ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

Knowledge Fusion Technique Using Classifier Ensemble by Different Classification Rules

J. B. Patil, Prof. V.S. Nandedkar

ME Student, Dept. of Computer Engineering, Savitribai Phule Pune University, Pune, Maharashtra of India

Assistant Professor, Dept. of Computer Engineering, Savitribai Phule Pune University, Pune, Maharashtra of India

ABSTRACT: Classification rules are extracted from sample data known as knowledge. If we extract these knowledge in a distributed way, it is necessary to combine or fuse these rules. The task of data fusion is to identify the true values of data items among multiple observed values drawn from different sources of varying reliability. In data mining applications knowledge extraction is splitted into subtasks due to memory or run-time limitations. Again, locally extracted knowledge must be consolidated later because communication overhead should be low. Extracting information from multiple data sources, and reconciling the values so the true values can be stored in a central data repository. But it's a problem of vital importance to the database and knowledge management communities.

In a conventional approach extracting knowledge is typically done either by combining the classifiers' outputs or by combining the sets of classification rules but in this paper, I introduce a new way of fusing classifiers at the level of parameters of classification rules. Here its focused around the utilization of probabilistic generative classifiers utilizing multinomial circulations and multivariate ordinary dispersions for the consistent ones. We are using these distributions as hyper distributions or second-order distributions. Fusing of these classifiers are can be done by multiplying the hyper-distributions of the parameters.

KEYWORDS: Knowledge Engineering, Training, Classifier Fusion, Probabilistic classifier, Knowledge Fusion, Generative Classifier, Bayesian techniques.

I. INTRODUCTION

Classification is a data mining function that assigns items in a collection to target classes. Goal of classification is to accurately predict the target class for each data case. Classification tasks are begins with a data set in which the class assignments are known. The simplest type of classification problem is binary classification in which the target attribute has only two possible values. In the model, a classification algorithm finds relationships between the values of the target and values of the predictors. Also Different classification algorithms use different techniques for finding relationships. We summarize these relationships in a model and then applied to a different data set in which the class assignments are unknown.

However, a more detailed analysis of current applied results does reveal some puzzling aspects of unlabeled data. Researchers have reported cases where the addition of unlabeled data degraded the performance of the classifiers when compared to the case in which unlabeled data is not used. These cases were not specific to one type of data, but for different kinds, such as sensory data , computer vision , and text classification .To explain the phenomenon, we began by performing extensive experiments providing empirical evidence that degradation of performance is directly related to incorrect modelling assumptions. Here we are going to estimate the parameters of a Naive Bayes classifier with 10 features using the Expectation-Maximization (EM) algorithm with varying numbers of labelled and unlabeled data.

Section 2 briefly issues the related work. In section 3 we mainly introduce proposed system and framework. Module wise experimental results are shown in section 4.. And finally the conclusion is introduced in section 5.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

II. RELATED WORK

In many real world applications, the information from heterogeneous sources has to be fused for obtaining optimal results and decisions respectively.

By the transformation of the available knowledge, the corresponding uncertainties, and the given dependencies into Degree of Belief (DoB) distributions. Here likelihood probabilistically draws conclusions from the given information to an underlying state of the world that is not directly observable.

Existing system was done with the help of Bayesian theory. Which states that the information provided by the sources influences the posterior DoB distribution only. Bayesian inference Results consists of the complete posterior DoB distribution of given the processed information.

Disadvantage:
The mean squared error for continuous quantities or the maximum.
The classification error probability in the discrete case.(estimates no longer lossless.)
The error correction and data loss may become higher.

Also any application in the field of distributed intrusion detection in computer networks, it is very much impossible to exchange the raw data because of limited communication bandwidth and also a central unit would constitute a single point of failure. In data mining applications, extraction of knowledge is split into subtasks due to runtime limitations or memory. Second order-distributions (Hyper-distributions) are retained through the fusion process. It is very useful to weight single decisions in the class posterior probabilities when several classifiers are combined. Also a rejection criterion could easily be defined which allows to refuse a decision if none of the class posteriors reaches a pre-specified threshold.[1].

Distributed intrusion detection: An application in the field of distributed intrusion detection in computer networks is described in [2]. it is impossible to exchange the raw data because of a limited communication bandwidth. And its also a central unit would constitute a single point of failure. Knowledge extraction is split into subtasks due to runtime limitations in other data mining applications.

According to [7], the conjugate prior of a multinomial is a Dirichlet distribution and the conjugate prior of a multivariate normal is a normal-Wishart distribution. The Bayesian knowledge fusion focuses on the 2nd category. The term “Bayesian knowledge fusion” (which we also claim for our work) is often associated with first category. Several variants can be found in Bayesian estimation techniques, while the most interesting ones are sequential Bayesian estimation techniques [9] or the fusion of several likelihood functions as in the case of the independent likelihood pool. Robotics, multimedia, or target detection are outline more details on this technique which is quite distinct from ours as it addresses the first order and not the second-order distributions can be found in [10].

Work in category 3 fuses, the output of “low-level” classifiers by averaging their labels or using their labels as input of a decision unit that could also be trained. Also complex approaches are bagging or boosting which are often motivated by the idea that an ensemble of “weak classifiers”. Work in category-2 essentially depends on the kind of knowledge representation. Normally two main fields can be identified:

1. Knowledge is often equated with constraints and there is some work focusing on fusion of constraints.
2. Knowledge is often represented by graphical models that are subject to fusion, for example, Bayesian networks, (intelligent) topic maps.

The Bayesian fusion approach based on hyper distributions. There is an article that is closely related to approach which also describes a Bayesian fusion approach based on hyper parameters and it also exploits the concept of conjugate priors. This work is much more concrete than the hyper parameter consensus technique concerning the derivation of fusion formulas and the application to classifier fusion.

III. PROPOSED SYSTEM AND FRAMEWORK

We proposed an probabilistic generative classifiers using multinomial distributions. Knowledge represented by components of classifiers fused at a parameter level in knowledge fusion. Probabilistic classifiers provide outputs that can be interpreted as conditional probabilities as they model the conditional distribution of classes given an input

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

sample. Generative classifiers aims at modelling the processes from which the sample data are assumed to originate. The posterior class probabilities are very useful to weight single decisions when several classifiers are combined together. And the rejection criterion could easily be defined which allows to refuse a decision if none of the class posteriors reaches a pre-specified threshold. In case of dynamic environments it is possible to detect novel situations.

A. System architecture

Proposed System working on the basis of fundamental models these are.

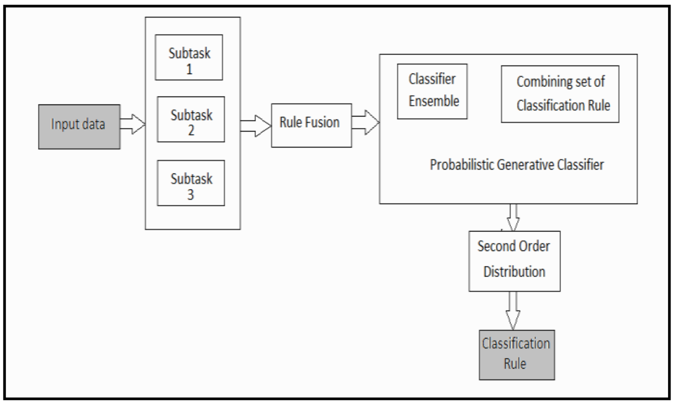

Fig. 1. Proposed System Architecture

Input data : Classification rules are extracted from input sample data in a distributed way, it is necessary to combine or fuse these classification rules.

Subtask : In various machine learning applications, the task of knowledge extraction (e.g., classification rules) from input sample data is divided into a number of subtasks. In the data mining applications, due to runtime limitations or memory, knowledge extraction is split into subtasks.

Rule Fusion : At some point, there is necessity to fuse or to combine the knowledge that is now “contained” in a number of classifiers in order to apply it to new data. Our fusion process uses the second order distribution (hyper distributions) obtained in training process. Here retain these hyper distributions throughout the fusion process, and having several advantages over a simple linear combination of classifiers parameters.

Probabilistic Generative Classifier: The probabilistic classifiers offer the possibility to combine classifiers at the level of components of the mixture models (in the following these components are also referred to as “rules”). This can be accomplished by taking the union of all component sets and renormalizing the mixture coefficients.
Components or rules may be fused at the level of parameters. In this case, it is necessary to “average” the parameters of two or several components in an appropriate way if these components are regarded as being “sufficiently” similar.

a. Classifier Ensemble: First, the classifiers can be used in the form of ensembles, an idea for which a number of realizations exist. For probabilistic classifiers the outputs can be interpreted as posterior probabilities.
b. Combining Classification Rule: If for a component of the first classifier, corresponding component of the second classifier not exists or vice versa. These components are only combined in the resulting classifier. That means union of these component sets is built and the mixture coefficients are adapted accordingly.

*Second Order Distribution :*The key contribution of this work is that it shows an actual fusion of classifiers (or, components of classifiers) can be accomplished essentially by multiplying the second-order distributions if the classifier is based on certain members of the exponential family of distributions. Distributions such as Dirichlet or normal-Wishart distributions over parameters of the classifier.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

Classification Rule : Output of the system is Knowledge represented by components of classifiers fused at a parameter level.

b. Mathematical Model

Here We are going to represent mathematical model by using state diagrams where possible states are represented by nodes and operations or activities represented by arcs.

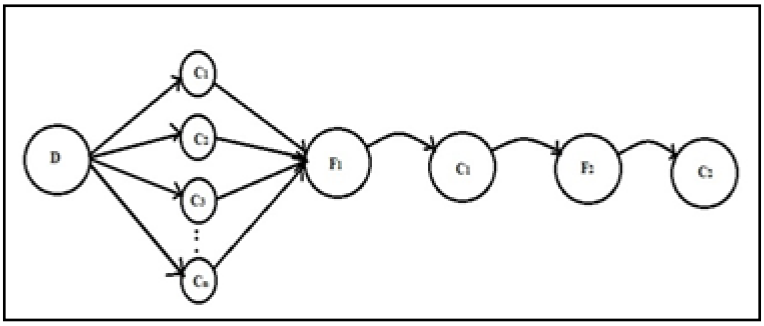

Fig 2. Mathematical Model

States:
D: Raw data
C(Clusters)={C1,C2,...Cn}
F1: Fusion Output
C1:Output of probabilistic generative classifier.
F2:Second order distributions output
C2: Second classifiers output (required value)

Operations:
O1: Extract datasets in distributed manner.
O2: Apply clustering algorithm.
O3: Apply Rule Fusion (Normal Wishart and Dirichlet algorithm-algorithm2).
O4: Use Probabilistic Generative classifier for ensemble.
O5: Use VI algorithm for training (Second Order Distribution).
O6: Use Dirichlet algorithms.(Second order classifier).

Datasets(D): Datasets for which this system works may be either artificial or real world datasets. eg. adult dataset.

Clusters(C): Which is used for forming number of clusters which is needed for classifiers.
C={C1,C2,...Cn}

Fusion(F):
Algorithm: FUSION AND COMBINATION
Input : Two sets of distribution C1and C2
Output : Fused classifier

1. C` ← θ
2. foreach c1in C1do
3. found ← false
4. for each c2in C2do
5. dist ←_H_dist(c1; c2)
6. if H_dist > v_H and class(c1) == class2 then
7. C`.add(false(c1; c2))
8. C2.remove(c2)
9. found ← true
10. break;
11. if not found then

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

```
C'.add(c1)
for each c2in C2do
C`.add(c2)
classifier ← θ
for each component in C` do
classifier.add ( )
return classifier;
```

IV. SIMULATION RESULTS WITH MODULES

We distribute our system into five modules as follows

1. Implementation of Knowledge Classification:

First, the classifiers can be used in form of ensembles ,in which a number of realizations exist . For probabilistic classifiers, where the outputs interpreted as posterior probabilities. Second, probabilistic classifiers offer the possibility to combine classifiers at the level of components of the mixture models. Third, components or rules may be fused at the level of parameters. In this case it is necessary to "average" the parameters of two or several components in an appropriate way if these components are regarded as being "sufficiently" similar.

In our module we used our trained dataset which is get extracted and viewed to an user and thus the data set is get classified into classifier-1 and another classification as classifier-2 where these two classification are done with the help of our classification rule based on the position and parameters and thus the classification are get fused at the final step of our implementation based on our rule(Fig.3).

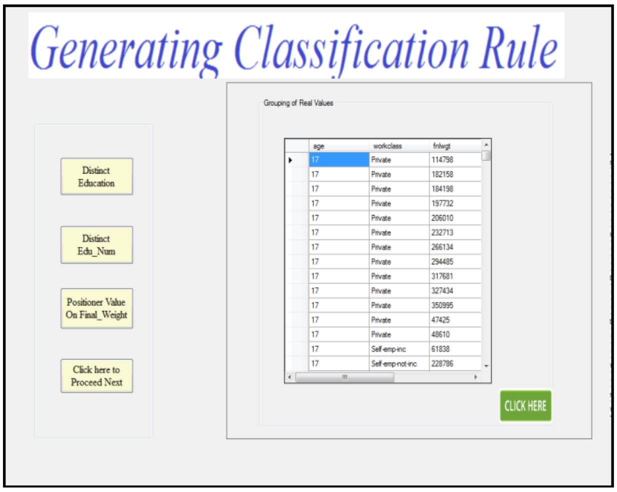

Fig 3. Knowledge classification

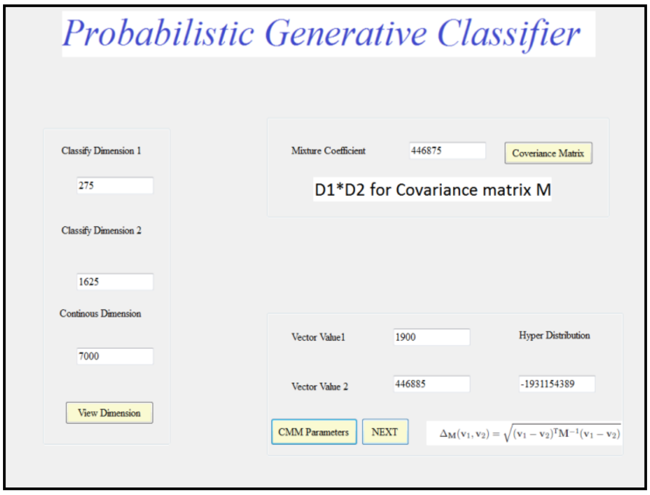

Fig4.Probabilistic Classification

2. Implementation Generative Classifier

In this module we designed the generative classifier where each classification count is get taken that is based on the user view the distance count is get calculated for an classification 1 and similar to that the distance count is get calculated to the classification 2 and from that formulae we found an Mahalanobis distance is then get calculated. And the CMM Parameters also get calculated with the help of Vectors and Parameters we found with the help of implementing our algorithm (Fig.4).

3. Implementation of VI Training Algorithm.

The hyper-distributions of the CMM classifier are trained using a Variation Inference (VI) algorithm. (VI) algorithm can be seen as a Bayesian variant of the well-known Expectation Maximization (EM) algorithm. The basic idea of VI is to find the joint posterior distribution. In case of the VI algorithm it is convenient to work with the precision matrix instead of the covariance matrix contains the training data is a set of latent variables each of which describes the gradual "assignment" of a sample to the components of the classifier (Fig 5).

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

Fig.5 Variation Inference Training

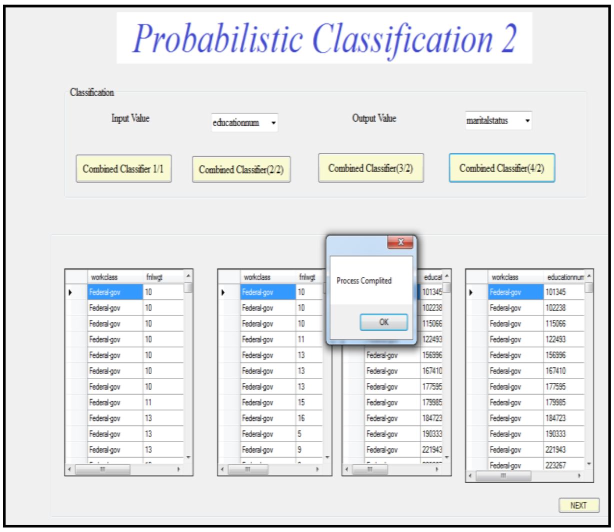

Fig 6 Probabilistic Classification

4. Probabilistic Classification:

The Main work of this module is, we divide our classifies into 4 division based on the probability each classification that is classifier 1 and classifier 2 is get divided and from that the fusion is being get formed. And based on the likelihood and the positioner value the classifier is get fusion based on our formula (Fig 6).

5. Fusion Training and Analysis:

The conjugate prior distribution that must be used to estimate the parameters of a multinomial distribution is a Dirichlet distribution. In order to fuse two Dirichlet distributions we multiply their density functions and then divide the result by the prior. The knowledge that we have a certain distribution type implicitly gives us a suitable normalizing factor for the fused distribution. In this module we fuse two classifiers based on the likelihood and positioner value the entire data where get viewed and from the fusion value we found an error detecting code for each classifiers with the generated value and the fusion value is get formulated (Fig 7 & Fig 8) .

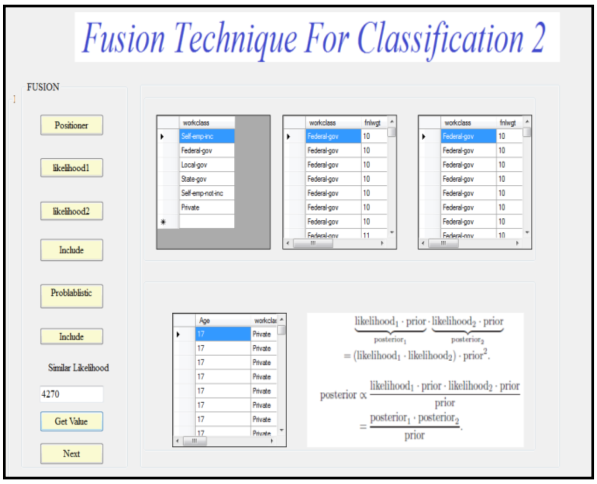

Fig 7 Likelihood function Implementation

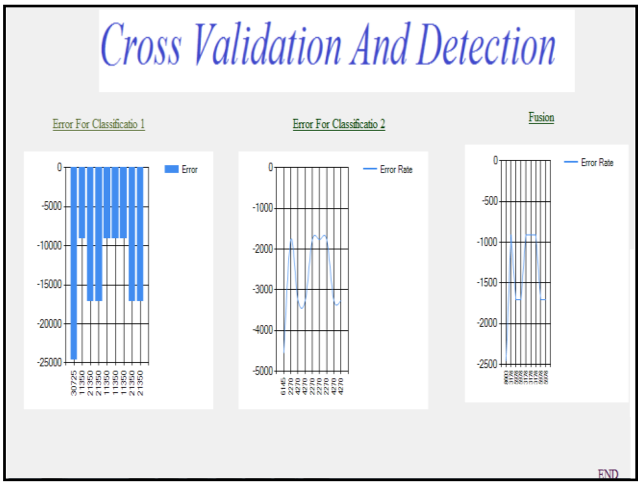

Fig 8 Cross verification

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

V. CONCLUSION

In this system we are going to present a novel technique to fuse two probabilistic generative classifiers into one even if these classifiers work on different distributions. A Classifier Mixture Model consists of several components each of which may in turn consist of one multivariate normal distribution modeling continuous dimensions of the input space and multiple multinomial distributions, one for each categorical dimension of the input space. To identify components of two classifiers that shall be fused, I suggested a similarity measure that operates on the distributions of the classifier. The actual fusion of components works one level higher on hyper-distributions. And these are the result of the Bayesian training of a CMM using the VI (variation Bayesian inference) algorithm. We use formulas to fuse both Dirichlet and Normal-Wishart distributions. These formulas are the conjugate prior-distributions of the multinomial and normal distributions of a CMM in a very elegant way with benchmark datasets outlined the properties of this new knowledge fusion approach.

REFERENCES

1. Dominik Fisch,Eldar kalskowski and Bernhad Sick,"Knowledge Fusion for Probablistic Generative Classifiers with Data Mining Applications",IEEE Transactions On Knowledge And Data Engineering, Vol. 26, NO. 3, March 2014.
2. Rupali Vij, Suresh Kumar "Improved K- means Clustering Algorith For-Two Dimensional Data", CCSEIT-12, October 26-28, 2012, Copyright © 2012 ACM 978- 1-4503-1310-0/12/10.
3. You Li, Kaiyong Zhao, Xiaowen Chu, and Jiming Liu, "Speeding up K-Means Algorithm by GPUs" 2010 10th IEEE International Conference on Computer and Information Technology (CIT 2010).
4. Brijesh Verma Ashfaqur Rahmanseniou Member, IEEE."Cluster oriented ensemble Classifier: Impact of multicluster characterization on ensemble classifier learning" Vol.24 No. 4.April 2012.
5. Dominik Fisch, Student Member, IEEE, Thiemo Gruber, and Bernhard Sick, Member, IEEE "SwiftRule: Mining Comprehensible Classification Rules for Time Series Analysis" IEEE Transactions On Knowledge And Data Engineering, Vol. 23, NO. 5, MAY 2011.
6. JiyanyongWang, George Karypis,IEEE seniour member, "ON Mining Instance- Centric Classification Rules", IEEE Transactions On Knowledge And Data Engineering, Vol. 18, NO. 18, Nov 2006..
7. D. Fisch, B. Ku¨ hbeck, B. Sick, and S.J. Ovaska, "So Near and Yet So Far: New Insight into Properties of Some Well-Known Classifier Paradigms," Information Sciences, vol. 180, no. 18, pp. 3381-3401, 2010.
8. J. Sander and J. Beyerer, "Fusion Agents—Realizing Bayesian Fusion via a Local Approach," Proc. IEEE Int'l Conf. Multisensor Fusion and Integration for Intelligent Systems, pp. 249-254, 2006.
9. A. Makarenko and H. Durrant-Whyte, "Decentralized Data Fusion and Control in Active Sensor Networks," Proc. Seventh Int. Conf. Information Fusion, pp. 479-486, 2004.
10. H. Durrant-Whyte and T. Henderson, "Multisensor Data Fusion," Springer Handbook of Robotics, B. Siciliano and O. Khatib, eds hapter 25, pp. 585-610, Springer, 2008 .
11. L. Iocchi, N. Monekosso, D. Nardi, M. Nicolescu, P. Remagnino, and M. Valera, "Smart Monitoring of Complex Public Scenes," Proc. Assoc. for the Advancement of Artificial Intelligence 0(AAAI) Fall Symp., 2011.

BIOGRAPHY

Mr. J. B. Patil is a ME student in the Computer Engineering Department, PVPIT Bavdhan Pune, Savitribai Phule Pune University. He received Bachelor of Information Technology degree in 2010 from Shivaji University, Kolhapur, MS, India. His research interests are Computer Networks (wireless Networks), Data mining, Computer Algorithm etc.

Prof. V. S. Nandedkar is an Assistant professor with more than 11 Years Teaching Experience in Savitribai Phule Pune University. She received Master of Engineering in Computer Science and Engineering and doing PhD work. Her research interests are Image Processing and Software Engineering.