

July 2013

Application of Data Mining Using Bayesian Belief Network To Classify Quality of Web Services

M.Swami Das

Computer Science Department, MLR Institute of Technology, Dundigal, Hyderabad-500 043, AP, India., msdas.520@gmail.com

Ramakanta Mohanty

Computer Science Department, MLR Institute of Technology, Dundigal, Hyderabad-500 043, AP, India., ramakanta5a@gmail.com

D. Vijayalakshmi

Department of IT, MGIT, Gandipat, Hyderabad – 500 075,(AP) India, psridhar_babu_2k@yahoo.com

Govardhan A

JNTUH College of Engineering, Karimnagar, Andrapradesh, India., govardhan_cse@yahoo.co.in

Follow this and additional works at: <https://www.interscience.in/ijcsi>



Part of the [Computer Engineering Commons](#), [Information Security Commons](#), and the [Systems and Communications Commons](#)

Recommended Citation

Das, M.Swami; Mohanty, Ramakanta; Vijayalakshmi, D.; and A, Govardhan (2013) "Application of Data Mining Using Bayesian Belief Network To Classify Quality of Web Services," *International Journal of Computer Science and Informatics*: Vol. 3 : Iss. 1 , Article 1.

DOI: 10.47893/IJCSI.2013.1109

Available at: <https://www.interscience.in/ijcsi/vol3/iss1/1>

This Article is brought to you for free and open access by the Interscience Journals at Interscience Research Network. It has been accepted for inclusion in International Journal of Computer Science and Informatics by an authorized editor of Interscience Research Network. For more information, please contact sritampatnaik@gmail.com.

Application of Data Mining Using Bayesian Belief Network To Classify Quality of Web Services

¹M.Swami Das, ²Ramakanta Mohanty, ³D.Vijayalakshmi & ⁴A. Govardhan

^{1&2}Computer Science Department, MLR Institute of Technology, Dundigal, Hyderabad-500 043, AP, India.

³Department of IT, MGIT, Gandipat, Hyderabad – 500 075,(AP) India

⁴SIT, JNT University, Hyderabad 500 085, A. P, India

E-mail : msdas.520@gmail.com, ramakanta5a@gmail.com, psridhar_babu_2k@yahoo.com, govardhan_cse@yahoo.co.in

Abstract - In this paper, we employed Naïve Bayes, Augmented Naïve Bayes, Tree Augmented Naïve Bayes, Sons & Spouses, Markov Blanket, Augmented Markov Blanket, Semi Supervised and Bayesian network techniques to rank web services. The Bayesian Network is demonstrated on a dataset taken from literature. The dataset consists of 364 web services whose quality is described by 9 attributes. Here, the attributes are treated as criteria, to classify web services. From the experiments, we conclude that Naïve based Bayesian network performs better than other two techniques comparable to the classification done in literature.

Keywords- *Web Services, Quality of Services (QoS), Bayesian Network, Data Mining, Naïve based Bayesian, Markov Blanket, Supervised Learning*

I. INTRODUCTION

Services are tendered and availed in almost all the business and industries. The growth and proliferation of IT across industries and business appear to have fuelled the requirement as well as the delivery of services profoundly. Delivering services has been an attractive business proposition for many industries lately. The latest development in the systems is a new paradigm, called web services [22]. Web Services heralded another significant mile stone in the history of IT. Earlier, Internet catered mostly to the business to Customer (B2C) category of the users on the web. As against this, Web Services enable B2B interaction as well Web. They are independent of platform and natural languages, which is suitable for accessing from heterogeneous environments. With the rapid introduction of web services technologies, researchers focused more on the functional and interfacing aspects of web services, which include HTTP and XML-based messaging. They are used to communicate by employing pervasive standards based web technologies. Web services are based on XML and three other core technologies: WSDL, SOAP, and UDDI. WSDL is a document which describes the services' location on the web and the functionality the service provides. Information related to the web service is to be entered in a UDDI registry, which permits web service consumers to find out and

locate the services they required. With the help of the information available in the UDDI registry based on the web services, client developer uses instructions in the WSDL to construct SOAP messages for exchanging data with the service over HTTP attributes. A Web service is a software application that can be accessed remotely using XML-based languages. It represents a communication interface offered by the server, through that the clients (programs on other systems) may require different information [23].

In this study, we address this problem of efficiently identifying a set quality attributes by employing Bayesian Networks vs. Naive Bayes, Markov Blanket and search Bayesian. Naive Bayes is special form of Bayesian network that is widely used for classification [9] and clustering [6], but its potential for general probabilistic modeling (i.e., to answer joint, conditional and marginal queries over arbitrary distributions) remains largely unexploited. Naive Bayes represents a distribution as a mixture of components, where within each component all variables are assumed independent of each other. The Markov Blanket of a variable Y , $MB(Y)$, by definition, is the set of variables such that Y is conditionally independent of all the other variables given $MB(Y)$. A *Markov Blanket Directed Acyclic Graph* (MB DAG) is a Directed Acyclic Graph over that subset of variables. When the parameters of the MB

DAG are estimated, the result is a *Bayesian Network*, defined in the next section. Recent research by the machine learning community [13, 14, 15] has sought to identify the Markov Blanket of a target variable by filtering variables using statistical decisions for conditional independence and using the MB predictors as the input features of a classifier. However, learning MB DAG classifiers from data is an open problem [7]. There are several challenges: the problem of learning the graphical structure with the highest score (for a variety of scores) is NP hard [7] for methods that use conditional independencies to guide graph search, identifying conditional independencies in the presence of limited data is quite unreliable the effectiveness of the MB DAG as a classifier, with conventional Bayesian updating and a heuristic method of estimating parameters.

Classification using the Markov Blanket of a target variable in a Bayesian Network has important properties: it specifies a statistically efficient prediction of the probability distribution of a variable from the smallest subset of variables that contains all of the information about the target variable; it provides accuracy while avoiding over fitting due to redundant variables; and it provides a classifier of the target variable from a reduced set of predictors. The TS/MB procedure proposed in this paper allows us to move through the search space of Markov Blanket structures quickly and escape from local optima, thus learning a more robust structure.

The rest of paper is organized as follows: Section II presents the quality issues in web services and QWS dataset. Section III describes the overview of Bayesian Network Section IV presents the results and discussions, and Section V concludes the paper.

II. QUALITY ISSUES IN WEB SERVICES

QoS plays an important role in finding out the performance of web services. Earlier, QoS has been used in networking and multimedia applications. Recently, there is a trend in adopting this concept to web services [21]. The basic aim is to identify the QoS attributes [5, 12, 19, 29] for improving the quality of web services through replication services [5], load distribution [8], and service redirection [3]. To measure the QoS of a web service, attributes like Response Time, Throughput, Availability, Reliability, Cost, and Response Time are considered.

A. QoS attributes

According to Kalepu et al. [7], quality of service (QoS) is a combination of several qualities or properties of a service, such as: (i) Availability (ii) Reliability (iii) Price (iv) Throughput (v) Response Time (vi) Latency

(vii) Performance (viii) Security (ix) Regulatory (x) Accessibility (xi) Robustness/Flexibility (xii) Accuracy (xiii) Servability (xiv) Integrity and (xv) Reputation. QoS parameters determine the performances of the web services and find out which web services are best and meet user's requirements.

Users of web services are not human beings but programs that send requests for services to web service providers. QoS issues in web services have to be evaluated from the perspective of the providers of web services (such as the airline-booking web service) and from the perspective of the users of these services (in this case, the travel agent site) [4]. There are other models available related to the quality of web services issues. A QoS model [4] represented in Table 2 shows that the main classification of QoS attributes is based on internal attributes, which are independent of the service environment, and external attributes that are dependent on the service environment. The attributes of the model in Table 3 are almost similar to the attributes of QWS Dataset used in this paper.

B. Description of QWS Dataset

QWS dataset [18] consists of different web service implementations and their attributes as presented in Table 3. The classification is measured based on the overall quality rating provided by all the attributes. The functionality of the web services can be helpful to differentiate between various services. The attributes G1 to G10 are used as explanatory variables and the attribute G11 is used as the target variable. However, attributes G12 and G13 are ignored as they do not contribute to the analysis.

The web services in the QWS dataset are classified into four categories, such as (i) Platinum (high quality) (ii) gold (iii) silver and (iv) bronze (low quality). The classification is measured based on the overall quality rating provided by WSRF. It is grouped into a particular web service based on classification. The functionality of the web services can be helpful to differentiate between various services [3].

III. OVERVIEW OF BAYESIAN NETWORKS

A Bayesian network is a directed acyclic graph model that represents conditional independencies between a set of variables [17]. It has two constituents: One is a network graphical structure which is a directed acyclic graph with the nodes of variables and arcs of relations. The other is the conditional probability table associated with each node in the model graph. Machine learning techniques are able to estimate the structure and the conditional probability table from the training data. Based on the Bayesian probability inference, the conditional probability can be estimated from the

statistical data and propagated along the links of the network structure to the target label. By setting a threshold of confidence, the final probability value can be used as the indication for the classification decision. The Bayesian formula can be mathematically expressed as below:

$$P(H | \vec{E}) = \frac{P(\vec{E} | H) \times P(H)}{\sum_{i=1}^n P(\vec{E} | H_i) \times P(H_i)} \quad (1)$$

j = 1. 2n

According to the basic statistical theory, e.g., the Chain Rule and independency relation derived from the network structure, the joint probability of \vec{E} can be calculated by the production of local distributions with its parent nodes, i.e.

$$P(\vec{E}) = \prod_{i=1}^n P(E_i | \text{ParentOf}(E_i)) \quad (2)$$

In the above formulas,

\vec{E} denotes a set of variable values, i.e.

$\vec{E} = \{E_1, E_2 \dots E_n\}$. H is termed as hypothesis. H is

called the prior probability and $P(H | \vec{E})$ is called posteriori probability of H given \vec{E} . If E_i has no parent nodes, $\text{ParentOf}(E_i)$ is equal to $P(E_i)$. The basic graph of Bayesian network are presented in Fig. 1 and the graph of QWS dataset for naïve Bayes are depicted in Fig.4. and Fig.5.

B. MARKOV BLANKET

The Markov Condition implies that the joint distribution P can be factorized as a product of conditional probabilities, by specifying the distribution of each node conditional on its parents [17]. In particular, for a given DAG S , the joint probability distribution for X can be written as

$$P(X) = \prod_{i=1}^n P_i(X | Pa_i) \quad (3)$$

where Pa_i denotes the set of parents of X_i in S ; this is called a Markov factorization of P according to S . The set of distributions represented by S is the set of distributions that satisfy the Markov condition for S . If P is faithful to the graph S , then given a Bayesian Network (S, P) , there is a unique Markov Blanket for Y consisting of Pa_Y , the set of parents of Y , ch_Y , the set of children of Y , and $Pa\ ch_Y$, the set of parents of children of Y .

For example, consider the two DAGs in Figure.1. and Fig. 3. The factorization of p entailed by the Bayesian Network (S, P) is

$$P(Y, X_1, \dots, X_6) = P(Y | X_1) \cdot P(X_4 | X_2, Y) \cdot P(X_5 | X_3, X_4, Y) \cdot P(X_2 | X_1) \cdot P(X_3 | X_1) \cdot P(X_6 | X_4) \cdot P(X_1), \quad (4)$$

The factorization of the conditional probability $p(Y | X_1, \dots, X_6)$ entailed by the Markov Blanket for Y corresponds to the product of those (local) factors in equation (2) that contain the term Y .

$$P(Y | X_1, \dots, X_6) = C_0 \cdot P(Y | X_1) \cdot P(X_4 | X_2, Y) \cdot P(X_5 | X_3, X_4, Y), \quad (5)$$

C. TREE AUGMENTED NAÏVE BAYES.

Tree Augmented Naive Bayes: Learning a Naive Bayes model, learn a naive Bayesian network based on a given training data set. The structure of the naive Bayes network is given as follows:

Estimate the parameters for the conditional probability distributions in the network using MLE on the training data. Based on the constructed naive Bayesian network you can classify samples by applying Bayes rule to compute conditional class probabilities

$P(C_j | A_1; A_2; A_3; A_4; A_5)$, and predicting the label with the highest probability. Learning a Tree Augmented Naive Bayes (TAN) model. Tree augmented naïve Bayes models are formed by adding directional edges between attributes. After removing the class variable, the attributes should form a tree structure (no V-structures). See Fig.2. as an example. Use the following procedure to learn the tree augmented naïve Bayes model for the training data, then draw the structure of the obtained model. The QWS data set for tree augmented Naïve Bayesian depicted in Fig.8. and Fig.9.

D. SONS AND SPOUSES

Here, the target node is the father of a subset of nodes possibly having other relationships (spouses). This algorithm has the advantage of showing the set of nodes being indirectly linked to the target. The time cost is of the same order as for the augmented naïve Bayes. The Bayesian graph of QWS dataset are depicted in Fig.10. and Fig.11 for Tree Augmented Naïve Bayes networks.

E. MARKOV BLANKET

This algorithm searches for the nodes that belong to the Markov Blanket of the target: fathers, sons and spouses nodes. With this structure's search, which is entirely focused on the target node, we can get the subset of the nodes that are actually relevant in a time frame that is much lower than with the other two architectures Augmented Naïve Bayes and Sons & Spouses. This method is a very good tool for analyzing one variable

markov blanket. The Markov Blanket QWS data set are depicted in Fig.12 and Fig.13.

F. AUGMENTED MARKOV BLANKET

This involves the Markov Blanket learning with the addition of an unsupervised search for the probabilistic relations between each of the variables belonging to the Markov Blanket. This additional search entails a supplementary cost but in exchange there is a gain in precision with respect to the simple version. The Augmented Markov Blanket QWS data set are depicted in Fig.14. and Fig.15.

G. MINIMAL AUGMENTED MARKOV BLANKET

Minimal Augmented Markov Blanket Learning, the selection of the variables that is realized with the Markov Blanket learning algorithm is based on a heuristic search. The set of the selected node scan then be non minimal, especially when there are various influence paths between the nodes and the target. In that case, the target analysis result takes into account too much nodes. By applying an unsupervised learning algorithm on the set of the selected nodes, the Min. Augmented Market Blanket learning allows reducing this set of nodes, and it results then in a more accurate target analysis. The QWS data set for Minimal Augmented Markov Blanket are depicted in Fig.16. and Fig.17.

H. SEMI SUPERVISED LEARNING

Supervised learning is the machine learning task of inferring a function from supervised (labeled) training data. A supervised learning algorithm analyzes the training data and produces an inferred function, which is called a classifier (if the output is discrete, see classification) or a regression function (if the output is continuous, see regression). The inferred function should predict the correct output value for any valid input object.

Semi-supervised learning: unsupervised learning algorithm that searches the relationships between the nodes that belong to a predefined distance of the target. This distance is computed by using the Markov Blanket learning algorithm. The semi-supervised learning algorithm allows learning a network fragment centered on the target variable [11]. This algorithm is very useful for tasks that involve a lot of nodes, as for example micro-arrays analysis (thousand of genes), and for prediction tasks where the Markov Blanket nodes have missing values, as these nodes do not allow to separate the target node from the other nodes. In this setting, the desired output values are provided only for a subset of the training data. The remaining data is unlabeled. The graph of QWS dataset for semi-supervised learning is depicted in Fig.18 and Fig. 19.

IV. RESULTS AND DISCUSSIONS

We employed Naïve Bayes, Augmented Naïve Bayes, Tree Augmented Naïve Bayes, Sons & Spouses, Markov Blanket, Augmented Markov Blanket, Minimal Augmented Markov Blanket and Semi Supervised techniques to classify web services. We note that the accuracy of Augmented Naïve Bayes classifier is 80.72%, followed by Tree Augmented Naïve Bayes, Sons & Spouse techniques presented in Table 1. In this context, we employed trained network data to find the importance of different attributes in web services. We observe that the average accuracy of Markov Blanket, Augmented Markov Blanket, and Minimal Augmented Markov Blanket Classifier accuracy is 72.73%. Minimum accuracy by semi supervised learning is 33.06%.

V. CONCLUSION

We presented Naïve Bayes, Naïve Bayes, Augmented Naïve Bayes, Tree Augmented Naïve Bayes, Sons & Spouses, Markov Blanket, Augmented Markov Blanket, Minimal Augmented Markov Blanket and Semi Supervised techniques to classify web services based on accuracy values. We observed that Augmented, Tree Naïve Bays, Sons & Spouses approach predicted better accuracy of 80.72% than other techniques.

ACKNOWLEDGEMENT

We are grateful to Dr. R. K. Mohanty for providing us the dataset related to the web services classification.

Table 1: Quality of web service data analysis

Classifier Technique	Accuracy in Percentage
Naïve Bayes	79.89
Augmented Naïve Bayes	80.72
Tree Augmented Naïve Bayes	80.72
Sons & Spouses	80.72
Markov Blanket	72.73
Augmented Markov Blanket	72.73
Minimal Augmented Markov Blanket	72.73
Semi Supervised Learning	33.06

TABLE 2: ATTRIBUTES OF QWS DATASET

ID	Attribute Name	Description	Units
G ₁	Response Time	Time taken to send a request and receive a response	ms
G ₂	Availability	Number of successful invocations/total invocations	%
G ₃	Throughput	Total number of invocations for a given period of time	Invokes/second
G ₄	Successability	Number of Response/Number of request messages	%
G ₅	Reliability	Ratio of the number of error messages to total messages	%
G ₆	Compliance	The extent to which a WSDL document follows WSDL Documentation	%
G ₇	Best Practices	The extent to which a web service follows	%
G ₈	Latency	Time taken for the server to process a given request	ms
G ₉	Documentation	Measure a documentation (i.e. description tags) in WSDL	%

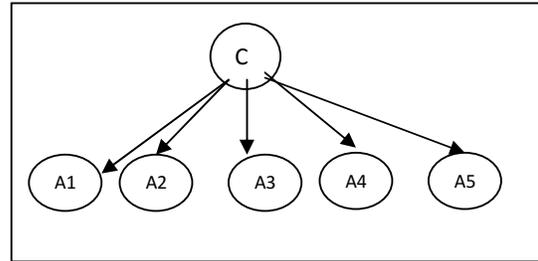


Fig.1. Naive Bayes network

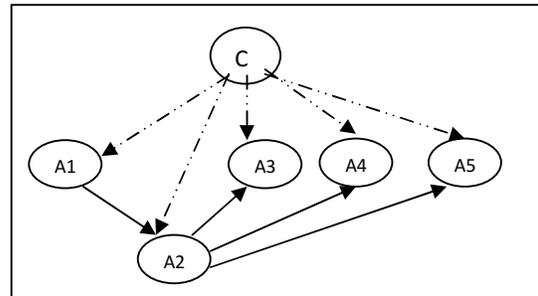


Fig.2. An example of a tree augmented naive Bayes network.

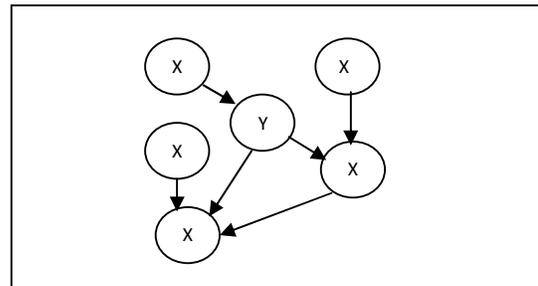


Fig.3. The Bayesian Network (S, P)

Table.3. QOS MODEL OF WEB SERVICES [4]

QOS Factor	Internal attributes (Metrics)	External Attributes (Metrics)
Reliability	Correctness (accuracy and precision)	Availability and consistency
Performance	Efficiency (Time and Space Complexity)	Load management (Throughput, waiting and response time security)
Integrity	-	Security
Usability	Input and output attributes	-

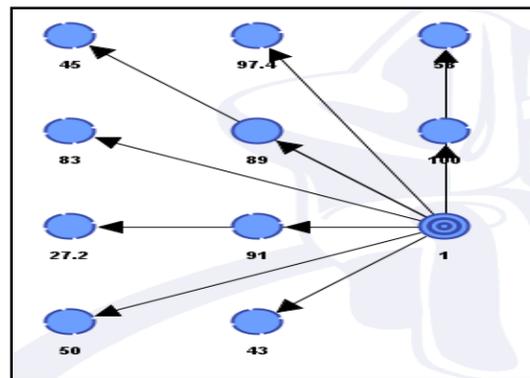


Fig. 4 QWS dataset for Naïve Bayes Bayesian network

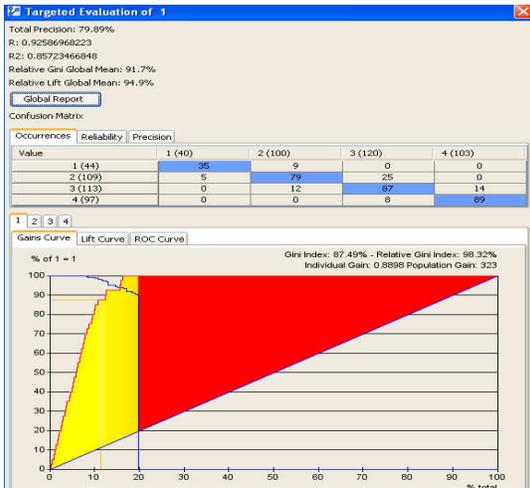


Fig.5. QWS data set Graph for Naïve Bayesian Network Accuracy is 79.89%

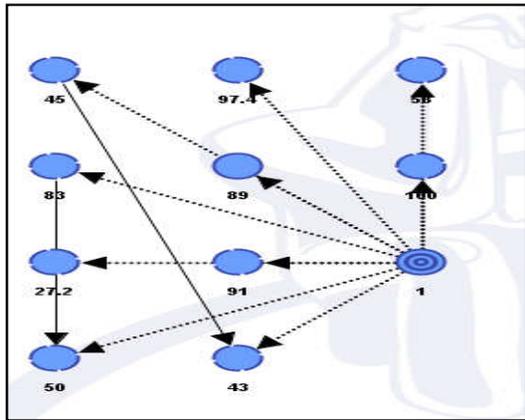


Fig.6. QWS data Augmented Naïve Bayes

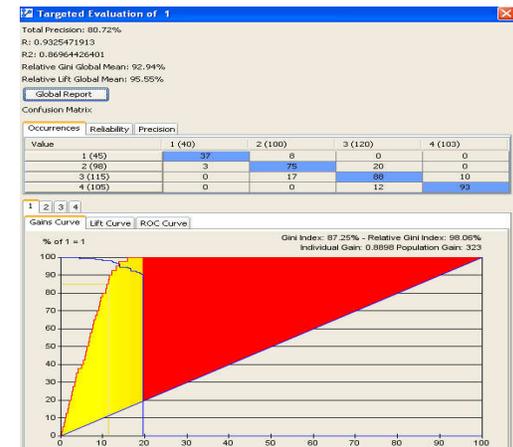


Fig.7. QWS data Graph Augmented Naïve Bayes Accuracy is 80.72%

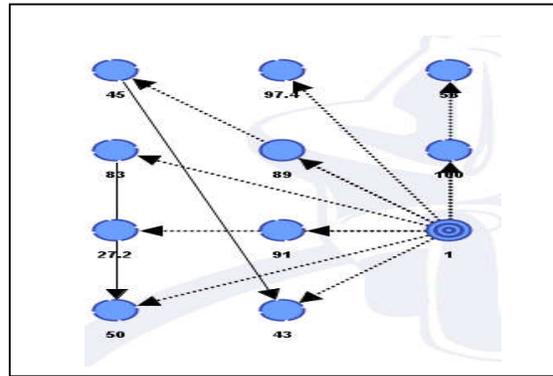


Fig. 8. QWS data Tree Augmented Naïve Bayesian

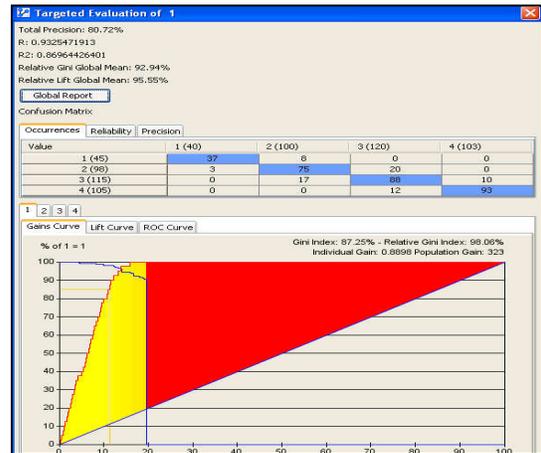


Fig.9. QWS data Graph Tree Augmented Naïve Bayesian accuracy is 80.72%

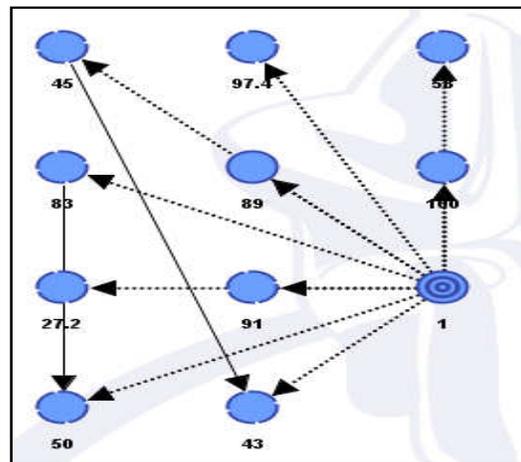


Fig.10. QWS data Sons and Spouses

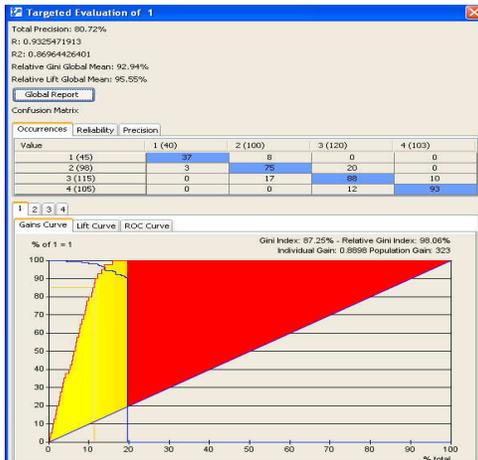


Fig.11. QWS data Graph Sons and Spouses accuracy is 80.72%

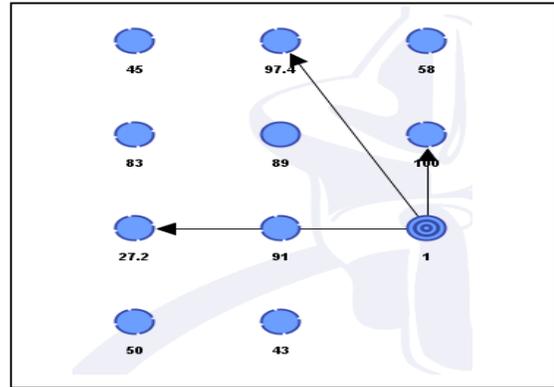


Fig.14. QWS data Augmented Markov Blanket

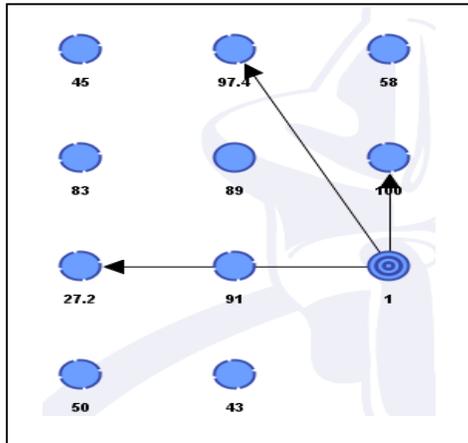


Fig12. QWS data Markov Blanket

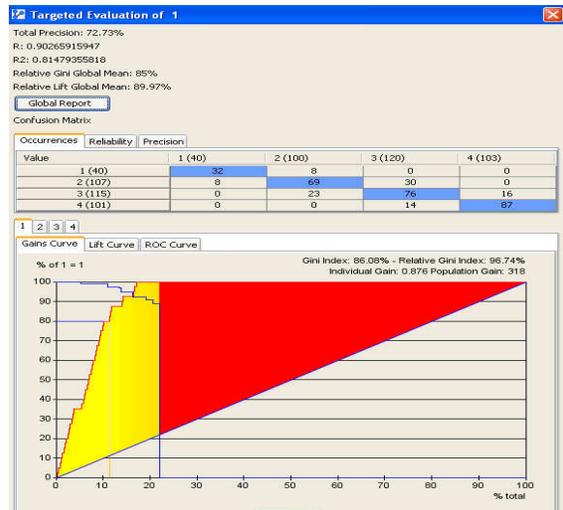


Fig.15. QWS data Graph Augmented Markov Blanket Accuracy is 72.73%

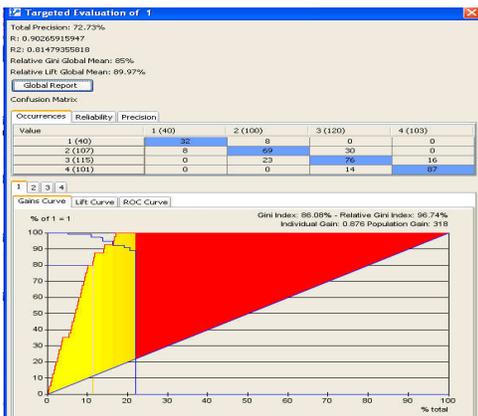


Fig.13. QWS data Graph Markov Blanket accuracy is 72.73%

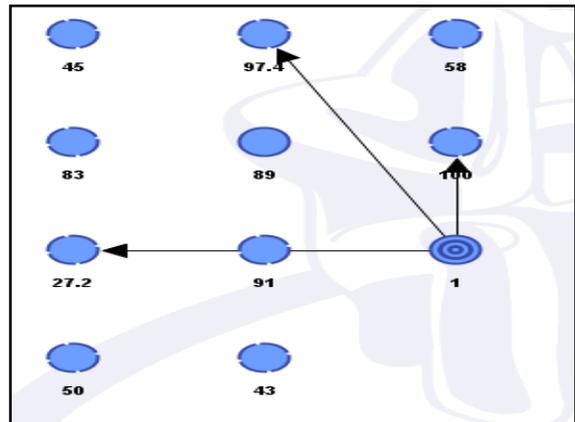


Fig16. QWS data Minimum Augmented Markov Blanket

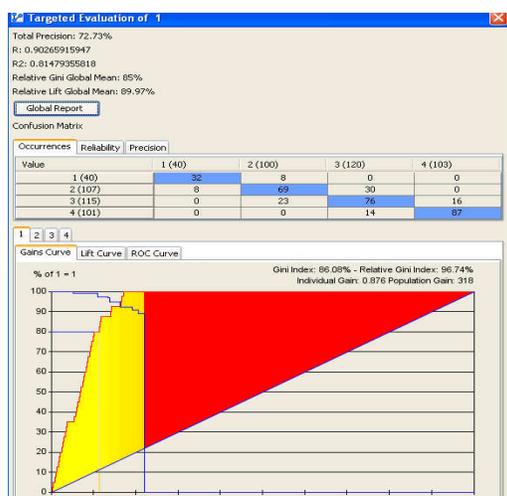


Fig.17. QWS data Graph Minimum Augmented Markov Blanket accuracy is 72.73%

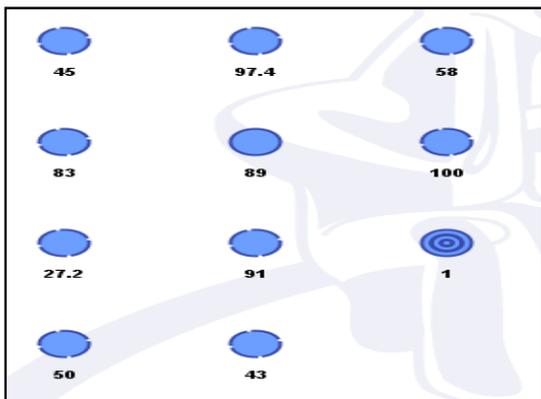


Fig.18. QWS data Semi supervised Learning

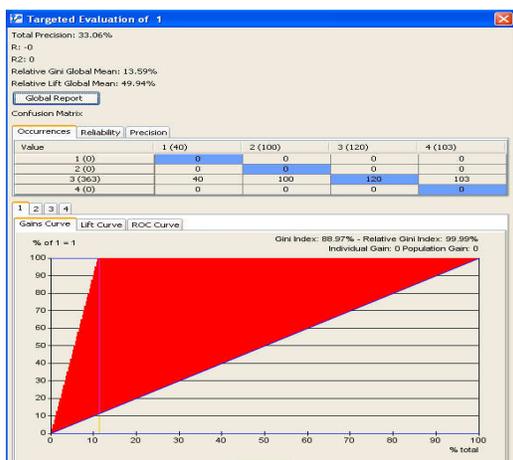


Fig.19. QWS data Semi supervised Learning accuracy is 33.03%

REFERENCES

- [1] E. AL-Masri, and Q. H. Mahmoud, "Investing Web services on the world wide web," The 17th International Conference on World wide web, Beijing, pp. 795-804, 2008.
- [2] E. AL-Masri, and Q. H. Mahmoud, "QoS based discovery and ranking of web services," IEEE 16th International Conference on Computer Communications and Networks (ICCCN), pp. 529 – 534, 2007.
- [3] O. Ardaiz, F. Freitag, and L. Navarro, "Improving service time of web clients using server redirection," ACM SIGMETRICS Performances Evaluation Review, 29(2), 2001.
- [4] S. Araban, and L. Sterling, "Measuring quality of service for contract aware web-services," In First Australian workshop on engineering service-oriented systems (AWESOS 2004). Melbourne, Australia, 2004.
- [5] G. V. Bochmann, B. Kerherve, H. Lutffiyya, M. M. Salem, and H. Ye, "Introducing QoS to electronic commerce applications," Berlin, Heidelberg: Springer- Verlag. pp. 138–147, 2001.
- [6] P. Cheeseman, and J. Stutz, "Bayesian classification (AutoClass): Theory and results". In *Advances in knowledge discovery and data mining*, 153–180. 1996, Menlo Park, CA: AAAI Press.
- [7] D. M. Chickering, Learning equivalence classes of bayesian-network structures. *Journal of Machine Learning Research*, 3:507–554, 2002
- [8] M. Conti, E. Gregori, and F. Panzieri "Load distribution among replicated web services: A QoS based approach," In Second workshop on internet server performance, Atlanta (GA): ACM Press, 1999.
- [9] P. Domingos and M. Pazzani, "On the optimality of the simple Bayesian classifier under zero-one loss", *Machine Learning*, 29, 103–130, 1997.
- [10] S. Kalepu, S. Krishnaswamy, and S. W. Loke, "Verity: A QoS metric for selecting web services and providers," In Proceedings of the 4th International Conference on web mining system engineering workshops (WISEW'03), pp.131-139, 2004..

- [11] Semi-supervised or Semi-unsupervised, Hal Daumé III School of Computing University of Utah, Salt Lake City.
- [12] A. Mani, and A. Nagarajan, "Understanding quality of services for web services," Available from: <http://www-106.ibm.com/developerworks/library/ws-quality>. (2002)
- [13] D. Margaritis, "Learning bayesian network model structure", 2003.
- [14] I. Tsamardinos, C. Aliferis, and A. Statnikov, "Large-scale feature selection using Markov blanket induction for the prediction of protein-drug binding", 2002.
- [15] I. Tsamardinos, C. Aliferis, and A. Statnikov, "Algorithms for large-scale local causal discovery in the presence of small sample or large causal neighborhood". Technical Report DSL-02-08, 2002 Vanderbilt University.
- [16] R. K. Mohanty, V. Ravi and M. R. Patra: "Web Services Classification using intelligent Techniques," Expert Systems with Applications, vol. 37, pp.5484-5490, 2010.
- [17]. Geiger, Dan; Verma, Thomas and Pearl, Judea (1990). "Identifying independence in Bayesian Networks". Networks 20: 507–534. Retrieved 6 October 2011.
- [18] QWS datasets: (<http://www.uoguelph.ca/~qmahmoud/qws/index.html>).
- [19] S. Ran, "A model for web services discovery with QoS," ACM SIGecom Exchange, vol.4, No.1, 2003.
- [20] A. Tsalgatidou, and T. Pilioura, "An overview of standards and related technology in web services," Distributed and Parallel Database, vol.12, pp. 135-162, 2002.
- [21] S. Vinoski, "Service discovery 101," IEEE Internet Computing, vol.7, No.1, pp.69–71, 2003
- [22] A.Zeng, B. Benatallah, M. Dumas, J. Z. Kalagnanam, and Q. Sheng, "Web engineering: Quality driven web service composition," In Proceedings of the 12th international conference on World Wide Web pp. 411–421, Budapest, Hungary: ACM Press 2003
- [23] Mironela, P. Fac. of Sci. & Technol. Inf., TITU MAIORESCU Univ., Bucharest, Romania pp 337-339, Kota Kinabalu, IEEE, 2009.

