MRE2C: A method for constructing multi relational ensemble classifier based on two-step combining classifiers

Raziyeh Zall, Mohammad Reza Keyvanpour

***Abstract*— In this work, we introduce MRE2C method for classifying multi relational data. Multi-relational data are stored on relational databases where they consist of multiple relations that are linked together by entity-relationship links. MRE2C creates multiple different feature subsets of relational database and then applies traditional classifiers as base classifiers. Final by using a proposed two-step combining classifier method, the results of base classifiers are combined. In first step, the proposed method uses local voting to create meta-features and then it learns meta learner to combine predication of base classifiers. Testing has been performed on two databases and six benchmark tasks. We compare our proposed method with other state-of-the-art multi relational classification methods which use different approaches to deal with multi relational setting. We showed that the proposed method achieves promising results in experiments.**

***Index Terms*— multi relational classification, relational database, ensemble learning, meta learning.**

I. Introduction

MOST real world data are stored in relational databases [1]. A relational database consists of multiple interconnected tables that are linked through Entity-relationship links in the relational database. One of tables is a target relation in which one of attributes is a class attribute or target attribute. A specific attribute in each table is used as primary key. Also, each table includes some foreign key attributes which foreign key attributes are primary keys in other tables in relational database. These foreign key attributes join tables together [2]. For dealing with such data, Multi relational classification methods are proposed to find patterns across multiple interconnected relations in relational databases [3]. For example, Fig 1 shows an example library database. A library database which contains six relations. In this relation database, book relation is target table and Barcode is primary key and PublisherID is foreign key and status is an attribute for class value, for concept to be learned, i.e. whether a book is a good or bad for this library (if we want to order books for this library which of books are better choice). In this paper, we use the example database in Fig1 as a reference to define some of the concepts.

Some researchers are shown that learning ensemble of different classifiers from different feature subsets can be improved accuracy rather than any of the individual classifiers [4] [5]. Especially a key for ensemble learning is to construct both accurate and diverse base learners [6]. In multi relational classification, by using different features in relations of a relational database, we can create a set of diverse base classifier. In fact, different subsets of features in relations of a relational database can create diverse base classifiers that combining of classifiers can be better for generating a classifier that is superior to any of the individual classifiers [5]. Some methods in this literature are proposed such as [1] [7] [3] [8] . The proposed method in this paper follows this line of thought.

In this paper, we propose an extension of the ensemble learning to the multi relational setting. We present a method for constructing multi relational ensemble classifier based on two-step combining classifiers called as MRE2C. Our proposed method can be applied on relational databases directly without need to convert relational databases to other representation such as logic and without need to convert multiple relations to single relation. In our method, we create different features that they describe tuples in target table by different feature sets. In this regard, we use tuple id propagation idea which is proposed in [9] for propagating label from target table to background tables based on foreign key paths and we call new background tables as propagated tables.

BOOK					
barcode	title	Number of version	Number of exist book	publisherID	status
B1	Relational Database	5	3	P1	Good
B2	Data structure	3	3	P2	Good
B3	History of 1995	2	2	P3	Bad
B4	Oracle database	5	5	P1	Good
B5	Mysql databases	6	4	P1	؟

MEMBER			
Name	MemberID	Member type	Start period
Sara	M1	Student	2011
Jim	M2	Student	2012
Oliver	M3	Professor	2015

AUTHOR				
Name	AuthorID	Phone number	citation	Number of publication
Sara	A1	882895	2000	500
Tim	A2	776321	1570	352
jack	A3	56789	500	226

Writing	
Barcode	AuthorID
B1	A1
B2	A2
B3	A3
B4	A1
B5	A2

BORROWED			
MemberID	barcode	Borrowed date	Must return data
M1	B1	2015/2/17	2015/2/21
M2	B5	2015/2/17	2015/2/21
M2	B2	2015/2/14	2015/1/19
M2	B1	2015/2/14	2015/1/19
M3	B4	2015/1/15	2015/1/20

PUBLISHER		
PublisherID	Address	Phone number
P1	Tehran	887695
P2	Esfahan	666321
P3	Tabriz	24789

Fig1- a small instance of a library database

Some papers in this field are used aggregation functions after tuple id propagation to summarize information embedded in multiple tuples and squeeze them into one row [3] [7]. But, since using aggregation functions make information lost [1], we do not use aggregation function in creating different feature subsets after tuple id propagation. After tuple id propagation, we apply one of traditional classifier on data of target tables and data of propagated tables. In our method, all traditional classifier can be used as base classifier and we can use advantage of any traditional classifier. Some methods are

Manuscript received April 28, 2016; accepted July 26, 2016
Department of Computer Engineering, Alzahra University, Tehran, Iran.
(Email: Zall.raziyeh@gmail.com, keyvanpour@alzahra.ac.ir).

[Downloaded from mjee.modares.ac.ir on 2025-01-16]

proposed for combining the results of a classifier ensemble such as majority voting [10] and weighted voting [4] and meta learning [11] [12]. In this paper, we propose a two steps combining classifier method for multi relational classification that exploit advantage of both voting and meta-learning in multi relational classification. In first step, we use voting idea to create input for meta-leaner and in second step, we use meta-learning idea to combine base classifier and create final model of ensemble classifier. Meta learning is learning from base classifiers of ensemble classifier for combining results of base classifiers. The output of the base classifiers are inputs for meta-learner [11]. Meta learning are used to learn which base classifiers are reliable. In multi relational data classification, the results of base classifiers, that are learned by created different features, may de different and cannot be guarantee the comparable performances of the base classifiers, so the meta learning method is more suitable in this case [3]. In [3], multi relational classification based on multi view is proposed which it uses meta-leaner for combining trained classifiers of multiple views, But this method uses aggregation function. We compare our proposed method with this method to specify which using aggregation functions in creating multiple feature sets make information lost [1] so accuracy of classifiers reduced. In fact, we proposed a two-step combining classifier method to handle multiplicity of relationships.

The rest of the paper is organized as follows. In Section 2 we briefly review the related works. In Section 3, we introduce the proposed MRE2C method in detail. In Section 4, experimental results on several data sets are presented. Finally, concluding remarks and some discussions of future work will be given in Section 5.

II. Related work

Most real-world data are in the form of relational data which are stored in multiple relations, multi relational classification approaches are proposed to deal with such data. In this regard, two directions exist, one is to convert multiple relations into one relation by flattening and feature construction, called Propositionalization [13]. Most of the Propositionalization methods are heuristic, and the representation change is incomplete [2]. The flatting process of Propositionalization methods uses multiple SQL joins so can be computationally very expensive and the result tables take up more space [8]. Some method are proposed based on Propositionalization such as [14] , [15], [9], [16] [17]. The other direction is to upgrade traditional data mining algorithms which they do not need to flat and directly apply to relational databases [2]. A number of methods, which are considered as first-order upgrades of existing propositional learners, have been proposed in the Inductive Logic Programming (ILP) community [18]. C2D [19] combines rule extraction methods in ILP and Apriori-based specialization operator and uses support and confidence values for pruning the search space find frequent and strong concept definitions. CRIS [20] employes new search space pruning methods and metrics and eliminates the need for mode declarations.The system named Mr-SBC [21] is proposed to upgrade Naïve Bayes classification method to multi-relation setting. Mr-SBC algorithm integrates first-order classification rules with Naïve Bayesian classification so that the computation of probabilities of shared literals can be separated from the computation of probabilities for the remaining literals. Multi relational model tree induction tightly-coupled [22]proposes a model tree learner system, which deals with data stored in several tables of a tightly-coupled relational database. Relational model trees with both splitting nodes and regression nodes are built. BBN [23] generates dependency structure from Semantic relationship graph then it is processed. In this method uses tuple id propagation and findes conditional probability of each. BBN find out class label of unknown sample by using Forward or backward inference for every variable.

Upgrading approaches based on ILP techniques often have poor scalability when dealing with complex database schemas and also have unsatisfactory predictive performance while handling noisy or numerical values in real-world applications [3]. The above problem can be solved using relational classification algorithms that do consider detailed relationships between attributes [8]. In order to relational upgrading approaches are proposed based on relational database theory and use relational data model and primitives of relational database. Relational probabilistic Trees (RPTs) [24] is an approach based on decision tree that using idea of C4.5 and extending conventional probability estimation tree algorithms to work with relational data. SRG-BC [25] upgrades naïve bayesian and integrates relation and feature selection. The relation virtually joins together by tuple id propagation and provides necessary information for pruning. EDIT [26] extendes the relational algebra representation language for upgrading KNN to realtional data classification. It defines a new type of foreign keys associations, in addition to attributes of type set, gives rise to a new attribute of type list and extendes the well-known alignment-based edit distance measure on lists. In [26],the Bayesian network structure from functional dependencies implied are generatede, in which it constructes the corresponding Markov network.

Some approaches are proposed for upgrading ensemble learning to multi relational data classification. The MRC method is proposed based on multi view learning which used multi relational classification. It firstly explores the relational domain to partition its features space into multiple subsets. Subsequently, these subsets are used to construct multiple uncorrelated views, based on a heuristic view validation method, on the target concept [27]. In the information aggregation stage base aggregation functions of SQL is used [28].

The method CoTReC [7] combines transductive inference and co-training to mine the relational database. This method is based on multi view learning and exploits multi-views extracted from a relational database in a co-training schema.

In [8], two-phase hierarchical meta-classification algorithm (NBSplit-train and NBSplit-test) is presented for relational databases with a semantic divide and conquer approach. This algorithm; in NBSplit-train, does not use aggregation function. And in NBSplit-test, the proposed algorithm uses a recursive, prediction aggregation technique with a modified Naive Bayes algorithm over heterogeneous classifiers applied on individual database tables (views) that combine labels of each classifier

[Downloaded from mjee.modares.ac.ir on 2025-01-16]

for the final class label. In [29] also used Naive Bayesian Combination to combine classifiers of each of the data sets which constructed in the Information Propagation Stage to obtain output label. In addition to, it used Decision Template. In order to, decision profile is created to combine classifiers output and decision template and decision profile Based on similarity measure is compared to get final output. Simple decision forests for multi relational classification is proposed in [1]. The basic idea of it is to independently learn different decision trees for different related tables, and then combine their contributions in a new log-linear model to predict class probabilities. In this work, we also proposed a method for upgrading ensemble learning to deal with multi relational data. Our proposed method uses tuple id propagation for create background tables with target attribute and then applies traditional classifier on data in each table and uses two steps combining classifier method to predict label for test data.

III. PROPOSED METHOD: MRE2C

The structure of the proposed method, multi relational ensemble classifier based on two-steps combining classifiers (MRE2C) is shown in Fig 1.

It can be seen from Fig 1, MRE2C mainly consists of four stages: propagating phase, training phase, meta phase and testing phase. The input of MRE2C is relational database (D) which have target table (TT) with target attribute (TA) and background tables(BT_i, i=1,…,n where n is the number of background tables that are linked to TT based on foreign key paths). In the following, we describe each stage.

A. *Propagating phase*

In propagating phase, we use tuple id propagation idea [9] to create different feature sets in background relations which these created tables are called propagated tables, PT_i, i=1,..,n. Propagating phase of MRE2C includes following stages:

- Obtain join path based on relational database schema for each background table (BT): relational database scheme is designed by a domain expert to related attributes into relations with very close semantic meaning [7]. We obtain join paths which is foreign key paths from target table to background table based on relational database schema for each background tables.
- Propagating tuple id and target attribute (TA) of target table (TT): we create feature sets in relations based on foreign key paths (join paths). In this regard, we propagate tuple id (primary key) and target attribute (TA) of target table (TT) to background tables based on join paths. For each background table, we use "select" query to join tables based on join path which is obtain in previous stage of propagating phase. Lastly, we create a propagated table for each background table based on foreign key paths from target table to that background table. The join paths in library database is shown in fig 3. For example, for background table author, we use book, writing and author path and join these three table and select attributes of author table and tuple id of target table and target attribute in result joint table.

The result propagated tables of propagating process in library database is shown in fig 4.

The target table and propagated tables are used as input of base classifiers in training phase.

B. *Training phase*

In this phase of MRE2C includes of two stages:

- train base classifiers in target table and propagated table: we select one of traditional classifiers as base classifier and train base classifier on each propagated tables and target table and we learn BC_T model in target table and BC_i model i=1,..,n in propagated tables. This phase is very similar to any traditional classification procedure.
- Use each trained base classifier to assign label training samples of that base classifier: we apply trained base classifier BC_i on training samples of PT_i (i=,..,n) and BC_T on training samples in TT. These labels of training samples are input of meta phase and will be used to create meta features for combining base classifiers.

C. *Meta phase*

In meta-learning, we learn from base classifiers (ensemble components); the output of base classifiers create meta-features as input of meta-learner [11]. In Meta phase of MRE2C, a meta-classifier is trained to combine the predications of base classifiers of ensemble into a single predication. In training phase, base classifiers are trained and in this phase meta-classifier is trained. There are one to many or many to many relationships in relational databases so one target tuple may be associated with multiple tuples in a background relation. But in our proposed method, we do not use aggregation function so related tuples to target tuple do not squeeze into one row. Therefore, we proposed two step combining classifier method. In first step, we use voting method in propagated tables to create meta-features as input to a meta-learner and second step, we learn the meta-learner. In this phase of MRE2C, we follow below steps according to fig2:

Local voting in each propagated table for each target tuple: In first step of combining, voting is done in each propagated table so called local voting. Each target sample in each propagated table assigns to each class based on a weight that this weight is calculated based on assigned labels of tuples of propagated tables which are associated that target sample. The weight of each class of each target tuple is equal to number of related tuples that assign to that class based on trained base classifier

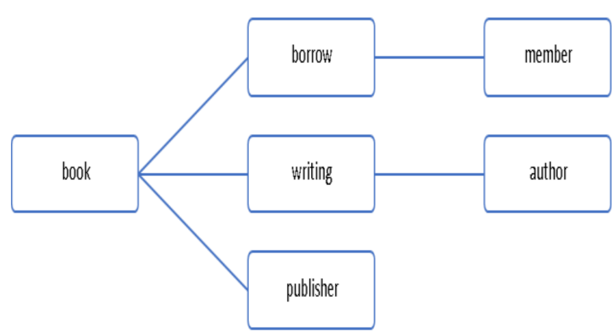


Fig2. Join paths in example library database

[Downloaded from mjee.modares.ac.ir on 2025-01-16]

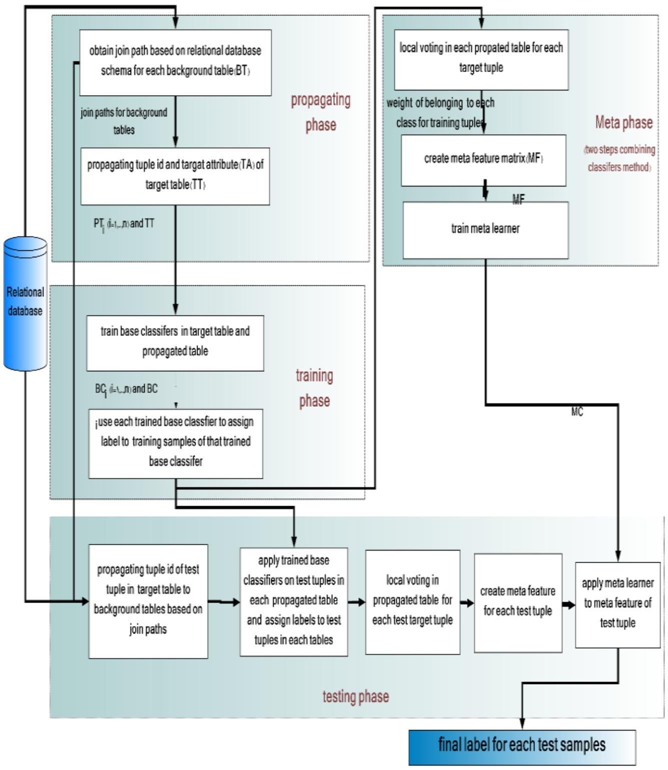


Fig 3. The framework of the proposed MRE2C method

- Create meta feature matrix(MF): We create meta-features to learn meta classifier for combining base classifiers. The n+1 sets of meta-features are extracted from the training data where n is the number of background tables, then a meta-feature matrix (MF) is created. MF is m by (n+1)×c+1 indication matrix where m is number of target tuples in target table and n is number of background table and c is number of class. The (i,j) entry of MF corresponds to the weight of belonging to the ith class(i=1,..,c) for jth(j=1,..,m) target tuples. In the fig 3, the structure of meta-feature matrix is shown.
- Train meta learner: we select one of traditional classifiers as meta classifer and we train meta-classifier (MC) on meta feature matrix and use this learned model(MC) in

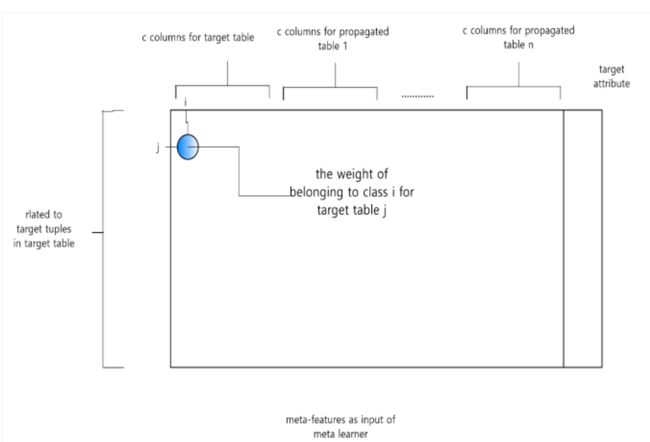


Fig 4- the structure of meta feature matrix

D. Testing phase

The learned model in training and meta phase and join paths of propagating phase are used as input of this phase and the following steps are follow,

Propagated BOOK table					
barcode	title	Number of version	Number of exist book	publisherID	status
B1	Relational Database	5	3	P1	Good
B2	Data structure	3	3	P2	Good
B3	History of 1995	2	2	P3	Bad
B4	Oracle database	5	5	P1	Good
B5	Mysql databases	6	4	P1	?

Propagated Writing table		
Barcode	AuthorID	status
B1	A1	Good
B2	A2	Good
B3	A3	Bad
B4	A1	Good
B5	A2	?

Propagated BORROWED table				
MemberID	barcode	Borrowed date	Must return data	Status
M1	B1	2015/2/17	2015/2/21	Good
M2	B5	2015/2/17	2015/2/21	?
M2	B2	2015/2/14	2015/1/19	Good
M2	B1	2015/2/14	2015/1/19	Good
M3	B4	2015/1/15	2015/1/20	Good

Propagated PUBLISHER table				
PublisherID	Address	Phone number	barcode	status
P1	Tehran	887695	B1	good
P2	Esfahan	666321	B2	good
P3	Tabriz	24789	B3	bad
P1	Tehran	887695	B4	good
P1	Tehran	887695	B5	?

Propagated MEMBER table					
Name	MemberID	Member type	Start period	barcode	status
Sara	M1	Student	2011	B1	Good
Jim	M2	Student	2012	B1	Good
Oliver	M3	Professor	2015	B4	Good
Jim	M2	Student	2012	B2	Good
Jim	M2	Student	2012	B5	?

Propagated AUTHOR table						
Name	AuthorID	Phone number	citation	Number of publication	barcode	status
Sara	A1	882895	2000	500	B1	Good
Tim	A2	776321	1570	352	B2	Good
jack	A3	56789	500	226	B3	Bad
Sara	A1	882895	2000	500	B4	Good
Tim	A2	776321	1570	352	B5	?

Fig 5- result propagated tables of propagating process in library database

- Propagating tuple id of test tuple in target table to background tables based on join paths: we propagated tuples to join tables are linked together in relational database schema.
- Apply trained base classifiers on test tuples in each propagated table and assign labels to test tuples in each tables: In this step, we apply BC_T on test data of target table and BC_i , i=1,..,n , on test data of propagated table i.
- Local voting in propagated table for each test target tuple: we use assigned labels of test tuples in previous step to do local voting and assign weights related to each class for target samples in each propagated table. This weight is calculated based on assigned labels of tuples of propagated tables which are associated that target sample.
- Create meta features for each test tuple: meta features of each test tuple are similar to meta feature in meta phase but have not value for target attribute.
- Apply learned meta learner to meta features of test tuple: in this step we apply meta leaner (MC) which is learned in meta phase on meta features of test tuple and obtain class label for each test target tuple.

[Downloaded from mjee.modares.ac.ir on 2025-01-16]

IV. Experimental results

In this section, several experiments are performed for illustration. There are mainly some commonly used data, financial database and mutagenesis database. The first database is used is used in the PKDD 1999 discovery challenge. The loan table is target table and have 682 records. The status of loan table is target attribute which indicates the status of the loan which have four mode i.e. A (finished and good), B (finished but bad), C (good but not finished), or D (bad and not finished). The background tables are relations Account, Client, Order, Transaction, Credit Card, Disposition and Demographic that are linked to the target table through directed or undirected foreign key chains. In this experiment, we classify samples of financial database based on which of loan is good or bad, regardless of whether the loan is finished or not. The data distribution of samples is quite skewed so randomly selected 324 positive instances, and all the negative loans to make the numbers of positive tuples and negative tuples more balanced same as [9].

The second database is used benchmark Mutagenesis dataset which it is composed of the structural descriptions of 188 Regression Friendly molecules that they are to be classified as mutagenic or not (Of the 188 instances 125 tuples are positive and 63 are negative). The background relations consist of descriptions about the atoms and bonds that make up the molecules, which include 4893 atoms and 5244 bonds.

In all experiment, the datasets are analyzed by means of ten-fold cross-validation. We compare our proposed method with following method:

- FOIL [30]: is popular ILP system which aims to induce complete and the consistent hypothesis to define a target relation.
- TIDLE [4]: is a method that upgrades traditional decision tree classifier to multi relational classification.
- RelAggs [32]: is a flatting approach that uses aggregation functions to transform multiple relations to one relation.
- CrossMine [9]: is an efficient ILP rule learner algorithm that it
  uses tuple ID propagation to virtually join tables, and randomly selects instances.
- Graph-NB [33]: is an upgrading approach that upgrade naïve Bayesian to relational settings.
- Simple decision forest [1]: independently learn different decision trees for different related tables, and then combine their contributions in a new log-linear model to predict class probabilities
- MRC [3]: is based on multi view learning and create different views and learn base classifier on each view and combine them by using meta-learner. It uses aggregation function.

This design evaluates some methods with different ways which they deal with multi relational data, upgrading approaches such as TIDLE, Graph_NB, Propositionalization approaches such as RelAggs, ILP based approaches such FOIL and CrossMine, approaches which are based creating multiple feature sets such as Simple decision forest, MRC and our proposed method.

The average classification rate of all methods in financial and mutagenesis database are reported in table 1 and 2. In two databases, accuracy of or proposed method is superior to other methods.

In financial database, FOIL which is based on ILP, has lower accuracy comparing with other methods. In this database, CrossMine is better than FOIL, since CrossMine use tuple id propagation idea to virtually join tables. In financial database, RelAggs has better accuracy when uses decision tree but in mutagenesis database, accuracy of RelAggs, whether uses decision tree or naïve Bayesian as traditional classifier, is not different.

TABLE I
RESULTS FOR FINANCIAL DB. FOR RELAGGS, THE TRADITIONAL CLASSIFIER IS SPECIFIED AND FOR MVC AND MRE2, THE BASE LEARNER AND META-LEARNER ARE SPECIFIED RESPECTIVELY

Algorithm	Accuracy(%)
TILDE	89
FOIL	72.8
CrossMine	85.8
Graph-NB	81
Simple decision forest	92
RelAggs(decision tree)	89
RelAggs(naïve Bayesian)	73.3
MRC(decision tree-decision tree)	87.8
MRC(decision tree-naïve Bayesian)	86.9
MRC(naïve Bayesian-naïve Bayesian)	89.8
MRE2C(decision tree-decision tree)	98.58
MRE2C(decision tree- naïve Bayesian)	99.5
MRE2C(naïve Bayesian - naïve Bayesian)	94.5

TABLE II
RESULTS FOR MUTAGENESIS DATABASE. FOR RELAGGS, THE TRADITIONAL CLASSIFIER IS SPECIFIED AND FOR MVC AND MRE2, THE BASE LEARNER AND META LEARNER ARE SPECIFIED RESPECTIVELY

Algorithm	Accuracy(%)
TILDE	85.6
FOIL	85.7
CrossMine	86.2
Graph-NB	86.2
Simple decision forest	89.1
RelAggs(decision tree)	85.1
RelAggs(naïve Bayesian)	85.1
MRC(decision tree-decision tree)	86.7
MRC(decision tree-naïve Bayesian)	88.9
MRC(naïve Bayesian-naïve Bayesian)	87.3
MRE2C(decision tree-decision tree)	91.7
MRE2C(decision tree- naïve Bayesian)	91.3
MRE2C(naïve Bayesian - naïve Bayesian)	89.89

We can also observe from Table 1 and 2 that simple decision forest, MRC and MRE2C achieve improved accuracy compared with flatting RelAggs method, two ILP based CrossMine and FOIL methods and upgrading Graph_NB and TIDLE methods. The simple decision forest, MRC and MRE2C methods create multiple feature sets based on relational database schema and train base classifiers on each feature set and combine base classifiers. In creating multiple feature sets, they use tuple id propagation idea but after propagating MRC uses aggregation function to squeeze related tuples to target tuple into one row but our proposed MRE2C method and simple decision forest are not use aggregation function due to information loss of using aggregation function. In two databases, simple decision forest and our proposed method are better than MRC(in three cases of MRC) since they do not use aggregation function so they do not lost information in creating multiple features step.

Simple decision tree use decision tree as base classifier and log-linear model to predict class probabilities for combing base

[Downloaded from mjee.modares.ac.ir on 2025-01-16]

classifier. We use a two-step combining method for aggregating predications of base classifiers which our proposed method uses local voting in result predications of learned model of each propagated table and then combines them by using meta-learner. Since accuracy of base classifiers are different may different and cannot be guarantee the comparable performances of the base classifiers, the meta learning method is more suitable so accuracy of or proposed method is superior to simple decision forest. Best accuracy in financial database is obtain by our proposed method when we use decision tree as base classifier and naïve Bayesian as meta-learner.

In order to validate the effectiveness of proposed MRE2C compared to MVC which use aggregation function in propagating phase, we compare this methods in six task of two data sets. We use three levels of background knowledge that appear in Table 2 for our experiments in mutagenesis database same as [25]. In financial database, we use three learned task same as [34]. The first learning task is to learn if a loan is good or bad from the 234 finished tuples. The second learning problem is to classify if the loan is good or bad from the 682 instances, regardless of whether the loan is finished or not. 400 examples in the target table are selected randomly same as previous experiment. The authors sampled the Transaction relation, since it contains an extremely large number of instances and discarded some positive examples from the target relation to make the learning problem more balanced.

Table 1 – description of Background knowledge for mutagenesis databases [25].

Background	Description
Bk_0	For each compound, it obtains atoms, bonds, bond types, atom types, and partial charges on atoms
BK_1	Consists of definitions in BK0 plus attributes indl and inda in the mole table
BK_2	Attributes logp and lumo are added to the mole table used in BK1

The average classification rate of MRE2C and MRC methods are reported in fig 6. We can also observe from fig 6 that MRE2C have better performance compared with MRC, so comparing MRE2C with MRC, do not using aggregation functions is clearly beneficial. The only exception is that the MRC and MRE2C algorithms achieved the same predictive performance against the Bk1 database.

In fig 7, we compare MRC and MRE2C based on different base classifier and meta learner on financial learning tasks. We can see that our proposed method in three leaning tasks is superior than MRC. In two task of three tasks (400 samples and 234 samples leaning tasks), MRE2C when uses decision tree as base classifier and naïve Bayesian as meta learner, is superior than others methods.

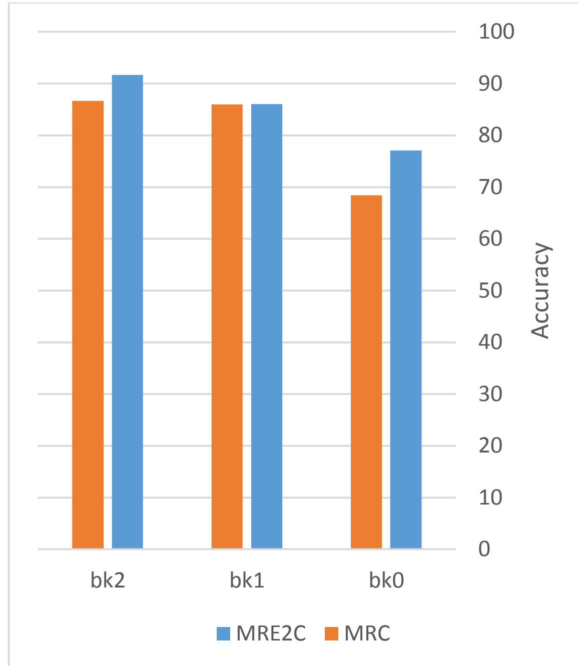
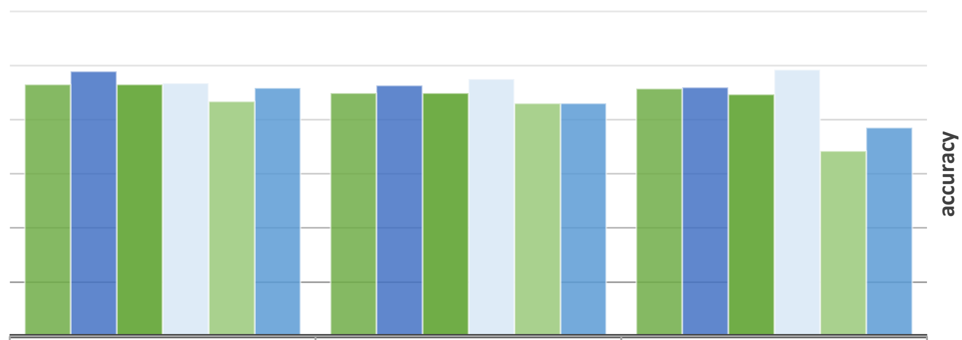


Fig 6- the accuracy of MRE2C and MRC in task learning of mutagenesis database

V. Conclusion

Many techniques are designed for relational data classification to deal multi relational data. some methods are proposed that create multiple feature datasets and select and train traditional classifiers as base classifier and combine result of base classifier to get the end result. In this regard we propose a method to constructing multi relational ensemble classifier based on two-step combining classifier. Firstly, we create different feature sets in relations based on foreign key paths (created tables are called propagated tables) and train traditional classifier as base classifier on each propagated tables and target table, then we create meta-features to learn meta classifier for combining base classifiers. The proposed method has been tested on two dataset and three benchmark tasks. Experiments performed on these real-world data sets show that the proposed method achieves promising results when compared with other popular relational data classification algorithms.

[Downloaded from mjee.modares.ac.ir on 2025-01-16]



	618 samples	400 samples	234 samples
MRE2C(decision tree-decision tree)	91.7	86.08	77.06
MRC(decision tree-decision tree)	86.7	86.01	68.4
MRE2C(decision tree-niave bayesin)	93.5	95	98.5
MRC(decision tree-niave bayesin)	93	89.8	89.3
MRE2C(niave bayesin-niave bayesin)	97.9	92.65	91.9
MRC(niave bayesin-niave bayesin)	93	89.8	91.5

Fig 7- the accuracy of MRE2C and MRC in task learning of financial database

References

[1] e. a. Bahareh Bina, "Simple decision forests for multi-relational classification," Decision Support Systems , vol. 54, no. 3, pp. 1269-1279., 2013.

[2] Raziyeh Zall, Mohammad Reza Keyvanpour, Multi Relational-upgraded methods: Classification and Analysis in supervised learning, The 1st Conference on Swarm Intelligence and Evolutionary Computation (CSIEC2016) , 2016

[3] H. Guo and H. Viktor, " Multi-relational classification: a multiple view approach," Knowledge and Information Systems, vol. 17, no. 3, 2008.

[4] L. I. Kuncheva and J. J. Rodríguez., "A weighted voting framework for classifiers ensembles," Knowledge and Information Systems , vol. 38 , no. 2, pp. 259-275, 2014

[5] H. Guo and H. L. Viktor., "Learning from Skewed Class Multi-relational Databases," Fundamenta Informaticae, vol. 89, no. 1, pp. 69-94, 2008.

[6] Z.-H. Zhou, "When semi-supervised learning meets ensemble learning," Frontiers of Electrical and Electronic Engineering in China , vol. 6, no. 1, pp. 6-16, 2011.

[7] M. Ceci, "Transductive relational classification in the co-training paradigm," Machine learning and Data mining in Pattern Recognition, pp. 11-25, 2012.

[8] G. Manjunath, M. M. Narasimha and D. Sitaram, "Combining heterogeneous classifiers for relational databases," Pattern Recognition, pp. 317-324, 2013.

[9] Y. Xiaoxin, "CrossMine: Efficient classification across multiple database relations," Springer Berlin Heidelberg, 2006.

[10] Q. Sun, S. Zeng, Y. Liu, P. Heng and D. Xia, "A new method of feature fusion and its application in image recognition," Pattern Recognition, vol. 38, no. 12, p. 2437–2448, 2005.

[11] E. Menahem, L. Rokach and Y. Elovici, "Combining one-class classifiers via meta learning," Proceedings of the 22nd ACM international conference on Conference on information & knowledge management. ACM, 2013.

[12] X. Liu, X. Wang, N. Japkowicz and S. Matwin, "An ensemble method based on adaboost and meta-learning," Advances in Artificial Intelligence. Springer Berlin Heidelberg, pp. 278-285, 2013.

[13] S. Kramer, N. Lavrač and P. Flach, "Propositionalization approaches to relational data mining," Springer Berlin Heidelberg, 2001.

[14] C. Perlich and F. Provost, "Distribution-based aggregation for relational learning with identifier attributes," Machine Learning, vol. 62, pp. 65-105, 2006.

[15] A. Knobbe, A. Siebes and B. Marseille, "Involving aggregate functions in multirelational search. In Principles of Data Mining and Knowledge Discovery," in Proceedings of the 6th European Conference, . Springer-Verlag, 2002.

[16] C. Vens, A. V. Assche, H. Blockeel and S. Dˇzeroski, "First order random forests with complex aggregates.," in In R. Camacho, R. King, and A. Srinivasan, editors,Proceedings of the 14th International Conference on Inductive Logic Programming, 2004.

[17] Q.-T. Dinh, C. Vrain and M. Exbrayat, "A Link-Based Method for Propositionalization.," ILP (Late Breaking Papers), 2012.

[18] W. Uwents and H. Blockeel, "Classifying relational data with neuaral networks," inductive logic programming, Springer Berlin Geidelberg, pp. 385-396, 2005.

[19] Y. Kavurucu, P. Senkul and I. H. Toroslu, "ILP-based concept discovery in multi-relational data mining," Expert Systems with Applications , vol. 36, no. 9,, pp. 11418-11428, , 2009.

[20] Y. Kavurucu, P. Senkul and I. Toroslu, "Concept discovery on relational databases: new techniques for search space pruning and rule quality improvement," Knowledge-Based Systems , vol. 23,, p. 743–756, 2010.

[21] M. Ceci, A. Appice and D. Malerba, "Mr-SBC: a MultiRelational Naive Bayes Classifier," in N. Lavrac, D.Gamberger, L. Todorovski & H. Blockeel (Eds.), Knowledge Discovery in Databases PKDD 2003, Lecture Notes in Artificial Intelligence, 2003.

[22] A. Appice, M. Ceci and D. Malerba, "Multi-Relational Model Tree Induction Tightly-Coupled with a Relational Database," Fundamenta Informaticae, vol. 129, no. 3, pp. 193-224, 2014.

[23] N. D. Bharwad and M. M. Goswami., "Proposed efficient approach for classification for multi-relational data mining using Bayesian Belief

[Downloaded from mjee.modares.ac.ir on 2025-01-16]

Network," Green Computing Communication and Electrical Engineering , 2014 International Conference , 2014.

[24] J.Neville, D.Jensen, L.Freidland and M.Hay, "Learning Relational Probability Trees," in knowledge discovery and data mining, 2003.

[25] H. Chen, H. Liu, J. Han, X. Yin and J. He, "Exploring optimization of semantic relationship graph for multi-relational Bayesian classification," Decision Support Systems, vol. 48, no. 1, pp. 112-12, 2009.

[26] W.-Y. Liu, K. Yue and W.-H. Li., "Constructing the Bayesian network structure from dependencies implied in multiple relational schemas," Expert Systems with Applications , vol. 38, no. 6, pp. 7123-7134,, 2011.

[27] H. Guo and H. L. Viktor., "Mining relational data through correlation-based multiple view validation," Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2006.

[28] H. Guo and H. Viktor, Mining relational databases with multi-view learning, Proceedings of the 4th international workshop on Multi-relational mining. ACM, 2005.

[29] U. Mistry and A. R. Thakkar, "Link-based classification for Multi-Relational database," Recent Advances and Innovations in Engineering (ICRAIE),2014. IEEE, 2014.

[30] J. R. Quinlan and R. M. Cameron-Jones, ""FOIL: A midterm report."," in Machine Learning: ECML-93. Springer Berlin Heidelberg, 1993.

[31] H. Blockeel and L. D. Raedt, " ‘Top-Down Induction of First-Order Logical Decision," Artificial Intelligence , vol. 101, no. 1-2, p. 285–297, 1998.

[32] M. Krogel and S. Wrobel, "Transformation-based learning using multirelational aggregation," Inductive logic programming. Springer Berlin Heidelberg, pp. 142-155, 2001.

[33] H. Liu, X. Yin and J. Han, "An Efficient Multi-relational Naïve Bayesian Classifier Based on Semantic Relationship Graph," USA, 2005.

[Downloaded from mjee.modares.ac.ir on 2025-01-16]