Journal of Theoretical and Applied Information Technology
10th February 2013. Vol. 48 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# CMER: CLASSIFICATION BASED ON MULTIPLE EXCELLENT RULES

[1]XUEJUN WANG, [2]ZHONGMEI ZHOU, [3]JUN ZHENG
[1, 2] Department of Compute Science and Engineering, Zhangzhou Normal University, Zhangzhou, 363000 China, [3]Department of Biomedical Engineering, The First Affiliated Hospital & College of Medicine, Zhejiang University ,Hangzhou, 310003, China
E-mail: lilium,snow@163.com

## ABSTRACT

One of the traditional rule-based classification tasks is to build a set of high quality classification rules for prediction. Traditional rule-based classification approaches can achieve high efficiency. However, some traditional rule-based classification methods usually generate few rules. They may miss some high quality rules, especially when the training data set is small. Therefore their accuracy may not be high in some data sets. In this paper, we propose a new classification approach called CMER (classification based on multiple excellent rules). CMER is distinguished from other traditional rule-based classification methods in three aspects. First, CMER constructs a candidate set and a seed set. Second, CMER connects the seed set with the candidate set to produce more classification rules at a time. Third, CMER uses the minimum support and foil gain to update the seed set. As a result, CMER generates more excellent classification rules, especially when the training data set is small. Our experimental results show that CMER gets higher accuracy than some traditional rule-based classification methods.

**Keywords:** *Date Mining(DM), Rule-based Classification(RC), Association Classification(AC)*

## 1. INTRODUCTION

Classification is an important task in data mining. It is generally divided into two steps. First, we construct classification model based on training data. Second, we use the model to classify the testing data. In 1998, Liu [1] proposed a new classification approach called associative classification. It generates a complete set of class association rules and then selects a set of high quality rules for prediction. In recent years, many improved associative classification approaches have been proposed, such as CMAR[2], CPAR[3], MMAC[4], MCAR[5] and CSMC[6] . Associative classification [7, 8, 11]can achieve higher accuracy than some traditional classification approaches, such as decision trees[9], FOIL[10]. However the accuracy of the associative classification depends on the settings of minimum support and minimum confidence. Moreover, when the minimum support is set to be low, it generates a large number of rules. It is difficult for us to select a set of high quality classification rules.

The traditional rule-based classification approaches [12]-[15] can achieve higher efficiency than associative classification. However, the accuracy of some traditional rule-based classification approaches may not be high in some data sets. One of the reasons is that they usually generate a small set of classification rules, especially when the training data set is small. For example, FOIL (First Order Inductive Learner) is one of the traditional rule-based classification approaches. FOIL selects the most important literal to generate one rule at a time. In this way, FOIL generates fewer rules. As a result, the accuracy of FOIL is not high in some data sets. So does the method of the decision tree. Each training example in the method of decision tree is covered by only one classification rule.

In order to solve these problems, we propose a new classification method called CMER (classification based on multiple excellent rules). In comparison with some traditional rule-based approaches, CMER has some advantages: (1) CMER constructs a candidate set and a seed set. Both the candidate set and the seed set are consisted of some important literals. (2) We connect the seed

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

set with the candidate set to produce more high quality rules at a time. (3) CMER employs the minimum support and foil gain to update the seed set. The experimental results show that CMER produces more high quality classification rules at a time and gets higher accuracy.

The outline of this paper is as follows. In section 2, we give some definitions and compare the differences between FOIL and CMER in rule extraction. We use an example to describe the main ideas of the CMER and develop the algorithm of CMER in section 3. We report the experimental results in section 4. We conclude the paper in section 5.

## 2. RULE-BASED CLASSIFICATION

In this part, we first give some definitions. Then we compare the differences between FOIL and CMER in rule exaction.

Let $T$ be a set of tuples with m district attributes $(A_1, A_2, \ldots, A_m)$ and the set of class be $C = \{C_1, C_2\}$. The measurement of literals and rules is defined as follows.

Definition 1. (Foil Gain) Suppose that $v$ is a literal. There are $|P|$ positive examples and $|N|$ negative examples in $T$. Then the Foil gain of literal $v$ is defined as follows:

$$Gain(V) = |v_1|(\log(|v_1|/(|v_1|+|n_1|)) - \log(|P|/(|P|+|N|))) \quad (1)$$

where $|v_1|$ means the number of positive examples which contain literal $v$, $|n_1|$ means the number of negative examples which contain literal $v$.

Definition 2. (support of a pattern ) The support of pattern $X$ is defined as follows:

$$\sup(X) = (count(X)/|T|) \times 100\% \quad (2)$$

where $count(X)$ means the number of examples in $T$ that contain pattern $X$, $|T|$ is the number of examples in a dataset.

Definition 3. (confidence of a pattern) The confidence of pattern $X$ is given as follows:

$$conf(X) = (count(X_c)/count(X)) \times 100\% \quad (3)$$

where $count(X_c)$ means the number of examples which contain pattern $X$ and have a class value $c$.

In the following, we give the differences between FOIL and CMER. First, FOIL uses the foil gain to select one literal with the best foil gain. CMER selects several important literals to construct the candidate set and the seed set. Second, FOIL selects one literal from the conditional base of the just found literal to generate one rule at a time. CMER connects the seed set with the candidate set to produce several rules at a time. By doing so, CMER can generate a larger set of rules than FOIL, especially when the training data set is small. The experimental results show that CMER can achieve higher accuracy than FOIL in most data sets.

## 3. CMER

In this section, first, we use an example to describe the process of rule mining in CMER. Second, we develop the algorithm of CMER. Finally, we give the measurement of classification rules and predict new examples.

### 3.1 Inducing Rules

We first construct a candidate set and a seed set. Then we connect the seed set with the candidate set to produce pattern and rules. Third, we update the seed set. Finally, CMER removes the examples covered by the just found rules and iterates the process. The detailed process of inducing rules in CMER is shown in the following example.

Example 1. The training data set $T$ is shown in Table 1. We suppose that the attribute $J$ is the decision attribute and others are the condition attributes. In this training data set, we suppose that all examples which have the class $(J = 89)$ are positive examples and all examples which have the class $(J = 90)$ are negative examples. Let the minimum support be $30\%$ and the minimum confidence be $100\%$. We construct classification rules for the class $(J = 89)$.

First, we construct a candidate set. If the foil gain of literal $p$ is greater than zero, then literal $p$ is positively correlated with class $(J = 89)$. We select literal $p$ as an element of a candidate set. By this way, we have the candidate set as shown in Table 2.

Second, CMER selects several important literals from the candidate set to form a seed set. The average foil gain of all literals in table 2 is 0.784. We use the average foil gain as the threshold of minimum foil gain. If the foil gain of a literal $p$ is greater than the threshold, and if the support of the literal p is greater than the minimum support, then

Journal of Theoretical and Applied Information Technology
10th February 2013. Vol. 48 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

literal $p$ is selected as an element of a seed set. We have the seed set as shown in the Table 3.

*Table 1. A Training Data Set*

| C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|
| 19 | 30 | 55 | 60 | 72 | 79 | 82 | 89 |
| 11 | 35 | 55 | 59 | 73 | 80 | 88 | 90 |
| 19 | 35 | 53 | 60 | 73 | 81 | 88 | 89 |
| 19 | 35 | 55 | 59 | 73 | 81 | 88 | 90 |
| 11 | 35 | 55 | 58 | 72 | 81 | 88 | 89 |
| 19 | 35 | 53 | 58 | 72 | 81 | 88 | 90 |
| 11 | 30 | 54 | 59 | 73 | 80 | 88 | 89 |
| 19 | 30 | 54 | 59 | 72 | 80 | 82 | 90 |
| 11 | 30 | 54 | 60 | 73 | 81 | 88 | 89 |
| 11 | 30 | 55 | 60 | 72 | 79 | 82 | 90 |

*Table 2. The Candidate Set*

| Attribute | Value | Gain | Support |
|---|---|---|---|
| F | 60 | 1.216 | 40% |
| B | 9 | 1.150 | 60% |
| I | 88 | 1.150 | 70% |
| E | 54 | 0.575 | 30% |
| C | 11 | 0.546 | 50% |
| D | 35 | 0.546 | 50% |
| G | 73 | 0.546 | 40% |
| H | 81 | 0.546 | 50% |

*Table 3. The Seed Set*

| Attribute | Value | Gain | Support |
|---|---|---|---|
| F | 60 | 1.216 | 40% |
| B | 9 | 1.150 | 60% |
| I | 88 | 1.150 | 70% |

*Table 4. The Generated Rules*

| Patterns | Confidence |
|---|---|
| F=60, B=9 | 100% |
| F=60, I=88 | 100% |
| F=60, E=54 | 100% |
| F=60, D=35 | 100% |
| F=60, G=73 | 100% |
| F=60, H=81 | 100% |
| B=9, H=81 | 100% |
| I=88, E=54 | 100% |

*Table 5. The New Seed Set*

| Patterns | Conf | Sup | Gain |
|---|---|---|---|
| B=9, I=88 | 75% | 40% | 1.217 |
| I=88,C=11 | 75% | 40% | 1.216 |

Third, CMER connects the seed set with the candidate set to produce patterns and classification rules. If the confidence of pattern $X$ is $100\%$, then $X \rightarrow C$ is a classification rule. We get a set of classification rules as shown in Table4. If the foil gain of pattern $X$ is greater than the threshold, and if the support of pattern $X$ is greater than the minimum support, then pattern $X$ is selected as an element of the new seed set. We have the new seed set as shown in Table5.

Fourth, CMER connects the new seed set with the candidate set to produce new rules until there is no new seed set generated.

Finally, CMER removes the examples that are covered by the just found rules and iterates the process. To learn rules for class $(J = 90)$, the process is repeated. We give the algorithm of rule extraction of positive examples in CMER below.

Input: Training data $T = P \cup N$ ( $P$ and $N$ are the sets of all positive and negative examples, respectively), minimum support, minimum confidence

Output: A set of classification rule

rule set $R \leftarrow \varnothing$, candidate set $cs \leftarrow \varnothing$, seed set $ss \leftarrow \varnothing$
while ( $|P| > 0$ ) do
 compute the foil gain of each literal $p$ in $P$
 if ( $gain(p) > 0$ ) then
  $cs \leftarrow cs \cup \{p\}$
 end if
 compute the average foil gain of all literals in $cs$
 if ( $gain(p) >$ average foil gain) then
  $ss \leftarrow ss \cup \{p\}$
 end if
  while( $ss \neq \varnothing$ and $\forall$ pattern $X \in ss$, $X\_length < \max\_rule\_length$ )do
  connect the seed set with the candidate set to generate patterns
  compute the confidence, support and confidence of each pattern $X$
   if ( $\mathrm{conf}(X) = 100\%$ ) then
    $R \leftarrow R \cup \{X\}$
   else if ( $\sup(X) > \min\sup$ and $gain(X) >$ average foil gain)
     $ss \leftarrow ss \cup \{X\}$
    else delete $X$
  end if
 end while
 remove from $P$ that all examples satisfying the rules
end while

*Figure 1: The Algorithm of CMER*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

### 3.2 Classification

In this part, first, we give the measurement of the classification rules. Second, we use the best k rules for prediction new examples.

For rule $r : X \rightarrow C$, we use the expected accuracy [3] to estimate the importance of rule $r$. The expected accuracy of rule $r$ is given by (4):

$$Laplacy(r) = (n_c + 1) \big/ (n_{tot} + k) \qquad (4)$$

where $k$ is the number of classes, $n_{tot}$ is the total number of examples which contain pattern $X$, $n_c$ is the total number of examples which contain pattern $X$ and have the class $c$.

For a testing example, we first select the best $k$ rules which are matched by the example. If all the best $k$ rules have the same class label, we just simply assign that label to the testing example. If we have $k1$ rules which belong to class 89 and $k2$ rules which belong to class 90 among the best k rules. If the average expected accuracy of k1 rules is higher than $k2$ rules, then we assign the class 89 to the testing example.

## 4. EXPERIMENTS

All the experiments are performed on a 1.83GHz PC with 2.49 G main memory, running Microsoft Windows XP. We tested our algorithm on the Mushroom data set.

In the table 6, we choose the size of the training data from 100 to 1000 in turn. We select the training data by random. We choose the size of the testing data 500. We select them from 0-5500 in turn. In the table 6, we give the average accuracy of FOIL and CMER respectively. Moreover, we also give the number of classification rules of CMER and FOIL respectively. We select the best 3 rules for prediction. The minimum support is set to be 5%. The minimum confidence is set to be 100%. The experimental results are shown in Table 6. From the Table 6, we can get the following conclusions: (1) the accuracy of CMER is higher than Foil no matter how large the training date is. (2) When the training data is small, CMER can achieve much higher accuracy than FOIL. (3) The number of rules in CMER is larger than FOIL.

In the table 7, the training data is the same as above. The minimum support is varied from $3\%$ to $9\%$. From table 7, we can see that the accuracy of CMER does not depend on the settings of the minimum support.

*Table 6: Accuracy and rules of FOIL and CMER*

| | FOIL | | CMER | |
|---|---|---|---|---|
| Size | Accuracy | Rules | Accuracy | Rules |
| 100 | 0.984009 | 9 | 0.995457 | 51 |
| 200 | 0.985465 | 11 | 0.996730 | 68 |
| 300 | 0.988375 | 9 | 0.995275 | 73 |
| 400 | 0.988014 | 11 | 0.997819 | 83 |
| 500 | 0.992008 | 12 | 0.995638 | 96 |
| 600 | 0.990919 | 14 | 0.997456 | 118 |
| 700 | 0.994912 | 10 | 0.996002 | 127 |
| 800 | 0.998545 | 15 | 1.000000 | 148 |
| 900 | 0.995277 | 15 | 0.998001 | 185 |
| 1000 | 0.996367 | 16 | 0.998182 | 231 |

*Table 7: Performance of CMER in Different Support*

| Size | sup=3% | sup=5% | sup=7% |
|---|---|---|---|
| 100 | 0.995457 | 0.995457 | 0.995457 |
| 200 | 0.996730 | 0.996730 | 0.996730 |
| 300 | 0.995275 | 0.995275 | 0.995275 |
| 400 | 0.997819 | 0.997819 | 0.997819 |
| 500 | 0.995638 | 0.995638 | 0.995638 |
| 600 | 0.997456 | 0.997456 | 0.997456 |
| 700 | 0.996002 | 0.996002 | 0.996002 |
| 800 | 1.000000 | 1.000000 | 1.000000 |
| 900 | 0.998001 | 0.998001 | 0.998001 |
| 1000 | 0.998182 | 0.998182 | 0.998182 |

## 5. CONCLUSIONS

Accuracy and efficiency are crucial factors in classification tasks in data mining. Association classification gets higher accuracy than some traditional rule-based classification in most cases. However, it generates a large number of rules. Thus, the efficiency of association classification is not high when the minimum support is set to be low. Although traditional rule-based classification methods can achieve high efficiency, some traditional rule-based classification methods usually generate few rules. As a result, some traditional rule-based classification methods may not achieve high in some cases. In this paper, a new approach CMER is proposed. CMER constructs the candidate set and the seed set. They are both consisted of important literals. CMER connects the seed set with the candidate set to produce more high quality rules at a time. CMER generates more classification rules than FOIL, especially when the data set is small. Moreover, in order to improve quality of rules, we use the minimum support and foil gain to update the seed set. Our experimental results show that techniques developed in this paper is feasible. Our experimental results also show that CMER achieves higher accuracy than FOIL.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

**ACKNOWLEDGEMENT**

This work is funded by China NSF program (No.61170129) and the postgraduate education innovation base for Computer application technology, Signal and Information Processing of Fujian province (No. [2008]114, high education of Fujian).

**REFRENCES:**

[1] Liu, B., W. Hsu and Y. Ma. "Integrating Classification and Association Rule Mining", in National Conferences on Aritificial Intelligence 1998-1999, 1998.

[2] Li, W., J. Han and J. Pei, CMAR: "accurate and efficient classification based on multiple class-association rules", in 2001 IEEEE International Conference on Data Mining, 2001, San Jose, California, USA.

[3] X. Yin, J.H., "CPAR:Classification based on Predictive Association Rules", Proceedings of the Third SIAM International Conference on Data Mining, San Francisco, CA, USA, 2003: pp. 208–217.

[4] F. Thabtah, P.C.Y.P., "A New Multi-Class, Multi-Label Associative Classification Approach", Proceedings of the 4th IEEE International Conference on Data Mining (ICDM'04), Brighton, UK, 2004, pp. 217–224.

[5] F. Thabtah, P.C.Y.P., "MCAR multi-class classiffication based on association rule", Proceedings of the 3rd IEEE International Conference on Computer Systems and Applications, Cairo, Egypt, 2005, pp. 1-7.

[6] Y. Z. Liu, Y.C.J., "CSMC-A combination strategy for multi-class classification based on multiple association rules", Knowledge-Based Systems Vol.21, No. 8, 2008, pp. 786-793.

[7] THABTAH, F., "A review of associative classification mining", Knowledge Engineering Review, Vol.22, No.1, 2007, pp. 37-65.

[8] A. H. L. Lim, C.S.L., "Processing online analytics with classification and association rule mining", Knowledge-Based Systems, Vol.23, No.3 p2010, pp. 248-255.

[9] Quinlan, J.R., "C4.5: programs for machine learning", 1993: Morgan Kaufmann Publishers Inc. 302.

[10] Cameron-Jones, J.R.Q.A., "FOIL: A midterm report", In Proc, 1993 European Conf. Machine Learning, 1993, pp.3-20.

[11] M. Antonie, O.Z., "An associative classifier based on positive and negative rules". Proceedings of the 9th ACM SIGMOD Workshop on Research Issues in Data, 2004: p. p.64-69.

[12] B. Liu, M. Hu, W. Hsu, "Multi-level organization and summarization of the discovered rules", Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2000), ACM Press, Boston, 2000, pp. 208–217

[13] K. Wang, S. Zhou, Y. He, "Growing decision tree on support-less association rules", Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, Boston, MA, 2000, pp. 265–269

[14] H.Y. Liu, J. Chen, G. Chen, "Mining insightful classification rules directly and efficiently", Proceedings of the 1999 IEEE International Conference on Systems Man and Cybernetics, IEEE Computer Society, Tokyo, 1999, pp. 911–916

[15] K. Wang, S. Zhou, Y. He, "Growing decision tree on support-less associationrules", Proceedings of the 6th ACM SIGKDD International Conference onKnowledge Discovery and Data Mining, ACM Press, Boston, MA, 2000, pp. 265–269