Using Association Rules to Improve Chinese Handwritten Character Recognition

Zhen Lixin, Dai Ruwei
Artificial Intelligence Lab. Institute of Automation Chinese Academy of Science
P.O. Box 2728, Beijing 100080 P. R. China

Abstract — Association rules are useful for determining correlation between attributes of a relation and have applications in marketing, financial and retail sectors. In this paper, we present an approach for combining handwritten character classifiers based on association rules, which reflect the correlation between the classifiers. The experimental results show that the association rules improve the performances of the integrated system significantly. An experimental comparison of two combination schemes (ILM, NISL) [1] is also provided.

I. INTRODUCTION

It is well known that machine recognition of handwritten Chinese character is a very difficult problem and regarded as one of the ultimate goals of character recognition research [2]. The main difficulties of handwritten Chinese character recognition are so many categories (classes) of Chinese characters, according to National Standard GB2312-80, more than 6000 commonly used Chinese characters, distortion of handwriting and similitude of Chinese characters. To achieve this complex pattern recognition task, designing a good single classifier is a tough problem [3]. The approach of classifier combination is widely used. Many combination algorithms have been suggested e.g. plurality, voting, Bonda count, averaging [4], Integration based on a Linear Model (ILM), and Network Integration based on Supervised Learning (NISL) [1], etc. The basic idea of classifier combination is to harness the compliment information of different classifiers.

Association rules are useful for determining the correlation between attributes of a relation. To combine the outputs of the individual classifiers, we mine the correlation between the outputs of the individual classifiers, expressed as association rules. These association rules served as decision rules are used to determine which output of the individual classifiers ought to be the result. In the system, the outputs of the individual classifiers are regarded as attributes of a relation or a transaction. Furthermore, the results of experiments, and the comparison with the ILM and NISL are presented.

Section 2 briefly introduces the three individual classifiers and the construction of the transaction database. Section 3 gives the procedures of mining and pruning association rules. Section 4 presents the combination approach based on association rules. Section 5 provides experimental results and discussion. The final section offers concluding remarks.

II. CONSTRUCTION OF TRANSACTION DATABASE

The original database consisting of the first level Chinese characters (3755 class), is a set of binary character images. For each class, 200 samples are stored. The image database is divided into training set and testing set, containing 180 samples per class and 20 samples, respectively. Three classifiers, C-HPS, C-FCS and C-LSD [1], are designed to extract features and provide ranking information for every import sample.

A Individual Classifiers

Taking account of the large variation of handwriting, which is probably the largest difficulty of handwritten character recognition, three groups of features, which are insensitive to writing variations, are selected and inputted into three different classifiers, respectively. These classifiers are in fact template-matching machines with similar architecture. Their function is to compare the features of the input character (image) with those of possible candidate templates, and output the candidates in descending order (as shown in Fig. 1) with respect to similarity between input features and those of the reference templates.

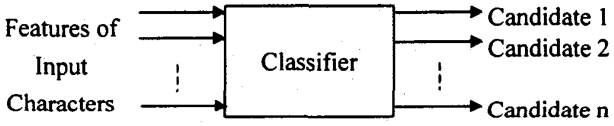

Fig. 1 Architecture of individual classifier

The classifier based on Hierarchical Peripheral Structures (C-HPS) extracts the features of the outermost and second outermost peripheral structures of the input image. The feature vector is 320 dimensions. C-FCS is the classifier based on Four Corner Structures represented by a feature vector with 120 dimensions. C-LSD is our third classifier based on Local Stroke Direction represented by a 256-dimension vector. For detailed description reference to [1].

B Transaction Database

The classes of the 3755 Chinese characters are represented by consecutive natural number. The category label ranges from 1 to 3755. The first candidate of the three individual classifiers is picked out to construct a transaction or a relation. A transaction should contain the correct class label of the input character. Table 1 illustrates some of transactions in generated database.

TID is a unique identifier of each transaction. C-1, C-2 and C-3 are the pair of classifier identifier (CID), given the value of 1, 2 or 3, and the first candidate generated correspondingly by one of three individual classifiers, ranging from 1 to 3755, i.e. <CID, the first candidate>. LS is

0-7803-5995-X/00/$10.00 ©2000 IEEE.

Table 1 Examples of transaction database

TID	C-1	C-2	C-3	LS
t_1	{1, 10}	{2, 10}	{3, 10}	{10}
t_2	{1, 345}	{2,722 }	{3, 2497}	{722}
t_3	{1, 2342}	{2, 1353}	{3, 2342}	{2342}
t_4	{1, 3264}	{2, 11}	{3, 257}	{3343}

the correct category label of input sample.

III. MINING ASSOCIATION RULES

Association rule is an important problem in the KDD (Knowledge Discovery in Database) community. Given a set of transactions *T*, and a set of items *I*, an association rule is an implication of the form $X \Rightarrow Y$, where X and Y are sets of items, and $X, Y \subseteq I$, $X \cap Y = \emptyset$. The *support* of the rule $X \Rightarrow Y$ is the percentage of transactions in *T* that contain both X and Y. The *confidence* of the rule $X \Rightarrow Y$ is c%, if c% of transactions in *T* that satisfy X also satisfy Y [5]. For example, 90% of customs who bought desk also bought chair (desk ⇒ chair (15% 90%)), where 90% is the confidence level of the rule and 15% is the support level of the rule indicating how frequent the rule holds.

The problem of mining association rules is to discover all association rules, in the given transactions T (the database), which have support and confidence greater than the user-specified minimum support (called minsup) and minimum confidence (called minconf). An association rule mining works in two steps:

1. Generate all large itemsets (which satisfy the *minsup*, i.e. the percentage of the transactions contained the itemset is greater than the *minsup*).
2. Generate all association rules that satisfy *minconf* according to the large itemsets.

A large itemset is an itemset that has transaction support above *minsup*. Association rule mining has been researched extensively, e.g. [5], [6], [7].

A Mining Association Rules

Many researchers have designed considerable algorithms for mining association rules. Mining association rules from the above transaction database *T* generated by individual classifiers belongs to typical association rule problem. In the paper, every different pair of <CID, the first candidate> is regarded as an unique item, which can be denoted by C-j-k, where $j \in \{1, 2, 3\}$, $k \in \{1, 2, ..., 3755\}$. And every different class label with respect to the input character is also mapped into different item denoted by LS-n, where $n \in \{1, 2, ..., 3755\}$. Therefore, the transaction database *T* contains 15020 (3755×4) items, the itemset:

I = {C-1-1, C-1-2, ..., C-3-3755, LS-1, LS-2, ..., LS-3755},

And *T* has 751000 (3755×200) transactions, i.e. $T = \{t_1, t_2, ..., t_m\}$, *m* = 751000. Each transaction consists of four items, for example, the transaction t_2 is represented as t_2 = {C-1-345, C-2-722, C-3-2497, LS-722}. The mining algorithm is constructed based on the algorithm Apriori [5]. The following will address the special problems in our system.

There will be a lot of association rules discovered by the mining algorithm. Considering the goal of predicting the class of the input character, many of these rules would be useless, for instance, the rule:

R1: (C-1-1092) Λ (C-3-2008) ⇒ (C-2-3261), (90.5%, 19)

States that 90.5% input characters that are regarded as No. 1092 character by classifier-1, according to the first candidate, and regarded as No. 2008 character by classifier-3, must be considered as No. 3261 character by classifier-2, and that 19 input characters actually satisfy the rule. The rule R1 is helpless for determining the input character's class. To constrain the algorithm to mine interesting rules a rule template is provided:

Rule Template: (C-1-a) Λ (C-2-b) Λ (C-3-c) ⇒ (LS-d), (c%, s).

The antecedent of discovered association rules must be one of the attributes C-1, C-2 and C-3, or a combination of them. The consequent of discovered association rules must be the attribute LS. The c% and s are *confidence* and *support* respectively.

Mining association rules from such a large database is still a tough task. In fact, another type of regularity is unnecessary:

R2: (C-1-334) Λ (C-2-334) Λ (C-3-334) ⇒ (LS-334), (98.1%, 52)

R3: (C-1-334) Λ (C-2-334) ⇒ (LS-334), (96.7%, 149)

Statistical analysis shows that while the plurality of the first candidates generated by three individual classifiers are same, say c^*, the input character's label is just the voting result c^* at the accuracy 95.2%. We are satisfied with the classification accuracy. The rules, like R2 and R3, can be embodied by the following voting rule.

Rv (Voting Rule): The combined decision is the class that gets more votes than any other class.

On the other hand, transactions satisfying the voting rules have no more information worthy to be mined. Therefore, the transactions, which satisfy the voting rule, can be removed from the transaction set *T* without losing important information. The transaction set *T* is trimmed tremendously. The minimum support (*minsup*) and minimum confidence

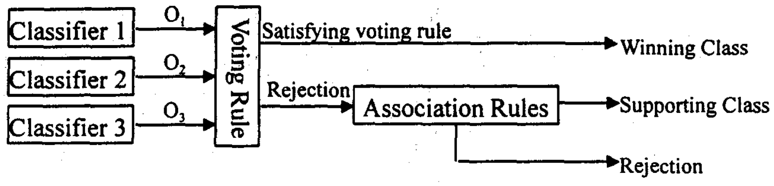

Fig. 2 Combination system model

(*minconf*), which ensure that the discovered rules have enough positive evidence, are selected empirically. Different *minsup* and *minconf* cause different number of rules discovered. If *minsup* and *minconf* are too low, samples can be overfitted. And if *minsup* and *minconf* are too high, a lot of information can be lost. In our system, *minsup* = 8 (cases), *minconf* = 75%.

B Pruning Association Rules

One of the fundamental problems in data mining is how to manage and use the discovered knowledge. Though the mining algorithm is constrained by the rule template and the transaction database is trimmed, the number of discovered association rule is still large. And discovered rules often are redundant; i.e. several rules describe the same transaction. For example:

R4: (C-1-3253) $\wedge$ (C-2-334) $\Rightarrow$ (LS-334), (88.9%, 16)

R5: (C-1-3253) $\wedge$ (C-2-334) $\wedge$ (C-3-2030) $\Rightarrow$ (LS-334), (85.7%, 6)

The rule R5 is more special than the rule R4, and has no additional predictive power over it. Every transaction described by R5 is also described by R4. The rule R5 can be pruned as redundant, though some information may be lost [6]. Some algorithm of pruning association rules are proposed [6], [7]. We adopt the algorithm RuleCover provided in [6] to pruning rules discovered from the transaction database *T*.

Assume Γ is a collection of rules with the same item (or itemset) as the consequent: $\Gamma = \{X_j \Rightarrow Y \mid j = 1, \ldots, n\}$, and Δ is a subset of Γ. The RuleCover algorithm is designed to find the minimum rule subset $\Delta \subseteq \Gamma$ such that Δ describes all transactions that the rule subset Γ describes.

The algorithm prunes the rules significantly, e.g. 27 rules, corresponding to No. 334 character, are discovered, and 8 rules remain after pruning. Totally, 4023 rules are mined out, and 1387 rules remain after pruning.

IV. COMBINATION SCHEME

Our combination model is mainly based on the following empirical knowledge:

1. For different classes, an individual classifier exhibits different classification abilities. Certain classifiers than others can distinguish some classes better. The classification abilities of classifiers are unbalanced between the input samples. Different classifiers complement each other to some extent.
2. The sequence of the classifiers' outputs implicitly encodes a certain amount of complementary information. Voting rules express explicit complementarity, while association rules describe implicit information.

The objective of combination is to obtain the final decision by synthesizing the outputs of the module classifiers. The structure of combination system is shown in Fig. 2.

In Fig. 2, the three individual classifiers output the first candidate labels respectively, which form vector λ, $\lambda = (O_1, O_2, O_3)$. The Voting Rules Module matches the vector λ with the voting rule Rv, and output the winning class if Rv matched, otherwise, put the vector λ into the Association Rule Module. The Association Rule Module searches the rule set Γ for the rule that support the vector λ. The supporting class is exported, if the vector λ is supported by a rule $Rk \in \Delta$, otherwise, rejection is the output.

While the vector λ is matched with the rules in the pruned rule set, say Δ, conflicts may occur. To solve the problem, we sort the rules in Δ in descending order with respect to the *support*. The rules with same *support* are sorted in descending order according to the *confidence*.

V. EXPERIMENT

The original character database is divided into training set (contains 180 characters per class) and testing set (contains 20 characters per class). Training set and testingset generate the training transaction database and the testing transaction database respectively. Three individual classifiers, introduced in section 2, are employed. Experimental results are shown in Table 2 and Table 3.

Table 2 Recognition rates of classifiers

Classifier	C-HPS	C-FCS	C-LSD
Rec. Rate (%)	83.4	80.1	78.9

In order to compare the performance of proposed combination scheme, the above three classifiers are also combined based on ILM, NISL [1] and voting, respectively. Combination system based on voting is a part of the system shown in Fig. 2. Comparing with the voting system, it is obvious that association rules play a concrete role for improving recognition rate. The experimental results show

Table 3 Recongnition rates of combination systems

Combination system	ILM	NISL	Voting	Rules
Rec. Rate (%)	89.9	90.6	85.3	89.5

that the performance of the combination system based on association rules is conspicuously better than any individual classifier, and almost equivalent to that of ILM and NISL.

VI. CONCLUSION

In this paper, we have presented the idea of combining classifiers based on association rules as a new method of combining multiple modules. Combining classifiers based on association rules can be considered as based on knowledge. The combination system harnesses the correlation of classifier modules explicitly, different from traditional methods. And we believe that it is a promising strategy to cope with some other complex decision making problems.

VII. REFERENCE

[1] Hong Wei Hao, Xu Hong Xiao and Ru Wei Dai, "Handwritten Chinese Character Recognition by Metasynthetic Approach." Pattern Recognition, Vol. 30, No. 8, p1321-1328, 1997.

[2] S. Mori, K. Yamamoto and M. Yasuda, "Research on machine recognition of handprinted characters." IEEE Trans. Pattern Analysis Mach. Intell. PAMI-6 (4), p386-450, 1984.

[3] Nei Kato, Masato Suzuki, Shinichiro Omachi, Hirotomo Aso and Yoshiaki Nemoto, "A Handwitten Character Recognition System Using Directional Element Feature and Asymmetric Mahalanobis Distance." IEEE Trans Pattern Analysis Mach. Intell. PAMI 21 (3), 1999.

[4] Khaled Al-Ghoneim and B. V. K. Vijaya Kumar, "Unified Decision Combination Framework." Pattern Recognition, Vol. 31, No. 12, p2077-2089, 1998.

[5] R. Agrawal and R. Srikant, "Fast Algorithm for mining Association Rules." In Proceedings of 20^{th} international Conference on Very Large Databases Santiago, Chile, 1994.

[6] H Toivonen, M Klemettinen, P Ronkainen, K Hatonen and H Mannila, "Pruning and Grouping Discovered Association Rules." In ECML-95, Workshop on Statistics, Machine Learning, and Knowledge Discovery in Databases, 47-52, Greece, April 1995.

[7] Bing Liu, Wynne Hsu and Yiming Ma, "Pruning and Summarizing the Discovered Associations", KDD-99, San Diego, August 1999.