American Journal of Systems and Software, 2014, Vol. 2, No. 5, 127-130
Available online at http://pubs.sciepub.com/ajss/2/5/3

DOI:10.12691/ajss-2-5-3

Performance and Cost Evaluation of Query Plans from the Student Database Using Specific G.A. Technique

Mishra Sambit Kumar[1,*], Pattnaik Srikanta[2], Patnaik Dulu[3]

[1]Department of Computer Sc. & Engg, Ajay Binay Institute of Technology, Cuttack, Odisha, India
[2]S.O.A. University, Bhubaneswar, Odisha, India
[3]Government College of Engineering, Bhwanipatna, Odisha, India
*Corresponding author: sambit_pr@rediffmail.com

Received October 20, 2014; Revised November 07, 2014; Accepted November 10, 2014

Abstract Many educational institutions in India have already established online teaching and learning methodologies with different capabilities and approaches. After inspired from foreign universities, they have successfully adopted the learning online network with computer assisted personalized approaches. Usually, two kinds of large data sets are involved with the system, e.g. educational resources such as web pages, demonstrations, simulations, and individualized problems designed for use on homework assignments, and information about users who create, modify, assess, or use these resources. Genetic Algorithms (GAs) may be implemented as an effective tool to use in pattern recognition. The important aspect of GAs in a learning context is their use in pattern recognition. There are two different approaches for application of GA in pattern recognition. First of all apply a GA directly as a classifier. In this case G.A. may be applied to find the decision boundary in N dimensional feature space. Then use a GA as an optimization tool for resetting the parameters in other classifiers. Most applications of GAs in pattern recognition optimize some parameters in the classification process. In many applications of GAs, feature selection has been used. GAs has also been applied to find an optimal set of feature weights which improve classification accuracy. In this paper, it is intended to use a GA to optimize a combination of classifiers. The objective is to predict the students' semester grades of a reputed Engineering college of India based on some acquired features. It is also intended to evaluate the size of each chromosome, e.g. student level at each query level as well as cost of query plans which may be associated with the student level. The objective is also aimed to evaluate the performance of the queries as well as query plans associated with the student database.

Keywords: *query plan, pattern, plan select, function_value, g.a., classifier, web based learning*

Cite This Article: Mishra Sambit Kumar, Pattnaik Srikanta, and Patnaik Dulu, "Performance and Cost Evaluation of Query Plans from the Student Database Using Specific G.A. Technique." *American Journal of Systems and Software*, vol. 2, no. 5 (2014): 127-130. doi: 10.12691/ajss-2-5-3.

1. Introduction

Now a day, the Internet has become an effective medium which has changed completely and irreversibly the way information and knowledge are transmitted and shared throughout the world. The education community has been at the forefront of most of the changes. The web-based education and e-teaching with large amounts of information describes the continuation of the teaching-learning interactions which are endlessly generated. This type of problems may be solved through Genetic algorithm by evaluating a series of pattern classifiers. It is understood that not just as a collection of data analysis methods, but as a data analysis process it encompasses anything from data understanding, pre-processing and modelling to process evaluation and implementation. The implementation of Genetic algorithm with a series of pattern classifiers may be used to extract knowledge from e-learning systems through the analysis of the information available in the form of data generated by their users. In this case, the main objective becomes finding the patterns of system usage by teachers and students and, perhaps most importantly, discovering the students' learning behavior patterns. It aims to provide an as complete as possible review of the many applications of Data Mining to e-learning.

2. Review of Literature

Guerra-Salcedo C. et. al. [1] have discussed in their paper that most applications of GAs in pattern recognition optimize some parameters in the classification process. GAs has been applied to find an optimal set of feature weights that improve classification accuracy.

Freitas et. al. [2] have used a GA for selecting the features as well as selecting the types of individual classifiers in their design of a Classifier Fusion System. GA is also used in selecting the prototypes in the case-based classification. Genetic Algorithms have already been proved to be an effective tool to use in data mining and pattern recognition.

Chapman et. al [3] have discussed in their paper about e-learning which is rather difficult to define. Not because of its intrinsic complexity, but because it has most of its roots in the ever-shifting world of business. It can be understood not just as a collection of data analysis methods, but as a data analysis process that encompasses anything from data understanding, pre-processing and modelling to process evaluation and implementation.

Ha, S.H et. al [4] have reviewed the possibilities of the application of Web Mining techniques to meet some of the current challenges in distance education. They analyzed that the proposed approach could improve the effectiveness and efficiency of distance education in two ways. On the one hand, the discovery of aggregate and individual paths for students could help in the development of effective customized education, providing an indication of how to best organize the educator organization's courseware.

Srivastava et. al [5] have discussed in their paper that application of data mining techniques to analyze Web logs, in order to discover useful navigation patterns, or deduce hypotheses can be used to improve web applications, is the main idea behind Web usage mining. Web usage mining can be used for many different purposes and applications such as user profiling and Web page personalization, server performance enhancement, Web site structure improvement, etc.

Tang, T.Y et. al [6] have discussed in their paper about how Data Mining techniques could successfully be incorporated to e-learning environments and how they could improve the learning tasks were carried out. In their paper data clustering was suggested as a means to promote group-based collaborative learning and to provide incremental student diagnosis.

Stathacopoulou. G.D et. Al [7] have suggested a neuro-fuzzy model for the evaluation of students in an intelligent tutoring system. Fuzzy theory was used to measure and transform the interaction between the student and the ITS into linguistic terms. Then, Artificial Neural Networks were trained to realize fuzzy relations operated with the max-min composition. These fuzzy relations represent the estimation made by human tutors of the degree of association between an observed response and a student characteristic.

Hwang, G.J et. al [8] have developed a fuzzy rules-based method for eliciting and integrating system management knowledge that is proposed and served as the basis for the design of an intelligent management system for monitoring educational Web servers. This system is capable of predicting and handling possible failures of educational Web servers, improving their stability and reliability. It assists students' self-assessment and provides them with suggestions based on fuzzy reasoning techniques.

Markham, S. et. al [9] have proposed software agents as an alternative for data extraction from e-learning environments, in order to organize them in intelligent ways. The approach includes pedagogical agents to monitor and evaluate Web-based learning tools, from the educational intentions point of view.

Liang, A et. al [10] have proposed an e-learning model for the personalization of courses, based both on the student's needs and capabilities and on the teacher's profile. Personalized learning paths in the courses were modeled using graph theory. They have applied decision trees as classification models. The implementation of the distance learning algorithm uses rough set theory to find general decision rules. The decision tree was used to adequate the original algorithm to distance learning issues. On the basis of the obtained results, the instructor might consider the reorganization of the course materials.

Prentzas. J A et. al [11] have developed a system that evaluates the students' performance in Web based e-learning. Its functioning is controlled by an expert system using "neurules": a hybrid concept that integrates symbolic rules and neural computing. Internally, each "neurule" is represented and considered as an Adaline neuron.

Van Rosmalen et, al [12] have focused in their paper that practically many systems claim to be innovative and stress the importance of content but, they hardly provide any information about which didactical methods and models they implement; it is therefore difficult to assess them. As far as adaptation is an integral part of the systems, it would require extensive customization. Most of the surveyed systems do support collaborative learning tasks; however they do not allow the use of any specific scenario.

O. Gorlitz et. Al [13] have discussed in their paper that in case of distributed database systems, an optimizer needs to identify relevant sources and determine subqueries that can be evaluated directly at the sources without having to transfer the actual data.

O. Hartig et. al. [14] have focused on explorative query processing, that exploits the linked data by iteratively evaluating and downloading data for URIs representing results for parts of the query. The process requires only little cooperation from the sources. Instead of determining relevant sources based on intermediate query results, alternatively indexes describing the data provided by the sources may also be used.

K. Hose et. al [15] have discussed in their paper that the scalable query processing may only be achieved in a distributed environment. A main condition in order to hold this assumption is that global query optimization may take place.

N. Papailiou et. al [16] have focused on multiple query processing that adopts the idea of data warehouses, collects all the data in advance and combines it into a centralized triple store. The centralized query processing along with the cluster technology has already been proposed to increase the efficiency of these solutions.

3. Problem Formulation

The test data of the students of 5th Semester Computer Sc.& Engg. of an Engineering college have already been selected..

The students may be grouped regarding their final grades in many ways, out of which Some are defined as follows.

1. Suppose the 5 possible category labels be the same as students' grades, as referred in table-3.1.

2. The students in relation may be labeled as per their grades and grouped into three classes, "high" representing grades greater than 7.0, "middle" representing grades from

6.0 to 7.0, and "low" representing grades less than 5.0, as referred in table 3.2.

Table 3.1. Selecting 5 category labels regarding to students' grades in 5th Semester (CS& Engg.)

Category	SGPA	No.of Students	Percentage(%)
1	4.5	2	4
2	6.0	12	20
3	7.0	24	40
4	8.0	18	30
5	9.0	04	7

Table 3.2. Selecting 3 class labels regarding to students' grades in 5th Semester (CS& Engg.)

Category	SGPA	No. of Students	Percentage(%)
High	>=7.0	46	77
Middle	>6.0 and <7.0	24	40
Low	<5.0	2	4

It may be predicted that the error rate in the first class grouping should be higher than the others, because the sample size of the grades over 5 categories differs considerably. It is clear that the less data for the first three categories in the training phase are available, and so the error rate may likely be higher in the evaluation phase.

In general, Pattern recognition may have variety of applications, such that it may not possible to come up with a single classifier which may produce better results in all the cases. The optimal classifier in every case is usually dependent on the problem domain. So it may be possible to come across a case where no single classifier may classify with an acceptable level of accuracy. In such cases it may be better to accumulate the results of different classifiers to achieve the optimal accuracy.

3.1. Algorithm

Table 3.3. Evaluating Function_value and cost of query plan

Sl.No.	Size of query Plan	Plan_select value	Function_value	Est_cost of query plan
1	921	33.491	0.038008	0.19818
2	1070	38.909	0.038261	0.20088
3	1099	39.969	0.0383	0.20092
4	1340	48.727	0.038562	0.21418
5	1677	60.982	0.039036	0.21939
6	1955	71.091	0.039375	0.22422
7	1984	72.165	0.039356	0.22922

Step 1: The maximum generation considered=20;

Step 2: The total number of relations related to student database=20;

Step 3: The total number of queries =50;

Step 4: Size of Chromosome (Student record at each query level) =11;

Step 5: Population=round (rand (total number of queries, size of chromosome));

Step 6: Crossover probability, pc=0.9; Mutation probability, pm=0.007;

Step 7: Perform Crossover operation

Step 8: Perform Mutation operation

Step 9: Evaluate size of query plan related to student record at each query level

Step 10: Evaluate the plan select value by considering the query plans along with total number of queries and size of chromosome

Step 11: Calculate the cost of query plans by monitoring the CPU time and considering plan select value and size of query plan from the total number of queries;

Step 12: Evaluate the Function_value by calculating the weight of the query plans as well as plan fitness value

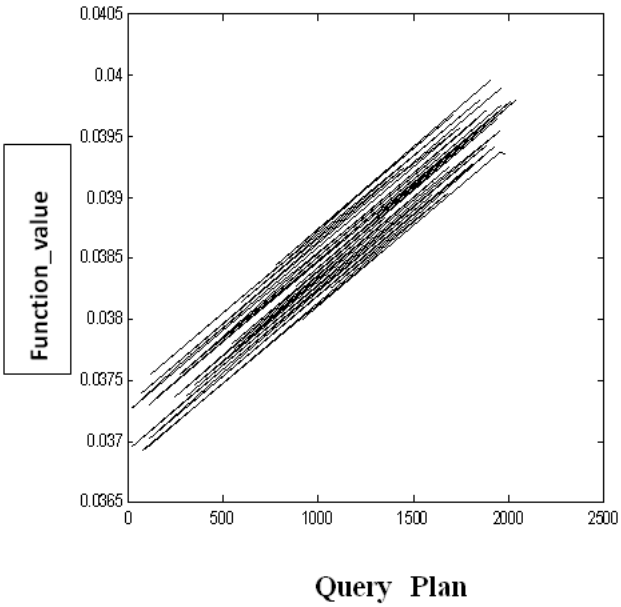

Figure 3.1. Query Plan VS Function_value

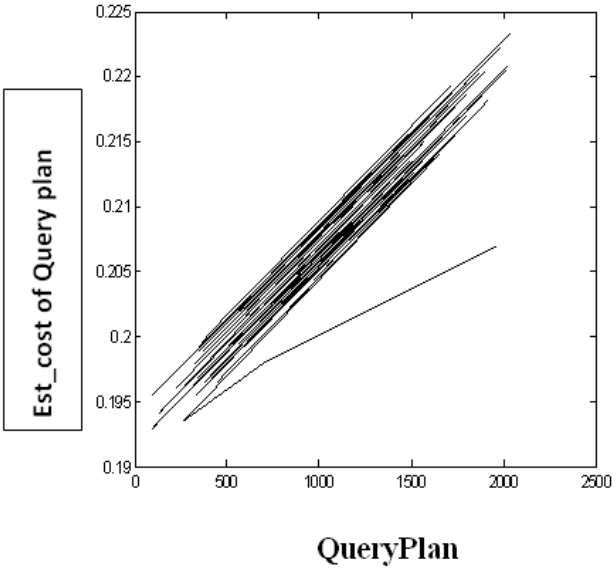

Figure 3.2. Query Plan VS Est_cost of plan

4. Experimental Analysis

In this case 20 number of relations related to student database along with 50 number of queries are considered.

The minimum size of student record at each query level is considered as 11.

It is seen that with the increasing size of query plans, the plan select value is directly proportional to the query plans. Also the size of query plans is directly proportional to the function_value retrieved from the student database with considerable number of queries and relations.

The estimated cost of query plans is also dependent over the number of queries as well as the size of student record within the relation and corresponding CPU time.

5. Discussion and Future Direction

One of the most difficult and time-consuming activities for teachers in distance education courses is the evaluation process, due to the fact that, in this type of course, the review process is better accomplished through collaborative resources such as email, discussion forums, chats, etc. As a result, this evaluation has usually to be carried out according to a large number of parameters, whose influence in the final mark is not always well defined or understood.

It has been observed that the e-learning course offerings are very much fruitful, and quite new e-learning platforms and systems have been developed and implemented now a day. These systems generate an exponentially increasing amount of data to improve all instances of e-learning.

6. Conclusion

It has been seen that while considering the e-learning problems, most work deals with students' learning assessment, learning materials and course evaluation, and course adaptation based on students' learning behavior. In this paper, the size of student record at each query level has been found out. Also the cost of query plan at each query level of a signified relation has been evaluated. It is also seen that the cost of query plan is directly proportional to the query plan associated with each query level.

References

[1] Guerra-Salcedo C. and Whitley D. "Feature Selection mechanisms for ensemble creation: a genetic search perspective". Freitas AA (Ed.) Data Mining with Evolutionary Algorithms: Research Directions-Papers from the AAAI Workshop, 13-17. Technical Report WS-99-06. AAAI Press, 1999.

[2] Freitas, A.A. "A survey of Evolutionary Algorithms for Data Mining and Knowledge Discovery", See: www.pgia.pucpr.br/~alex/papers. To appear in: A. Ghosh and S. Tsutsui. (Eds.) Advances in Evolutionary Computation. Springer-Verlag, 2002.

[3] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R.: CRIPS-DM 1.0 Step by Step Data Mining Guide. CRISP-DM Consortium (2000).

[4] Ha, S.H., Bae, S.M., Park, S.C.: Web Mining for Distance Education. In: IEEE International Conference on Management of Innovation and Technology, ICMIT'00. (2000) 715-719.

[5] Srivastava, J., Cooley, R., Deshpande, M., Tan, P.: Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data. ACM SIGKDD Explorations. 1 (2) (2000)12-23.

[6] Tang, T.Y., McCalla, G.: Smart Recommendation for an Evolving e-Learning System: Architecture and Experiment. International Journal on e-Learning 4 (1) (2005) 105-129.

[7] Stathacopoulou, G.D., Grigoriadou, M.: Neural Network-Based Fuzzy Modeling of the Student in Intelligent Tutoring Systems. In: International Joint Conference on Neural Networks. Washington (1999) 3517-3521.

[8] Hwang, G.J., Judy, C.R., Wu, C.H., Li, C.M., Hwang, G.H.: Development of an Intelligent Management System for Monitoring Educational Web Servers. In: 10th Pacific Asia Conference on Information Systems, PACIS 2004. (2004) 2334-2340.

[9] Markham, S., Ceddia, J., Sheard, J., Burvill, C., Weir, J., Field, B., et al.: Applying Agent Technology to Evaluation Tasks in e-Learning Environments. In: Proceedings of the Exploring Educational Technologies Conference. Monash University, Melbourne, Australia (2003) 31-37.

[10] Liang, A., Ziarco, W., Maguire, B.: The Application of a Distance Learning Algorithm in Web-Based Course Delivery. In: Ziarko, W., Yao, Y. (eds.): Second International Conference on Rough Sets and Current Trends in Computing. Lecture Notes in Computer Science. Springer-Verlag, Berlin Heidelberg New York (2000) 338-345.

[11] Prentzas, J., Hatzilygeroudis, I., Garofalakis, J.: A Web-based Intelligent Tutoring System Using Hybrid Rules as its Representational Basis. In: Cerri, S.A., et al. (eds.): Intelligent Tutoring Systems, ITS 2002. LNCS Vol. 2363. Springer-Verlag, Berlin Heidelberg New York (2002) 119-128.

[12] Van Rosmalen, P., Brouns, F., Tattersall, C., Vogten, H., van Bruggen, J., Sloep, P., Koper, R.: Towards an Open Framework for Adaptive, Agent-Supported e-Learning. International Journal Continuing Engineering Education and Lifelong Learning 15 (3-6) (2005) 261-275.

[13] O. Gorlitz and S. Staab. SPLENDID: SPARQL Endpoint Federation Exploiting VOID Descriptions. In COLD' 11, 2011.

[14] O. Hartig. Zero-Knowledge Query Planning for an Iterator implementation of Link Traversal Based Query Execution. In ESWC (1), pages 154{169, 2011.

[15] K. Hose, R. Schenkel, M. Theobald, and G. Weikum. Database Foundations for Scalable RDF Processing. In Reasoning Web, volume 6848 of Lecture Notes in Computer Science, pages 202{249. Springer, 2011.

[16] N. Papailiou, I. Konstantinou, D. Tsoumakos, and N. Koziris. H2RDF: adaptive query processing on RDF data in the cloud. In WWW, pages 397{400, 2012.