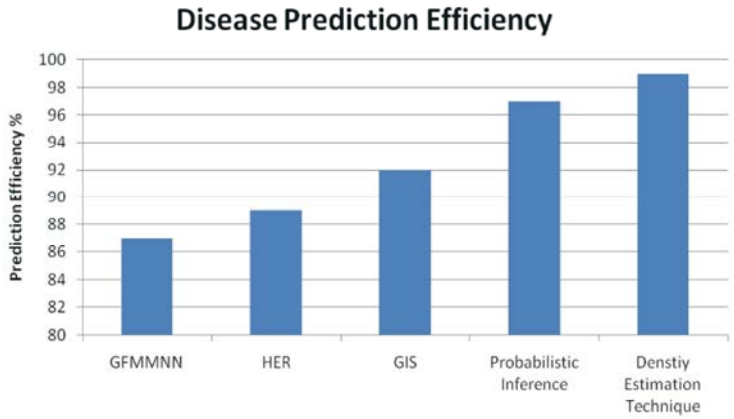
Performance Analysis: The performance of the proposed methods are evaluated with the existing methods with difference data sets based on accuracy and false prediction rate are shown in Graph 4 and Graph 5. This analysis shows that the proposed methods provide high prediction accuracy and less false prediction rate, less time and space complexity.

Table 2: Comparison of performance analysis

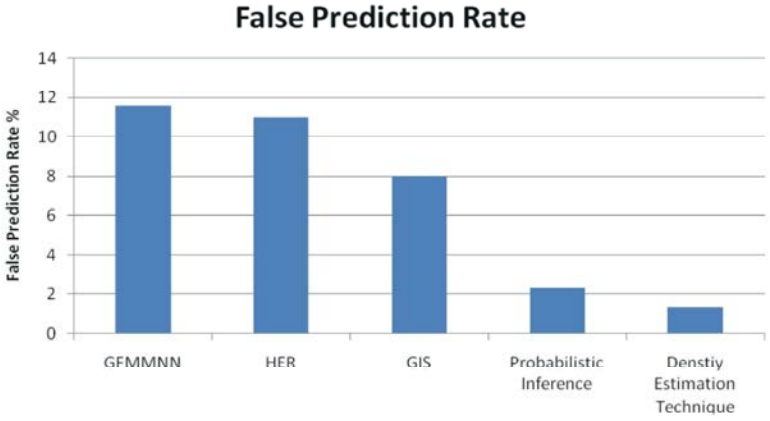
Method Name	Dataset Name	No of Instances	Prediction Accuracy	False Classification Rate
GFMMNN	Abalone	4177	83	17
	Arrhythmia	452	79.5	20.5
	Breast Cancer	286	75.2	24.8
EHR	Abalone	4177	86	14
	Arrhythmia	452	84.6	15.4
	Breast Cancer	286	83.8	16.2
GIS	Abalone	4177	89.2	10.8
	Arrhythmia	452	87.8	12.2
	Breast Cancer	286	87.4	12.6
Probabilistic	Abalone	4177	96.8	3.2
	Arrhythmia	452	97.4	2.6
	Breast Cancer	286	97.3	2.7
Density Estimation Based	Abalone	4177	98.5	1.5
	Arrhythmia	452	98.4	1.6
	Breast Cancer	286	98.3	1.7

Table 3: Efficiency of Multi-Agent Probabilistic Inference model –Time Complexity

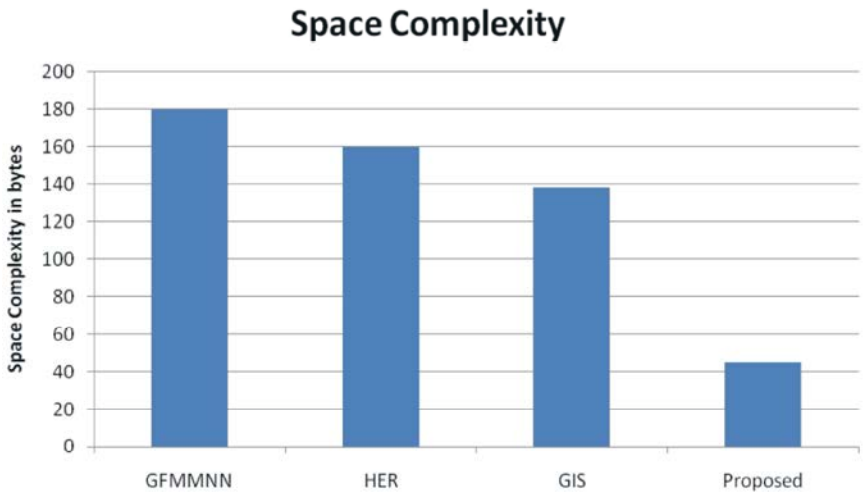
	Multi-Agent Probabilistic Inference Model approach	Multi-Attribute Density Estimation Location Selection approach
No. of Locations	Time Complexity (in Seconds)	
50	7	6
100	12	9
500	17	14
1000	27	21



Graph 3: Comparison of disease prediction efficiency



Graph 4: Comparison of false prediction rate



Graph 5: Comparison of space complexity

The Graph 3, shows the comparison of disease prediction efficiency produced by different methods and it shows clearly that the proposed method has produced efficient result than other methods.

The Graph 4, shows the comparison of false prediction rate produced by different methods and it shows clearly that the proposed method has produced less false classification than other methods.

The Table 3, shows the time complexity of the proposed approach and it shows clearly that the proposed approach has less time consuming even at many locations.

The Graph 5, shows the space complexity produced by different methods and it shows clearly that the proposed approach has produces less space complexity than other methods.

CONCLUSION

The proposed multi attribute density estimation based decisive support system for the prediction of disease has implemented and produced efficient result with increased prediction accuracy and decrease the space, time and false prediction rate. The method uses the meta data to choose the set of locations and the agent generation based on MADE factor. The agents generate disease pattern according to the support values and finally, with the generated pattern and the disease support factor (DSF) is computed and choose the maximum valued disease. This algorithm provides the accuracy of (98.5 %), time taken (21 seconds) and the memory usage is (49 bytes). Moreover, the performance of the proposed methods are compared and evaluated with the other existing methods such as GFMMNN, HER and GIS. From this analysis, it is observed that the proposed methods provide the better performance and less false prediction.

REFERENCES

1. Kumar Prakash, Pradeep Kumar and Vikas Kumar, 2013. An Effective Dynamic Load Balancing Algorithm for Grid System, International Journal of Engineering Trends and Technology (IJETT), 4(8): 3713-3718.
2. Arsini, 2012. Decision Support System for Water Distribution Systems Based on Neural Networks and Graphs, Computer Modelling and Simulation (UKSim), pp: 315-323.
3. Silva, M.V.D., 2012. A Semi-Automated Distributed Decision Support System Virutal Enterprises, Latin America Transactions, IEEE, 10(1): 1235-1242.
4. Shaker, H. and El-Sappagha, 2014. A distributed clinical decision support system architecture, Journal of King Saud University, Computer and Information Sciences, 26(1): 69-78.
5. Sonam Pal, 2013. An Intelligent Decision Support System for Establishment of New Organization on Any Geographical Area Using GIS, International Journal of Advanced Research in Computer Science and Software Engineering, 3(8).
6. Huang Ivy B., Jeffrey Keisler and Igor Linkov, 2011. Multi-criteria decision analysis in environmental sciences: Ten years of applications and trends, Science of The Environment, 409(19): 3578-3594.

7. Joao Coutinho-Rodrigues, Ana Simao, Carlos Henggeler Antunes, 2011. A GIS-based multicriteria spatial decision support system for planning urban infrastructures, Decision Support Systems, 51(3): 720-726.
8. Gerard C. Kelly, Marcel Tanner, Andrew Vallely and Archie Clements, 2012. Malaria elimination: moving forward with spatial decision support systems, Trends in Parasitology, 28(7).
9. Prasuhn Volker, Hanspeter Linigerb, Simon Gisler, Karl Herweg, Anton Candinas and Jean-Pierre Clément, 2013. A high-resolution soil erosion risk map of Switzerland as strategic policy support system, Land Use Policy, 32: 281-291.
10. Gawali Tukaram K. and Mrs. Ujwala M. Patil, 2012. The first way to implement smooth spatial query processing in spatial database, World Journal of Science and Technology, 2(3): 99-102.
11. Zhengmeng, Chai and Zhang Xingling, 2012. On Decision Support Systems and It's Application to the Clinical Decisions, 2012 International Conference on Innovation and Information Management (ICIIM 2012), IPCSIT), IACSIT Press, Singapore.
12. Christina Albert Rayed, 2012. Using GIS for Modeling a Spatial DSS for Industrial Pollution in Egypt, American Journal of Geographic Information System, 1(3): 33-38.
13. Allen, M., 2013. Water Distribution System Monitoring and Decision Support Using a Wireless Sensor Network, Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2013 14th ACIS International Conference on, pp: 641-646.
14. Agrebi, M., 2013. Decision support system for auditing distribution logistics information systems, Industrial Engineering and Systems Management (IESM), 1: 8.