Entropy Based Feature Selection For Multi-Relational Naïve Bayesian Classifier

Vimalkumar B. Vaghela
Karpagam University

Kalpesh H. Vandra
C.U Shah College of Engineering & Technology

Nilesh K. Modi
S.V Institute of Computer Studies, S.V Campus, Kadi, Gujarat

Entropy Based Feature Selection For Multi-Relational Naïve Bayesian Classifier

Vimalkumar B. Vaghela
Department of Computer Science & Engineering
Karpagam University
INDIA

Kalpesh H. Vandra
C. U. Shah College of Engineering & Technology,
INDIA

Nilesh K. Modi
S. V. Institute of Computer Studies, S. V. Campus, Kadi, Gujarat
INDIA

ABSTRACT

Current industries data's are stored in relation structures. In usual approach to mine these data, we often use to join several relations to form a single relation using foreign key links, which is known as flatten. Flatten may cause troubles such as time consuming, data redundancy and statistical skew on data. Hence, the critical issues arise that how to mine data directly on numerous relations. The solution of the given issue is the approach called multi-relational data mining (MRDM). Other issues are irrelevant or redundant attributes in a relation may not make contribution to classification accuracy. Thus, feature selection is an essential data pre-processing step in multi-relational data mining. By filtering out irrelevant or redundant features from relations for data mining, we improve classification accuracy, achieve good time performance, and improve comprehensibility of the models. We had proposed the entropy based feature selection method for Multi-relational Naïve Bayesian Classifier. We have use method InfoDist and Pearson's Correlation parameters, which will be used to filter out irrelevant and redundant features from the multi-relational database and will enhance classification accuracy. We analyzed our algorithm over PKDD financial dataset and achieved the better accuracy compare to the existing features selection methods.

INTRODUCTION

The term data mining refers to the extraction of valuable knowledge from large amounts of data. Data mining is the process of discovering knowledge from data. With the massive quantity of data stored in repositories, it is progressively more significant to develop powerful analysis and decision making tool for the extraction of interesting knowledge. The task of classification is concerned with predicting the value of one field from the values of other field. The target field is called the class. The other fields are called attributes. Propositional machine learning algorithms assume the input data is represented in a simple attribute-value format. Most existing data mining algorithms (including algorithms for classification, clustering, association analysis, outlier detection, etc.) work on single tables. For example, a typical classification algorithm (e.g., C4.5 or SVM) works on a table containing many tuples, each of which has a class label, and a

value on each attribute in the table. In recent years, there has been growing interest in multi-relational classification research and application, which address the difficulties in dealing with large relation search space, complex relationships between relations, and a daunting number of attributes involved. Most structured data is stored in relational databases, which is stored in multiple relations by their characters. Conventionally, many classification approaches can only be applied to a single relation. When performing these approaches on multi-relational data, it often requires transferring data into a single table by flattening and feature construction, which is known as Propositionalization. However, many of these methods are heuristic, so flatten may cause some problems such as time consuming and statistical skew on data. Multi-relational data mining (MRDM) has been successfully applied in a variety of areas, such as marketing, sales, finance, fraud detection, and natural sciences. Multi-Relational data mining looks for patterns that involve multiple relations in a relational database, its main difference with traditional data mining approaches is that it does not need to transform the data into a single table, it learns from the data in its original form preserving its structure and incorporating such structure into the learning process.

RELATIONAL DATABASES

A *relational database* is a collection of tables called *relations*, each of which is assign a unique name. Each relation consists of a set of *attributes* and stores a large set of *tuples*. Every tuple in a relational table represents an object which is used to identifying by a unique *key to* describe by a set of attribute values. Often one uses a semantic model to represent relational databases, allowing one to describe and design the database without having to pay attention to the physical database. Such a model is often referred to as a *database scheme*. One of the most common models is the *Entity-Relationship* (ER) model (Figure 1).

A relational database typically consists of several tables (relations) and not just one table. A schema for a relational databases describe a set of entities DB = {E1, E2, …, En}, and set of relationships between entities. Each row in a relation is a tuple. Each relation has at least one primary key attributes. The other attributes are either descriptive attributes or foreign key attributes. Foreign key attributes link to primary key attribute of other relations. A relational database contains multiple interconnected relations, each of which represents a certain kind of objects or a type of relationships. A relational database consists of a set of named tables, often referred to as *relations* that individually behave as the single table that is the subject of Propositional Data Mining. Data structures more complex than a single record are implemented by relating pairs of tables through so-called *foreign key relations*. Such a relation specifies how certain columns in one table can be used to look up information in corresponding columns in the other table, thus relating sets of records in the two tables.

Figure1. The schema of a financial database (from PKDD CUP 1999).
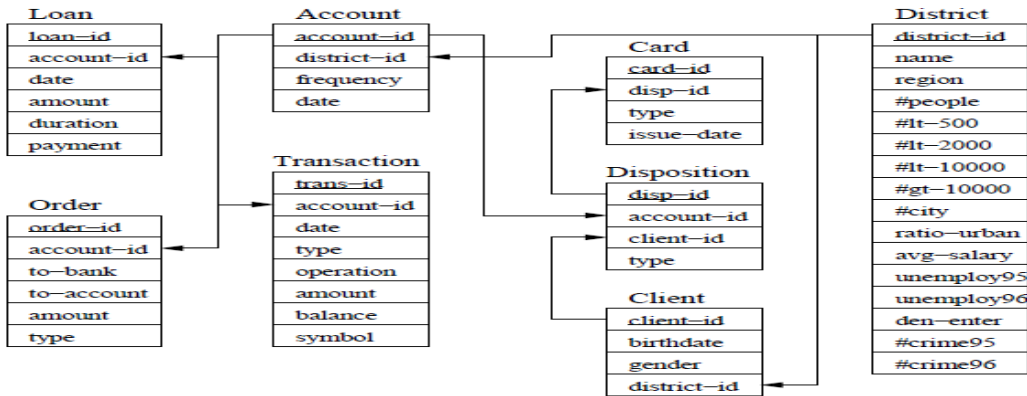

SEMENTIC RELATIONSHIP GRAPH

For a classification task in a multi-relational database, there is usually one table containing the class label attribute. We call this table as *target table*, and call the class label attribute as *target attribute*. Apart from the target table, there are usually many other tables linked to the target table directly or indirectly through arbitrarily long chains of joins. In order to represent this kind of relationship between tables, we use a graph, which is called a semantic relationship graph. Semantic Relationship Graph is kind of similar to ER diagram, which usually can be automatically generated from those common commercial database systems.

Definition: (**Semantic Relationship Graph**) Semantic Relationship Graph is a directed acyclic graph SRG (*V*, *E*, *W*), where *V* is a set of vertices, each of which corresponding to a table in the database. *E* is a set of directed edges, and an edge (*v*, *w*) means table *w* can be linked to table *v* by directly joining these two tables. W is a set of attributes, each of which links two tables. We call this kind of attribute link attribute.

Each edge of the semantic relationship graph represents one of the following two relationships between tables *v* and *w*:

(1) *Primary-key to foreign-key relationship*, indicating that table *w* contains foreign-key referring to primary-key in table *v*.

(2) *Foreign-key to primary-key relationship*, indicating that table *v* contains foreign-key referring to primary-key in table *w*.

The reason we define a directed graph instead of undirected graph is that we need to start from the target table and link other tables with the target table step by step. We can also relax the constraints of semantic relationship graph by allowing the existence of cycle. If so, in order to avoid the iteration doing too many times, we can also set a parameter to control the iteration times. In the following sections, we only regard SRG as an acyclic graph. SRG facilitates the process of virtually joining the relations and acts just like road maps for the entire algorithm. One example of SRG for a financial database from PKDD CUP99 is given in Figure 2.

Figure 2. Semantic relationship graph for the financial database from PKDD CUP99.
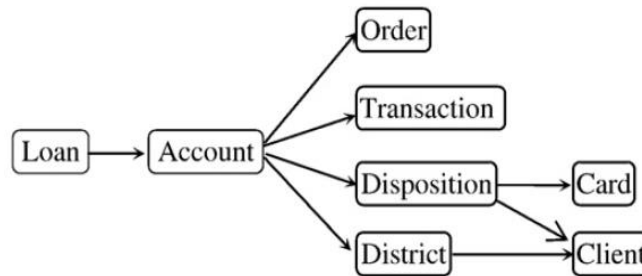
TUPLE ID PROPAGATION

Tuple ID propagation is a method for virtually joining non-target relations with the target relation. It is flexible and efficient method and it avoids the high cost of physical join. Suppose the primary key of the target relation is an attribute of integers, which represent the IDs of the target tuples. We use the ID of each target tuple to represent that tuple. This process takes only small amount of time and space compared to the physical joins used by the existing classifiers and it will boost up the effectiveness of the multi-relational classification techniques. Tuple ID propagation approach reveal to search in the relational database and which is observed that less costly than physical joins in both time and space.

Definition: ID propagation. Suppose we have relation R_1 and R_2, which can be joined by attributes $R_1.A$ and $R_2.A$. Each tuple in R_1 is associated with some IDs in the target relation. For each tuple t in R_2, we set t's IDs to be the union of $\{u's\ ID \mid u \in R_1, u.A = t.A\}$.

FEATURE SELECTION PROCESS

With the creation of huge databases and the consequent requirements for good machine learning techniques, new problem arise and novel approaches to feature selection are demand. Feature selection plays an important role in classification. Feature selection is an important preprocessing step to machine learning. It selects an effective subset from the original features according to a certain criterion so that it can improve the performance of later data processing, such as classification and clustering. In real-world applications, there are many irrelevant and redundant attributes in relations of relational database, in which are little contribution to classification accuracy. Hence, feature selection is an essential data processing step in multi-relational data mining. By applying feature selection techniques, we can improve classification accuracy, achieve good time performance, and enhance comprehensibility of the models. Feature selection reduces the number of features, removes irrelevant, redundant, or noisy data, and brings the immediate effects for applications: speeding up a data mining algorithm, improving mining performance such as predictive accuracy and result comprehensibility.In fact, feature selection techniques have been widely employed in a variety of applications, such as genomic analysis, information retrieval, and text categorization.

Feature selection is a process that selects a subset of original features. The optimality of a feature subset is measured by an evaluation criterion. As the dimensionality of a domain expands, the

number of features *N* increases. Finding an optimal feature subset is usually intractable and many problems related to feature selection have been shown to be **NP**-hard.

Feature selection algorithms designed with different evaluation criteria broadly fall into two categories: the *filter* model and the *wrapper* model.

Filter Model: The filter model relies on general characteristics of the data to evaluate and select feature subsets without involving any mining algorithm.

Wrapper Model: The wrapper model requires one predetermined mining algorithm and uses its performance as the evaluation criterion. It searches for features better suited to the mining algorithm aiming to improve mining performance, but it also tends to be more computationally expensive than the filter model.

Feature selection is defined by many authors by looking at it from various angles. But as expected, many of those are similar in intuition and/or content. The following lists those that are conceptually different and cover a range of definitions.

 (1) **Find the minimally sized feature subset that is necessary and sufficient to the target concept.**

 (2) **Select a subset of *M* features from a set of *N* features, *M <N*, such that the value of a criterion function is optimized over all subsets of size *M*.**

 (3) **The aim of feature selection is to choose a subset of features for improving prediction accuracy or decreasing the size of the structure without significantly decreasing prediction accuracy of the classifier built using only the selected features.**

 (4) **The goal of feature selection is to select a small subset such that the resulting class distribution, given only the values for the selected features, is as close as possible to the original class distribution given all feature values.**

Feature selection attempts to select the minimally sized subset of features according to the following criteria. The criteria can be:

 (1) **The classification accuracy does not significantly decrease; and**

 (2) **The resulting class distribution, given only the values for the selected features, is as close as possible to the original class distribution, given all features.**
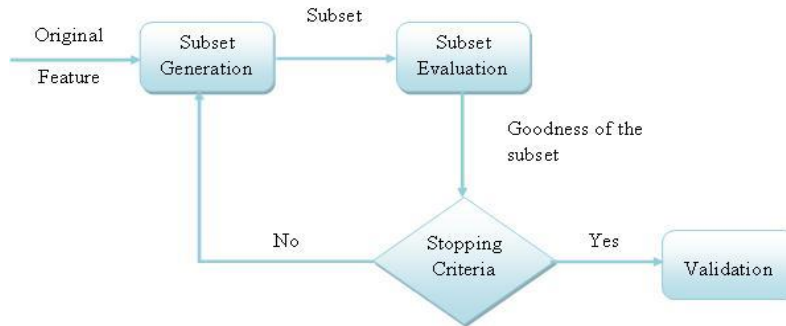
Figure 3. Feature selection process.

Table 1. Analysis of feature selection method.

Feature selection method	Author Name and year of publication	Description	Drawback
Feature and Relation Selection (FARS)	Hu, Liu, He & Du (2008)	Evaluated by using table symmetrical uncertainty (TSU) which is symmetrical uncertainty (SU) value between relation and class. (over multi-relational dataset)	Discrete values are not handled
Feature Selection using InfoDist	Sha, Qiu, & Zhou (2007)	Evaluated by InfoDist which based on information theory.	Discrete values are not handled
Wilk's Lambda criterion method	Ouardighi, Akadi, & Aboutajdine (2007)	Evaluated by a statistical value used in discriminant analysis.	Insufficient to improve the classifier performances only by relevance criterion.
MR^2 feature selection method	Unler, Murat, & Chinnam (2007)	This method uses InfoDist and Pearson Correlation to calculate the relevant features (over multi-relational dataset)	Cutoff value is hard to decide
Fast Correlation Based Filter (FCBF)	Yu & Liu(2003)	Evaluated by information gain combines optimal subset and feature relevance weight method.	Discrete values are not handled

OUR PROPOSED ENTROPY BASED FEATURE SELECTION ALGORITHM

In feature selection, first we use InfoDist to evaluate the distance between feature and class label. If a feature x_i has less distance $d(x_i,C)$ with the class label C, we thought it is more relevant to the class label. We define a cutoff distance based on standard deviation. These features with distance larger than mean distance plus cutoff value are regarded as irrelevant and are removed. In experiment, we observe the effect to classification accuracy with respective

to different cutoff values.

Second, we use Pearson's correlation to evaluate the correlation between features. Two features with high correlation are redundant each other. We select the minimum redundancy features according to the correlation between the features. We select three different feature sets according to InfoDist distances and Pearson's correlations for our experiments. The three selection methods are described as follows:

(1) Maximum relevant feature set (MaxRel): We use cutoff value to discard irrelevant features from the sorted InfoDist feature list. These features which have the smallest Pearson's correlation with respective to each feature in the remaining feature list are appended in the selection feature list with duplicates eliminated.

(2) Minimum redundancy feature set (MinRed): We discard irrelevant features from the sorted InfoDist feature list using cutoff value as in maximum relevant feature set. The feature which has the smallest Pearson's correlation with the listed feature is put immediate following the listed feature included in the selection list. The selected feature lists are primary based on less redundancy between features.

InfoDist calculation

InfoDist is based on information theory. The main concept of information theory is entropy, which measures the expected uncertainty or the amount of information provided by a certain event. The entropy of a random variable X is defined as follows:

$$H(X) = -\sum_{x} P(X = x) \log P(X = x)$$

(1)

, where P(X=x) is the prior probability of x.

Entropy *H(Y /X)* of a random variable *Y* given *X* is defined as follows:

$$H(Y \mid X) = -\sum_{x,y} P(x,y) \log P(y \mid x)$$

(2)

Mutual information is a measure of how much the probability distribution for a random variable changes when the value of another random variable is known. The mutual information between two random variables *X* and *Y* is defined in the following:

$$I(X;Y) = \sum_{x,y} P(x,y) \log \frac{P(x,y)}{P(x)p(y)}$$
$$= H(X) - H(X \mid Y)$$
$$= H(X) + H(Y) - H(X,Y)$$

(3)

InfoDist adopt the conditional entropy to measure the relevance between a feature and the class label. The distance, *d(X, C)*, of a feature and the class label is evaluated by

$$d(X, C) = H(X|C) + H(C|X)$$

(4)

Pearson's correlation calculation

We adopt Pearson's correlation to measure the redundancy between features. If variables *X* and *Y* are continuous, the correlation is calculated by formula defined in the following:

$$r_{XY} = \frac{\sum xy}{n\sigma_x \sigma_y} \tag{5}$$

If X is a discrete feature with k values, and Y is a continuous feature. The correlation is calculated by formula defined in the following:

$$r_{XY} = \sum_{i=1}^{k} P(X = x_i) r_{X_{b_i} Y} \tag{6}$$

W here X_{bi} is a binary feature that takes value 1 when X has value x_i; otherwise, 0. If variables X and Y are both discrete, the correlation is calculated by formula defined in the following:

$$r_{XY} = \sum_{i=1}^{k} \sum_{j=1}^{t} P(X = x_i, Y = y_j) r_{X_{b_i} Y} \tag{7}$$

There are two types of methods to deal with multiple relations by Naive Bayes. One is to convert multiple relations into a single relation; and the other is to deal with each relation directly. We prefer the latter method because the advantages of MRDM.

Now, we need to extend the above formula of classification to deal with multiple relations. We assume that t is the target relation, and s is non-target relation that can be joined with the relation t. Assume the relation t has n attributes and the relation s has m attributes. For a tuple x in the relation t: $x=(x_1, x_2, \ldots, x_n)$, there are p tuples in the relation s which can be joined with the tuple x. These p tuples are $(yk_1, yk_2, \ldots, yk_p)$, where each tuple yk_i is represented by r attributes: $yk_i=(yk_{i1}, yk_{i2}, \ldots, yk_{ir})$. Then, the class label of the tuple x can be predicted according to the following formula:

$$
\begin{aligned}
C_{map} &= \arg\ \max_{c_j \in C}\ P(C_j|X) \\
&= \arg\ \max_{c_j \in C}\ P(C_j)\,P(X|C_j) \\
&= \arg\ \max_{c_j \in C}\ P(C_j)\,P(x_1, x_2, \ldots, x_n, y_{k_{1_1}}, \ldots, y_{k_{1_r}}, \ldots, y_{k_{p_1}}, \ldots, y_{k_{p_r}}|C_j) \\
&= \arg\ \max_{c_j \in C}\ P(C_j)\prod_{i=1}^{n} P(x_i|C_j)\prod_{q=k_1}^{k_p}\ \prod_{t=1}^{r} P(y_{q_t}|C_j)
\end{aligned}
\tag{8}
$$

In order to make the above expression operational, we should find a feasible way to specify the probability distribution for each attribute and compute the associated conditional probabilities. In our algorithm, we adopt the tuple ID propagation method to virtually join relations along each path and collect the required information for computation. To guide the search within the relation space, a Semantic Relationship Graph is also constructed to represent and summarize the relationships between various relations in the database.

By virtual join, tuple ids are propagated from the target relation to the non-target relations. The semantic relationships between relations remain unchanged. The storage space is cheaper than physical join. In feature selection, first we use InfoDist to evaluate the distance between feature and class label. If a feature x_i has less distance $d(x_i, C)$ with the class label C, we consider it is more relevant to the class label. We define a cutoff distance based on standard deviation. These features with distance larger than mean distance plus cutoff value are regarded as irrelevant and are removed. Second, we use Pearson's correlation, to evaluate the correlation between features. Two features with high correlation are redundant to each other. We select the minimum redundancy features according to the correlation between the features. Thus, feature which have the smallest Pearson's correlation with respective to each feature in the remaining feature list are appended in the selection feature list with duplicates eliminated. We select three different feature sets according to InfoDist distances and Pearson's correlations for our experiments. As a result, the best candidate features are produced to improve classification accuracy. Then these selected features are subjected for Multi-relational Naïve Bayes classification.

ENTROPY BASED FEATURE SELECTION ALGORITHM

The symbols and functions referred to the algorithm is as follows:

(1)   D, Relational database
(2)   M, Target table
(3)   Ri (1, 2 …n) Association tables
(4)   FeatureCount_Ri, Count of number of features in a table R_i
(5)   Function CreateRelationGraph (G) is generating a relation diagram.
(6)   Function Propagate (R_i, M) means to transmit the class label to The table R_i from the target table M;
(7)   Function InfoDist (A_j, C) means to calculate the InfoDist of the feature Aj w.r.t class label C, in a respective table.
(8)   Function PerCorr (A_j, C) means to calculate the Pearson's Correlation of the feature Aj w.r.t other features in a table.
(9) A_j, Feature of the table
(10)   C, Class label

The algorithm is as follows:

Input: Target table M, contingency tables R_i (i=1, 2 …n)
Output: S_{Ri}, Set of selected attributes for each table

Method
(1) CreateRelationGraph (G);
(2) Propagate (R_i, M);
(3) For k=0 to 3 do
      (i) Set cutoff value
      (ii) For each table R_i do
            (a) For j=0 to FeatureCount_Ri do
                  InfoDist_value[j] ←InfoDist (A_j, C);

End For

(b) Sort values of InfoDist_value[j] in ascending order;

(c) Discard the InfoDist values which are larger than cutoff value, and select the remaining respective InfoDist value features in set S_{Ri};

(d) For j=0 to FeatureCount_Ri do

   PerCorr_value[j] ←PerCorr (A_j);

End For

(e) Sort values of PerCorr_value in ascending order;

(f) Feature which has the smallest Pearson's correlation v a l u e with respective to each feature in the remaining feature list are appended in the selection feature list S_{Ri}, with duplicates eliminated.

   End For

End For

EXPERIMENTS, RESULTS AND DISCUSSION

For our experimental study we had used the well-known relational database PKDD Financial dataset as describe the Figure 1. Loan table is the target table and other table is consider as a non-target table for our work.

Table 2. PKDD Financial dataset description.

Relation	No. of Objects	Description
Account	4500	describes static characteristics of an account
Client	5369	describes characteristics of a client
Disposition	5369	relates together a client with an account
Order	6471	describes characteristics of a payment order
Transaction	1056320	describes one transaction on an account
Loan	682	describes a loan granted for a given account
Card	892	describes a credit card issued to an account
District	77	describes demographic characteristics of a district

Table 3 describes the performance comparisions of the existing feature selection with our proposed alogortihm Entropy based feature selection classifier. We had used the accuracy as our comparision parameter and we achieve the better accuracy compared to the existing methods.

Table 3. Performance comparisions of entropy based feature selection classifier with existing feature selection based classifier.

Data Set	Classifier	Accuracy (%)
PKDD Financial dataset	FARS	83
	Multi-relational Naïve Bayes Classifier with MR^2 feature selection and wrapper method	89
	Entropy based feature selection classifier	91

Table 4 describes the performance comparisions of the existing multi-relational classifiers (without feature selection approach). Still we are able to achieve the better performance compare to the existing methods on PKDD financial datasets.

Table 4. Performance comparisions of entropy based feature selection classifier with existing multi-relational classifier.

Data Set	Classifier	Accuracy (%)
PKDD Financial dataset	FOIL	71.5
	TILDE	81.3
	Graph-NB	85.25
	CrossMine	89.8
	Entropy based feature selection classifier	91

For calculating the classification accuracy we develop four join paths according to dataset relationships. Each path starts from Loan target table and follows Account table to branch paths. For easy reference, we call joins paths as OrderPath, TransPath, CardPath and ClientPath.

OrderPath->Loan, Account, Order
TransPath->Loan Account, Transaction
CardPath->Loan, Account, Disposition, Card
ClientPath-> Loan, Account, Disposition, Client

Figure 7. Performance comparisons of entropy based feature selection classifier with existing feature selection based classifier.
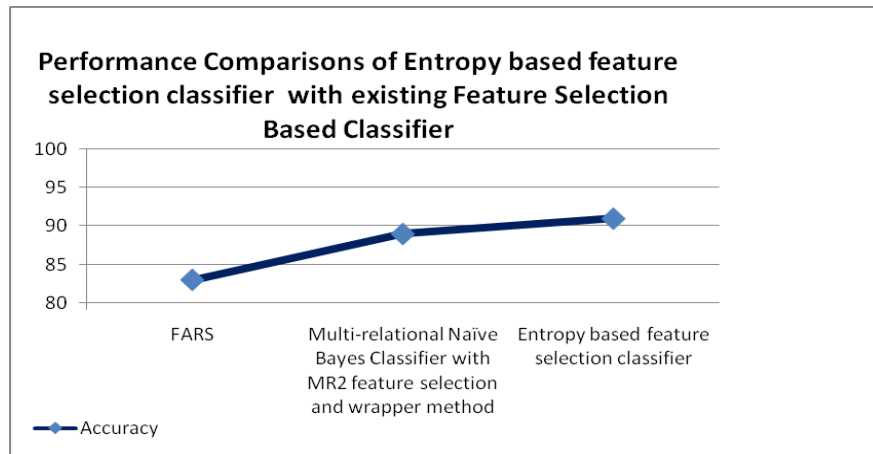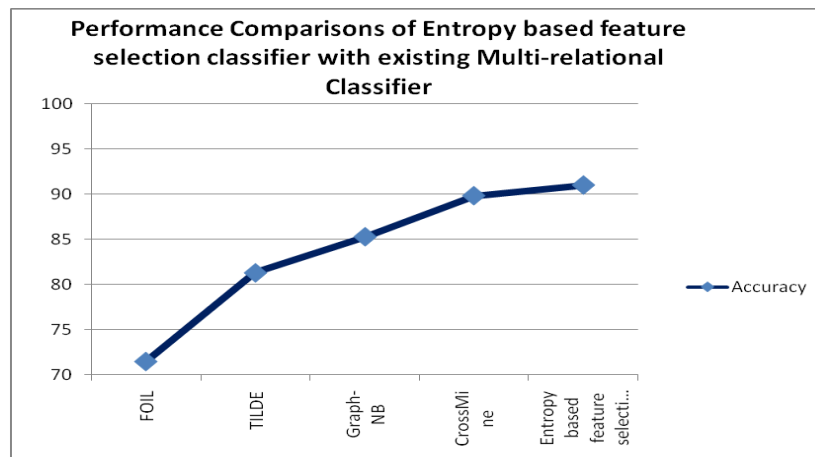

Figure 8. Performance comparisons of entropy based feature selection classifier with existing multi-relational classifier.

CONCLUSION AND FUTURE WORK

We proposed a entropy based feature selection method which having MaxRel feature selection method with Multi-relational Naïve Bayes classifier. Our proposed entropy based feature selection for mmulti-relational naive bayesian classifier improve the classification accuracy and enhance comprehensibility of the models. In pre-processing step, feature selection is done to select relevant features by using InfoDist values and remove redundancy features by using Pearson's correlation. In filter step, we select fewer relevant features in feature pool with respect to cut-off value. The experimental result shows that the our proposed classifier is effective in respect to the classification accuracy. For the future work, we can apply our proposed classifier to the more relational dataset to measure the performance of our classifier.

REFERENCES

Berka, P. (2000). Guide to the financial data set. In A. Siebes, and P. Berka, edited, *The ECML/PKDD 2000 Discovery Challenge*.

Chen, H., Liu, H., Han, J., Yin, X., & He, J. (2009). Exploring optimization of semantic relationship graph for multi-relational Bayesian classification. *Decision Support Systems*, 48(1), 112-121.

Dash, M., & Liu, H. (1997). Feature selection for classification. *International Journal of Intelligent Data Analysis*, 1(3), 131-156.

Ding, C., & Peng, H. (2003). Minimum Redundancy Feature Selection for Gene Expression Data. *Proceeding of IEEE Computer Society Bioinformatics Conference*, California, Stanford, 523-529.

Fayyad, U. & Irani, K. (1993). Multi-Interval Discretization of Continuous-Valued Attributes for Classification Learning. *Proceedings of the 13th International Joint Conference on Artificial Intelligence*, San Francisco, CA: Morgan Kaufmann, 1022-1027.

Gaertner, T., Flach, P., & Kowalczyk, A. (2002). Multi-instance kernels. *Proceedings of 19th International Conf. on Machine Learning,* 179-186.

Hall, M. (2000). Feature Selection for Discrete and Numeric Class Machine Learning. *Proceeding of Seventeenth International conference on Machine Learning*, San Francisco, California, 359-366.

He, J., Liu, Hu, B., & Wang, P. (2010). Selecting Effective Features and Relations For Efficient Multi-Relational Classification. *Computational Intelligence,* 26(3).

Horva, T., Wrobel, S., Bohnebeck, U. (2001). Relational Instance-based Learning with Lists and Terms. *Machine Learning,* 43, 53-80.

Hu, B., Liu, H., He, J., & Du, X. (2008). FARS: A Multi-relational Feature and Relation Selection Approach for Efficient Classification. *Advanced Data Mining and Applications: 4th International Conference, Lecture Notes in Computer Science*, 5139, 73-86.

Lavrac, N., & Flach, P. A. (2001). An Extended Transformation Approach to Inductive Logic Programming. *ACM Transactions on Computational Logic*, 2(4), 458~494.

Muggleton, S. & Feng, C. (1990). Efficient induction of logic programs. *Proceedings of the 1st Conference on Algorithmic Learning Theory*, Tokyo, Japan, 368-381.

Muggleton, S. (1995). Inverse entailment and Progol (1995). *New Generation Computing, Special issue on Inductive Logic Programming*, 13(3), 245-286.

Montanes, E., Quevedo, J. R., & Diaz, I. (2003). A Wrapper Approach with Support Vector Machines for Text Categorization. *Lecture Notes in Computer Science,* 2686, 230-237.

Ouardighi, A., Akadi, A., & Aboutajdine, D. (2007). Feature Selection on Supervised Classification Using Wilk's Lambda Statistic. *Proceeding of 2007 Computational Intelligence and Intelligent Informatics (ISCIII'07),* 51-55.

Pan, C., & Wang, H. (2009) . Multi-relational classification on the basis of the attribute reduction twice. *Communication and Computer,* 6(11), 49-52.

Sha, C., Qiu, X., & Zhou, A. (2008). Feature Selection Based on a New Dependency Measure. *Proceeding of 2008 International Conference on Fuzzy Systems and Knowledge Discovery (FSKD'08)*, 1, 266-270.

Tadeuchi, Y., Oshima, R., Nishida, K., Yamauchi, K., & Omari, T. (2007). Quick Online feature selection method for regression - A feature selection method inspired by human behavior. *Proceeding of the IEEE International Conference on Systems, Man and Cybernetics*, Canada, October 2007.

Vaghela, V. B., Ganatra, A., & Thakkar A.   (1993). Boost a weak learner to a strong learner using ensemble system approach. *Advance Computing Conference, 2009. IACC 2009. IEEE International*. IEEE.

Vaghela, V. B. (2013). MR2 Based Feature Selection for Multi-Relational Naïve Bayesian Classifier. *International Journal of Applied Research in Computer Science and Information Technology,* 2(1).

Vaghela, V. B., Vandra, K. H., & Modi, N. K. (2012). Analysis and Comparative Study of Classifiers for Relational Data Mining. *International Journal of Computer Applications* 55(7), 11-21.

Vimalkumar, B., Vaghela, K. H. V., Modi, N. K. (2012). Multi-Relational Classification Using Inductive Logic Programming. *International Journal of Engineering Research & Technology,* 1(3).

Xu, G. M., Yang, B. R., & Qin, Y. Q. (2008). New multi relational naïve Bayesian classifier. *Systems Engineering and Electronics,* 30(4), 655-655.

Yin, X., Han, J., & Yang. J. (2003). Efficient Multi-relational Classification by Tuple ID Propagation. *Proceedings of the KDD-2003 Workshop on Multi- Relational Data Mining*.