

## Performance Optimization Simulation Analysis of Artificial Intelligence Inference Engine Based on Data Mining

Shaodong Hu

Application of Atistics, Chengdu University of Technology, Chengdu of Sichuan 610059, China

774504616@qq.com

**Keywords:** Data Mining; Artificial Intelligence Reasoning Engine; Performance Optimization

**Abstract:** With the rapid growth of information on the network, search engine has become a necessary tool for knowledge search and knowledge discovery in order to query the required knowledge. Based on this, this paper applies data mining algorithm to objectify reasoning strategies in artificial intelligence, and proposes a unified reasoning strategy model. The purpose is to improve the application scope and code reuse of reasoning strategy, so as to reduce the complexity of reasoning strategy maintenance. The research results show that the quantitative uncertainty inference algorithm is used in the system, which changes the inaccurate results caused by qualitative reasoning, thus helping users to push data information more efficiently and accurately. The research shows that the automatic learning inference system is realized by using the theory and algorithm design in the fields of artificial intelligence and data mining. At the same time, the algorithm has improved efficiency in the process of clustering compared with the system clustering algorithm, and provides a good foundation for the next step of outlier data mining and classification.

### 1. Introduction

When artificial intelligence progresses to a certain extent, the requirements for the combination of symbol processing technology and neural network processing technology become more and more intense, and data mining is a good combination of the two [1]. Data mining is a non-trivial process that identifies effective, novel, potentially useful, and ultimately understandable patterns from large, incomplete, noisy, fuzzy, and random data sets [2]. Since the search engine system model requires a large amount of positive and negative sample information, one of the important tasks of the search engine system is the management of the sample library [3]. The traditional inference system is designed to integrate the inference strategy into the inference system. It cannot be customized to the inference strategy, so that the inference strategy can not be applied flexibly in various practical problems, which brings trouble to the maintenance of the inference system [4]. In knowledge acquisition, people expect to be able to automatically extract knowledge including examples from massive data. However, how to construct a rich, effective and redundant case base from traditional databases has always been one of the difficult problems in CBR [5-6]. In this paper, the classification and clustering technology in data mining is applied to search engine system, which improves the reliability of data search and the autonomy of data processing in search engine system, realizes the automatic optimization of search engine performance and improves the efficiency of the system [7].

After the introduction of the inference strategy model in the intelligent system, any new reasoning strategy needs to inherit the base strategy and overload the related functions when adding the inference system, or supplement the existing base strategy. Maintenance to the inference system [8]. Before the system is put into use, it is necessary to do a large amount of sample training for the sample library to establish a correct and reliable sample library [9]. At the same time, the user's continuous manual participation is required to modify and maintain the sample library. This system can provide technical methods and solutions for artificial intelligence inference engines [10-11]. Since 2013, related research on the optimization performance of electric drive systems based on multi-objective evolutionary algorithms and artificial neural networks has been proposed [12]. Later, efficient

modeling of multiple-input multiple-output (MIMO) systems was proposed by timing based on the motivation of the neuro-fuzzy inference engine B [13]. Next, the research of interval-valued fuzzy relation theory based on weighted reasoning engine is put forward [14]. This kind of research provides theoretical support for the construction of the system [15]. The system itself can also be regarded as a problem-solving system, which deduces knowable knowledge from the input knowledge and establishes the deduction process for every rule deduction [16]. The application of data mining technology in the construction of case base improves the degree of automation of knowledge acquisition, partially solves the bottleneck problem of knowledge acquisition, and speeds up the development cycle of CBR system [17-18]. When dealing with the above information with the method of artificial intelligence reasoning engine, firstly, the original data is processed to fill in missing data, eliminate abnormal data, smooth noise data and so on, so as to improve the validity and accuracy of data mining process [19].

The basic characteristics of object-oriented are objectivity, encapsulation, inheritance and polymorphism. Therefore, it is convenient and feasible to construct inference strategy model under object-oriented representation. Data mining has two forms of analysis and processing, supervised learning and unsupervised learning. The former is to build corresponding models through learning on the basis of known training sample set categories; the latter refers to learning with unknown training sample categories and numbers [20]. A series of ideas and methods, such as abstract modeling based on artificial intelligence, operation interface based on a series of third-party controls, and uncertain reasoning operation, are used to realize automatic learning reasoning system [21]. The essential difference between data mining and traditional data analysis query, report and online application analysis is that data mining is to mine information and discover knowledge without explicit assumptions, and the information obtained should be previously unknown [22-23]. In addition, the object-oriented reasoning system is different from the ordinary object-oriented program. The traditional object-oriented program only objects the operation object, but does not object the operation itself. In order to solve this problem, this paper proposes to apply the data mining technology to the management and maintenance of the artificial intelligence inference engine system. The system periodically clusters the groups in the sample library to further improve the performance optimization of the artificial intelligence inference engine.

## **2. Data Mining Technology**

Data Mining (DM) refers to the process of mining or extracting knowledge from a large amount of data or information. This includes data mining and intelligent information extraction process. Data mining technology has different classification standards. According to the type of knowledge discovered, it can be divided into association rule discovery, sequential discovery, classification, clustering, outlier analysis, time series analysis and so on. Cluster analysis is one of the important research contents of data mining, which involves many research fields including data mining, statistics, biology and machine learning [24]. Search technology is a very important technology in artificial intelligence technology, which plays a very important role in data mining. Typically, data mining first extracts data from a data warehouse into a data mining library or data mart as shown in Figure 1. There are many advantages to obtaining data for mining from the data warehouse. The data cleansing of the data warehouse is similar to the data cleansing of data mining. If the data has been cleaned up when importing into the data warehouse, there is no need to clean up when mining [25]. At the same time, it is divided into different clusters, so that the distance between individuals belonging to the same cluster is as small as possible, and the distance between individuals of different clusters is as large as possible, so that a small number of clusters represent the state of the entire data set, simplifying The representation of the data set.

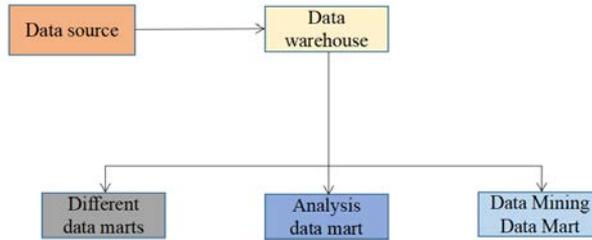


Fig. 1 Data Mining Database Derived from Data Warehouse

In data mining algorithm, the larger the value of alpha and beta, the larger the amount of calculation and the longer the calculation time. Therefore, in the case of obtaining satisfactory solutions, it is suggested that the value of alpha and beta should be relatively small. Table 1 and Figure 2 are the experimental results of the algorithm data simulation.

Table 1 Experimental results of algorithm data simulation

Value	Average value	Optimum solution
$\alpha$	62.33	88.75
$\beta$	56.12	74.96

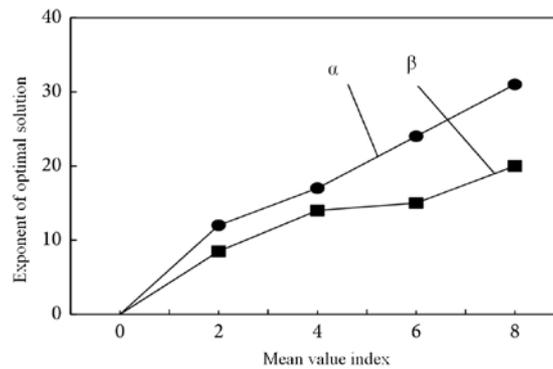


Fig. 2 Experimental results of algorithm data simulation

When using the general data mining-based AI search engine to process all users' label guesses, the system data search load will reach the maximum, because the information data needed by users is not cached, so every query and update of data must be carried out database operation. The test results are shown in Table 2 and Figure 3:

Table 2 Search performance test results

System running time	Number of word segmentation system	Number of tags
1 Days	489	362
2 Days	556	719

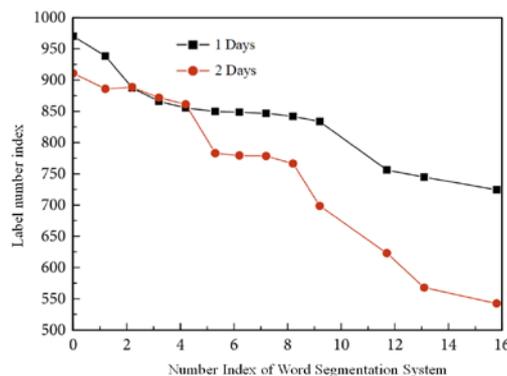


Fig. 3 Search performance test results

Symbolic representation uses various symbols with specific meanings to represent knowledge in different ways and in different orders. It is mainly used to represent logical knowledge. Connection representation is to connect various physical objects in different ways and in different order, and transfer and process various meaningful information among them. Data mining libraries may be a logical subset of data warehouses, rather than physically separate databases. For data mining, one or more transaction databases can be imported into a read database as a data mart, and then data mining can be carried out in it. Make the data in the same class the most similar, and the similarity between the class and the class is the smallest. The data processed by the cluster has no class identification, and is a method of automatically classifying according to the degree of similarity between samples without a tutor. General clustering methods and algorithms are for structured data, and general clustering methods and algorithms are for structured data. The mining of association rules in data mining uses symbolic representation. Association rule mining is to extract valuable knowledge about the interconnection between data items from a large amount of data. In the new era, artificial intelligence and data mining will become more and more intelligent, and more emphasis will be placed on increasing investment in smart technology.

At the 60MHz clock frequency, when IDMA exclusive CPM bandwidth, the performance of various modes of IDMA is shown in Table 3 and Figure 4. In the actual system, due to the existence of other serial channels in the CPM, such as the serial port and Ethernet in the system, the bandwidth utilization of the CPM cannot reach 99.99%, and the actual performance is often worse than that given in the manual.

Table 3 IDMA Performance at 60MHz

IDMA mode	Speed (MByte/s)	Utilization ratio(%)
Memory to Memory	10.55	89.76
Single Address Mode to Peripherals	6.34	97.96
Dual Address Mode to Peripherals	9.87	96.78

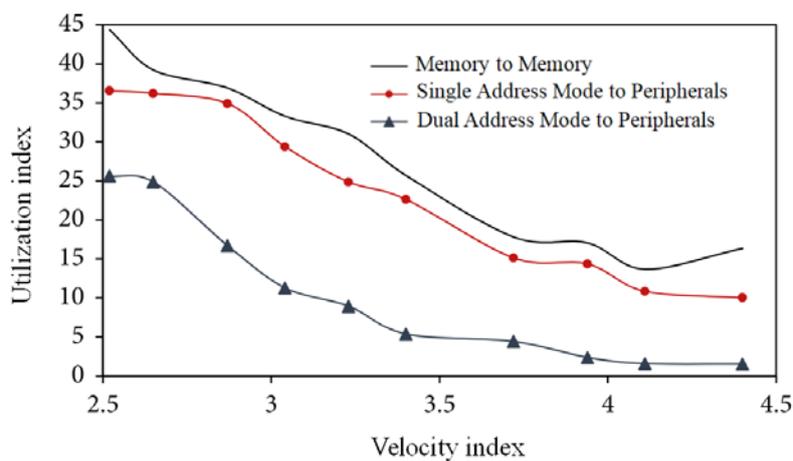


Fig. 4 IDMA Performance at 60MHz

Minimum support count is also a key problem in data mining. If the support degree is set too high and the frequent itemsets are too few to generate, the work of mining will be meaningless. After several trial analysis and comparison, it is reasonable to set the minimum support degree to 4 here. By exchanging frequent itemsets on three sets of data, the results are shown in Table 4 and Figure 5:

Table 4 Data mining results

Data code	Minimum Support Value	Number of frequent itemsets
M <sub>1</sub>	63	96
M <sub>2</sub>	74	91
M <sub>3</sub>	88	69

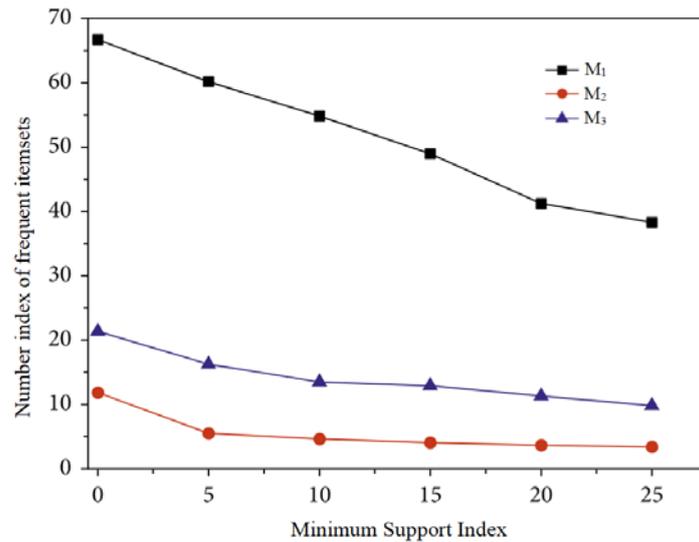


Fig. 5 Data mining results

Artificial intelligence and data mining are moving towards networking. The full use of artificial intelligence technology in the network can make the network like artificial intelligence. It also carries the characteristics of intelligence. Generally speaking, the basic idea of data mining in the process of processing is non-classical, and its "pruning" rule should be strictly proved by classical reasoning with its strict mathematical background. However, data mining is not a substitute for traditional statistical analysis technology, but an extension and extension of statistical analysis methodology. Most of the statistical analysis techniques are based on sound mathematical theory and superb skills. The accuracy of the prediction is satisfactory, but the requirements are very high. Each tuple in the database has a class identifier. It is necessary to analyze a training data set to establish a more accurate description pattern of each class, and then use this mode to classify the database. And quantizing the object into a finite number of cells to form a network structure. Usually the processing time is only related to how many cells are divided into each dimension in the quantization space. Its outstanding feature is that the processing speed is very fast. At this stage, data mining has obtained some specific acquisition, storage and analysis on the network, which can improve the visualization level of fault maintenance. At the same time, it provides a design solution for the performance optimization of the artificial intelligence inference engine.

### 3. Design of Performance Optimization Scheme for Artificial Intelligence Inference Engine

A new synthesis algorithm is proposed in this paper. Previous algorithms usually mining outlier data, then clustering analysis, then analysis, merge clustering, and finally extract cases to form a case base. The first two steps are replaced by the proposed algorithm. Figure 6 is the flow chart of the algorithm. That is, cluster analysis is carried out first, then outlier data mining is carried out. Clustering analysis will adopt improved data mining algorithm. Data information in search engine system is a very complex process. This paper proposes a model based on clustering analysis to explain the clustering process of text information. At the same time, it provides scheme design for the optimization of the performance of AI reasoning engine. The problem processing system is to initialize the user data into an object form that the system can understand, and write the initial situation into the object set, and on the other hand, after the inference system completes the reasoning process, the final result is output to the user interface. In the search of the output result, it is divided into blind search and heuristic search. The blind search is searched according to a predetermined control strategy, and the intermediate information obtained during the search process is not used to improve the control strategy. Heuristic search is to add heuristic information related to the problem in the search process, to guide the search in the most promising direction, accelerate the problem solving process, and find the optimal solution. The system should meet the requirements of entering the

inference data once and using the data multiple times. That is, the inference data is the basis of the system.

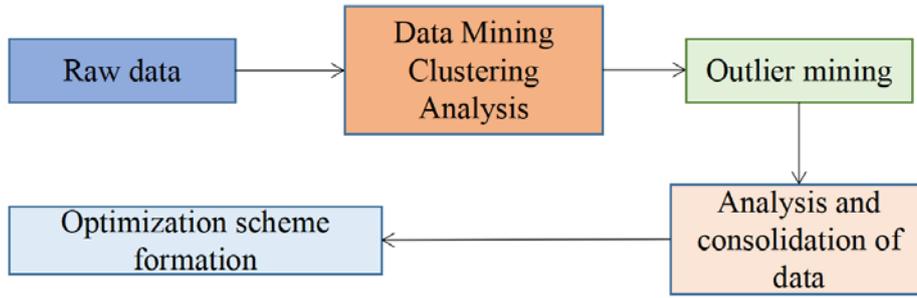


Fig. 6 Algorithm flow

Data can be regarded as information factors with different attributes. Clustering center is not only the source of information to be searched by software system, but also the process of data clustering can be regarded as the process of data being captured by search engine. Assume that the data object is:

$$\Delta y = M(t_0 + \Delta t) - M(t_0) \quad (1)$$

First, the algorithm is initialized, the pheromone of each path is set to  $t$ , the cluster radius  $R$  is set, and the statistical error is  $Q$ . Represents parameters such as object  $U$ . Calculate the pheromone  $M$  on each system path. Its expression is:

$$M(t_0 + \Delta t) = R_0 \left[ 1 + \frac{2\Delta R U}{Q_0 R_0^2} (t_0 + \Delta t) \right]^{0.5} \quad (2)$$

The heuristic function is constructed according to the distance-based attribute selection method, and the heuristic function value  $D$  of each attribute node  $X$  and  $Y$  is defined as:

$$D_k(x, y) = \begin{cases} 255 & |P_k(x, y) - B_k(x, y)| > T_h \\ 0 & \text{else} \end{cases} \quad (3)$$

Where  $P$  is the number of training set samples, and  $B$  is the number of samples in which the attribute  $t$  of the training set takes the value  $P$ . For a training set,  $N$  does not change during the data search. Therefore, in order to save computation time, this value can be calculated before the data search. Its calculation formula is:

$$N_{i0} = \frac{4i(t_0)}{\pi D^2 Z v_i e} \quad (4)$$

The roulette selection mechanism is used to select attribute nodes. For attributes that do not appear in the path, the probability of the selected attribute node  $E_c$  is calculated by the following formula:

$$E_c = 2 \sqrt{\frac{Z e N_i}{\varepsilon_0} (\sqrt{u(t) U_0 + U_0^2} - U_0)} \quad (5)$$

The selected attribute node will be added to the path. When all the attributes are included in the path, select a class label node to form a complete path, that is, a classification rule, and the rule is most effective. The validity of the rule  $P_d$  can be calculated by:

$$P_d = N_i v_i \left( \frac{M_i v_i^2}{2} + Z e u(t) \right) \quad (6)$$

Reasoning technology is based on known facts, using the knowledge already acquired, finding out the actual things involved, or summarizing new facts. The establishment of the artificial intelligence inference engine system is mainly to serve the software back-end maintenance staff and software users. After the operation and maintenance personnel need to build a series of knowledge systems such as classification and facts, all the results of the modeling are entered into the system. ALRS gets this knowledge and saves it in the database. In paradigm-based learning, the original nature of the target paradigm allows one to associate with the source paradigm in memory, but it is crude and not necessarily correct, and its analogy must be confirmed. In the system, each rule has a unique name. In reasoning, when the premise or condition of the information data search part can be satisfied, that is to say, when there are facts matching the premise or condition in the fact base, the output part of the rule will be executed. Because the reasoning strategy in the reasoning system is modular, it is necessary that the object in the object set should also follow a certain pattern. Therefore, in the object set of reasoning system, all objects come from a basic object. Usually, when using search engines, users will enter one or two keywords to query. According to the research survey, the average query sentence is 2.5 words. In addition, the inference result of the inference system is that the user adds attention to the software or annotates his hobbies, so the maintenance staff only need to input the facts they see into the artificial intelligence learning inference system to realize automatic inference.

The simplicity of rules is an important index to measure the efficiency of rules. Generally, simple rules are easier to understand and use. If the rule length is L, the path node R included in the rule has:

$$R = \omega L + \frac{1}{\omega C} \quad (7)$$

Where C is the validity of the rule. For nodes R that are not included in the rule:

$$\varepsilon = -L \frac{di}{dt} \quad (8)$$

According to the above algorithm, it is very likely that I and j have different clustering results, which is obviously unreasonable. Because the absolute error increases as the number of samples increases. Therefore, this paper proposes that the root mean square error should be used for the total error, and the formula is as follows:

$$U(x/\alpha) = \nabla \alpha_{i,j} \cdot \log P\left(\frac{X}{\lambda}\right) = \frac{\xi(x, s_j)}{\alpha_{i,j}} - \xi(s_j) \quad (9)$$

Combined with the data mining clustering algorithm, Q=3r is selected here, where r is the clustering radius in the data mining clustering algorithm, and q is the specified outlier criterion. If rt is outlier data if and only if it satisfies:

$$r_i(i) = P(q_t = s_i | y_i) \quad (10)$$

The distribution of the components of the sample data X in the simulation is (t,-t) uniform random distribution, where r = P, so the weight factor can be set to I (t). Because the weight factor is used, the simulation results are P-independent.

$$i(t) = \frac{1}{1 + \left(\frac{1}{i_0} - 1\right) \cdot e^{-\lambda t}} \quad (11)$$

Which represents the probability that t is merged into eyj, and Pi is the closed value set by the program. Since the algorithm is based on a weighted Euclidean distance, the weight Pi should also be introduced in the calculation of Yi. Therefore, the calculation formula in the above formula is modified as follows:

$$P_i = F(Y_i) = \frac{e^{Y_i}}{1 + e^{Y_i}} \quad (12)$$

The basic idea of the search engine association algorithm is to analyze the query history of the user in detail to mine the keywords with high correlation with those input keywords, so as to form the related vocabulary of the query and narrow the scope of the query. At the same time, in the object-oriented reasoning system, emphasizing the objectification of the strategy can greatly improve the maintenance efficiency of the inference system and reduce the maintenance difficulty. In the design of the finishing part, it is necessary to sort out the results of the previous work reasoning, and sort out the feedback of the software users, thereby obtaining the user's needs and the potential needs of the users. The inference engine is actually a collection of inference programs. It is mainly responsible for inferring the user's input and using the constructed reasoning tree and inference algorithm to derive the user's tendency. The clustering result is divided into a training set and a test set, and the label attribute in the training set is regarded as known, and the label attribute in the test set is unknown, and the category value of the clustering result in the test set is stored as the actual category value. At the same time, the outlier data is mined, the outlier data is extracted, and the outlier data is extracted so as to make the latter clustering have better effect. Even in the process of rule execution, the result of rule execution will not affect the current state of fact base, that is, it will not add, modify and delete the facts in fact base, but in knowledge inference network, it will record the reasoning process of this rule. It is a relatively new and challenging research field to apply data mining technology to search engines. This paper mainly analyses the application of clustering algorithm in data mining in search engines. Thus, it provides theoretical possibility for the design of performance optimization scheme of AI reasoning engine.

#### 4. Conclusions

In summary, with the development of information age, the use of software users is increasing, which requires the full play of the role of artificial intelligence in software data mining. To solve this problem, a new synthesis algorithm is proposed in this paper. Previous algorithms usually mining outlier data, then clustering analysis, then analysis, merge clustering, and finally extract cases to form a case base. The first two steps of the proposed algorithm are replaced by clustering analysis and outlier data mining. Clustering analysis will adopt improved data mining algorithm. It solves the problem of inaccuracy of the number of clusters in partitioning problem and can only be used to mine spherical clustering, so that partitioning method can also be used to discover clustering of arbitrary shape. At the same time, the user interest model is improved, and a new keyword expansion algorithm is implemented. In case-based reasoning, some decision theory can be embedded, as in the choice of alternative paradigms in the sample acquisition process. It can be seen that decision theory can be effectively applied to case-based reasoning, which makes up for some shortcomings in case-based reasoning. In the future work, methods and techniques such as machine learning, fuzzy reasoning, and uncertainty reasoning can be added to the inference system to make the inference system have more powerful decision-making ability and make the inference system more intelligent.

#### References

- [1] Mumini O O, Williams S O, John A E. A Genetic-Neuro-Fuzzy Inferential Model for Diagnosis of Tuberculosis [J]. *Applied Computing & Informatics*, 2017, 13(1):27-37.
- [2] Hybrid artificial intelligence approach based on neural fuzzy inference model and metaheuristic optimization for flood susceptibility modeling in a high-frequency tropical cyclone area using GIS[J]. *Journal of Hydrology*, 2016, 540:317-330.
- [3] Ghugare S B, Tiwary S, Tambe S S. Computational intelligence based models for prediction of elemental composition of solid biomass fuels from proximate analysis [J]. *International Journal of System Assurance Engineering and Management*, 2014:1-14.

- [4] Jacques J, Preda C. Model-based clustering for multivariate functional data [J]. *Computational Statistics & Data Analysis*, 2014, 71(3):92-106.
- [5] Quwaider M. Cloudlet-based efficient data collection in wireless body area networks [J]. *Simulation Modelling Practice & Theory*, 2015, 50:57-71.
- [6] Tea Tušar, Bogdan Filipič. Visualization of Pareto Front Approximations in Evolutionary Multiobjective Optimization: A Critical Review and the Prosection Method [J]. *IEEE Transactions on Evolutionary Computation*, 2015, 19(2):225-245.
- [7] Paudel S, Elmtiri M, Kling W, et al. Pseudo Dynamic Transitional Modeling of Building Heating Energy Demand Using Artificial Neural Network[J]. *Energy & Buildings*, 2014, 70(2):81-93.
- [8] Akoglu L, Tong H, Koutra D. Graph-based Anomaly Detection and Description: A Survey [J]. *Data Mining & Knowledge Discovery*, 2014, 29(3):626-688.
- [9] Stalidis G, Karapistolis D, Vafeiadis A. Marketing Decision Support Using Artificial Intelligence and Knowledge Modeling: Application to Tourist Destination Management[J]. *Procedia - Social and Behavioral Sciences*, 2015, 175:106-113.
- [10] Ma C, Xin M, Feldmann K A, et al. Machine Learning-Based Differential Network Analysis: A Study of Stress-Responsive Transcriptomes in Arabidopsis [J]. *The Plant Cell*, 2014, 26(2):520-537.
- [11] Acampora G, Pedrycz W, Vasilakos A V. Efficient modeling of MIMO systems through Timed Automata based Neuro-Fuzzy Inference Engine [J]. *International Journal of Approximate Reasoning*, 2014, 55(6):1336-1356.
- [12] ZVoianu A C, Bramerdorfer G, Lughofer E, et al. Hybridization of multi-objective evolutionary algorithms and artificial neural networks for optimizing the performance of electrical drives [J]. *Engineering Applications of Artificial Intelligence*, 2013, 26(8):1781-1794.
- [13] Lim C K, Chan C S. A weighted inference engine based on interval-valued fuzzy relational theory[J]. *Expert Systems with Applications*, 2015, 42(7):3410-3419.
- [14] Garg V K, Bansal R K, Garg V K, et al. Intelligent Computing Techniques for the Detection of Sleep Disorders: A Review [J]. *International Journal of Computer Applications*, 2015, 110(1):27-46.
- [15] Stylios G, Katsis C D, Christodoulakis D. Using Bio-inspired intelligence for Web opinion Mining [J]. *International Journal of Computer Applications*, 2014, 87(5):217-23.
- [16] Chougali S, Roudane A, Mansouri K, et al. New air traffic management approach based on expert system and using real-time scheduling algorithms [J]. *International Journal of Intelligent Engineering Informatics*, 2016, 4(3/4):305-321.
- [17] Maglogiannis I, Karpouzis K. Artificial Intelligence Applications and Innovations [J]. *Artificial Intelligence Review*, 2013, 42(3):309-311.
- [18] Nieves J C, Espinoza A, Peña Y K, et al. Intelligence distribution for data processing in smart grids: A semantic approach.[J]. *Engineering Applications of Artificial Intelligence*, 2013, 26(8):1841-1853.
- [19] Mahapatra G, Banerjee S. A Study of Bacterial Foraging Optimization Algorithm and its Applications to Solve Simultaneous Equations [J]. *International Journal of Computer Applications*, 2013, 72(5):1-6.
- [20] Khuntia S R, Panda S. ANFIS approach for SSSC controller design for the improvement of transient stability performance [J]. *Mathematical and Computer Modelling*, 2013, 57(1-2):289-300.
- [21] Ye F, Chen Q, Blanckaert K, et al. Riparian vegetation dynamics: insight provided by a process-based model, a statistical model and field data[J]. *Ecohydrology*, 2013, 6(4):567-585.
- [22] Taimoor K M, Mehr D, Shehzad K, et al. Online Knowledge-Based Model for Big Data Topic

Extraction:[J]. Computational Intelligence & Neuroscience, 2016, 2016(4-5):1-10.

[23] Gianni M, Kruijff G J M, Pirri F. A Stimulus-Response Framework for Robot Control [J]. ACM Transactions on Interactive Intelligent Systems, 2015, 4(4):1-41.

[24] Fu J, Tanner H G, Heinz J N, et al. Symbolic planning and control using game theory and grammatical inference[J]. Engineering Applications of Artificial Intelligence, 2015, 37:378-391.

[25] Gheisari S, Meybodi M R. BNC-PSO: Structure Learning of Bayesian Networks by Particle Swarm Optimization [J]. Information Sciences, 2016, 348:272-289.