Knowledge Discovery from Health Data Using Weighted Aggregation Classifiers

Toru Takae[1], Minoru Chikamune[1⋆], Hiroki Arimura[1], Ayumi Shinohara[1], Hitoshi Inoue[2], Shun-ichi Takeya[2⋆⋆], Keiko Uezono[3], and Terukazu Kawasaki[3]

[1] Dept. Informatics, Kyushu Univ., Hakozaki 6-10-1, Fukuoka 812–8581, Japan
{takae,chika,arim,ayumi}@i.kyushu-u.ac.jp
[2] Univ. Comput. Center, Kyushu Univ., Hakozaki 6-10-1, Fukuoka 812–8581, Japan
[3] Inst. Health Science, Kyushu Univ., Kasuga-koen 6-1, Kasuga 816–8580, Japan

Introduction. The automatic construction of *classifiers* is an important research problem in data mining, since it provides not only a good prediction but provides also a characterization of a given data in the form easily understood by a human. A *decision tree* [4] is a classifier widely used in real applications, which are easy to understand, and efficiently constructed by using a method based on entropy heuristics [4]. Fukuda *et al.* [1] have proposed an efficient algorithm (called DT in this abstract) for constructing a small and accurate decision tree with numeric attributes using optimized two-dimensional numeric association rules as node labels.

A problem is that at each node, DT generates many rules for possible pairs of numeric and ordered attributes, but selects only one optimized rule among them. Since this generation is time consuming, the construction may be inefficient when there are many numeric and ordered attributes. A possible approach is to build a one-level decision tree such as 1R [2]. We take another approach to aggregate the decisions made by all generated rules.

Weighted Aggregation Classifiers. In this abstract, we introduce weighted aggregation classifiers, which can be efficiently constructed as one-level decision trees but can provide highly accurate classification. Suppose that we have a set of all rules r_i, $i = 1, \ldots, k$, generated from a dataset, which are associated with the parameter c_i^1 (c_i^0) denoting the conditional probability of the target attribute is true given r_i is true (false) on an instance $\mathbf{x}$. Then, an *weighted aggregation classifier* (WA) is a collection $H = \{(r_i, c_i^1, c_i^0, w_i) \mid i = 1, \ldots, k\}$ of quadruples, where we associate with each rule r_i in H a real weight w_i representing its classification accuracy so that an accurate rule has a large weight. The decision $H(\mathbf{x})$ is made by the majority vote over the classifiers:

$$H(\mathbf{x}) = \left[\sum_i w_i \cdot conf_i(\mathbf{x}) \geq \sum_i w_i \cdot (1 - conf_i(\mathbf{x})) \right],$$

⋆ Presently, working for Mitsubishi Electronic.
⋆⋆ Presently working at Admission Center, Kyushu University.

S. Arikawa, K. Furukawa (Eds.): DS'99, LNAI 1721, pp. 359–361, 1999.

Table 1. The results of prediction on UCI repository [3] Acc and Time are in % and seconds, resp, and an underlined entry shows the winner.

Dataset	Size	Acc_{base}	WA	Acc	Time	DT	Acc	Time
Breast Canser	699	65.52		97.51	237		95.61	666
Liver Disorder	345	57.97		67.83	51		50.46	202
Pima Diabetes	769	65.10		69.40	234		69.92	1216
Balance Scale	625	53.92		85.59	31		79.36	106
Titanic	2201	67.70		77.60	0.5		79.05	2

Table 2. The results of High SBP prediction problem. Accuracies are measured in %.

Attributes	Acc_{base}	WA	Acc_{avr}	Acc_{max}	DT	Acc_{avr}	Acc_{max}
BMI	60.00		59.51	63.42		60.11	63.93
BMI+Others	60.00		70.71	74.13		62.92	70.09

where $conf_i(\mathbf{x})$ is c_i^1 if $\mathbf{x}$ satisfies r_i, c_i^0 otherwise, and $[P]$ is the characteristic function of a predicate P. In the experiments, we set $w_i = \max_i\{Ent(r_i)\} - Ent(r_i)$, where $Ent(r_i)$ is the entropy of r_i on a dataset S. If there is only one rule r_1 then we set $w_1 = 1$. Since WA makes a decision considering more than one rules, WA would perform better than 1R [2] when several attributes interact.

Experiments. We implemented and compared WA and DT experimentally as follows. First, we run experiments on five benchmark datasets from UCI repository [3]. The accuracies are evaluated by two-fold cross validation, and timing are taken on Solaris 2.6, Ultra Sparc IIi 300MHz. In Table 1, we observe that WA produced more accurate classifiers in shorter time than DT does.

Secondly, we run experiments on a large real dataset consisting of health condition records of around 300,000 Japanese national university students, which was obtained by a nationwide health science survey in 1995 conducted by the CSSH [5]. The task here is to predict an abnormality on Systolic Blood Pressure (SBP), called *High SBP* (SBP $\geq$ 140 mmHg), from Body Mass Index (BMI) and 13 ordered attributes on student's life-style attributes, which is a prevalent analysis in health science research. The training and the test sets consist of randomly chosen 75 and 3653 records, resp. The base line accuracy is Acc_{base} =60 %, and average Acc_{avr} and maximum Acc_{max} are taken through 100 trials.

In Table 2, we show the results of the prediction by BMI alone (the upper row) and the prediction by BMI and all life-style attributes (the lower row). In the case with BMI alone, both algorithms produce only trivial classifiers. In the case with BMI and life-style attributes, we can observe that in average WA produces nontrivial classifiers with accuracy 70.71% while DT produces classifiers with accuracy only 62.92%, which is only slightly above the baseline.

References

1. T. Fukuda, Y. Morimoto, S. Morishita, and T. Tokuyama. Constructing efficient decision trees by using optimized numeric association rules, In Proc. the 22th VLDB Conference, 146–155, 1996.
2. R. C. Holte. Very simple classification rules perform well on most commonly used datasets, Machine Learning, 11, 63–91, 1993.
3. P. M. Murphy and D. W. Aha. UCI repository of machine learning databases, 1994. `http://www.ics.uci.edu/mlearn/MLRepository.html`.
4. J. R. Quinlan. C4.5: Programs for machine learning, Morgan Kaufmann, 1993.
5. The Committee for Statistics of Student Health (CSSH) in National University. The data of health condition in students of national universities, 1997.