

Using an Adapted Classification Based on Associations Algorithm in an Activity-Based Transportation System

Davy Janssens, Geert Wets, Tom Brijs and Koen Vanhoof

Limburgs Universitair Centrum, Transportation Research
Institute, Universitaire Campus, gebouw D, B-3590 Diepenbeek, Belgium

Email: {davy.janssens; geert.wets; tom.brijs; koen.vanhoof}@luc.ac.be
Tel: +32(0)11 26 -- -- {86 43; 86 49; 86 33; 86 08}

Abstract

A lot of research has been carried out in the past by using association rules to build more accurate classifiers. The idea behind these integrated approaches is to focus on a limited subset of association rules. However, these integration approaches have not been tested yet within the context of transportation research. The aim of this chapter is therefore to evaluate the performance of an adapted well-known associative classification algorithm on the datasets that are used in the **Albatross** transportation modelling system. The presented work is an extension of previous research efforts in the sense that it now becomes possible to use the adapted CBA system for multi-class problems. Experiments showed that the original CBA system achieved the best average performance for the three classifiers under evaluation. While the adapted CBA still generated better average results than CHAID, the performance with respect to original CBA was slightly worse.

1 Introduction

Over the last decade, activity-based transportation models have received increased attention for modelling travel demand. The most important characteristic in these models is that travel demand is derived from the

activities that individuals and households need or wish to perform. The main advantage is that travel has no longer an isolated existence in these models, but is perceived as a way to perform activities and to realize particular goals in life.

One of the most advanced operational activity-based transportation models is the **Albatross** system (A Learning Based Transportation Oriented Simulation System), developed by Arentze and Timmermans (2000) for the Dutch Ministry of Transport. **Albatross** is a multi-agent rule-based system that predicts which activities are conducted where, when, for how long, with whom and the transport mode involved. It uses decision rules to predict each of those facets (where, when, etc.) and to support scheduling decisions. These decision rules were originally derived by means of a decision tree induction algorithm that is based on a chi-squared test for determining the subsequent splits in the tree (hereafter referred to as CHAID; Kass, 1980). Since the choice of the learning algorithm is one of the most important factors to determine the overall performance of a transportation model, several studies have been conducted in the past to measure the performance of different algorithms. For instance, in comparative studies by Wets, *et al.* (2000) evidence was found that different kinds of decision tree induction algorithms achieve comparable results; while improvements were found with respect to the reduction in complexity of the decision model in Moons *et al.* (2004). A previous study by Janssens, *et al.* (2004) suggested that Bayesian networks achieved better performance than decision trees and that they are better suited to capture the complexity of the underlying decision-making process. While the studies by Wets *et al.* and Moons *et al.* mainly focus on classification based (supervised) learning, the work by Janssens *et al.* is more an association based learning (unsupervised) approach. However, in recent years, extensive research has been carried out to integrate supervised and unsupervised learning (also referred to as associative classification). Despite the solid performance of these integration approaches in previous research efforts, the comprehensiveness and complexity of dealing with the often-large number of association rules have lead to difficulties and (accuracy versus generality) trade-off questions, which are part of a lot of research that is currently going on. In addition to this, it is important to state that these integration approaches have not yet been tested within the context of transportation research. The aim of this chapter is therefore to evaluate the performance of a well-known associative classification algorithm, as well as an adaptation of this algorithm within the context of the **Albatross** transportation model.

As mentioned above, association and classification rules are the two main learning algorithms in associative classification. The study of association rules is focused on using exhaustive search to find all rules in data that satisfy user-specified minimum support and minimum confidence criteria. On the other hand, classification rules aim to discover a small set of rules to form an accurate classifier. Given a set of cases with class labels as a training set, the aim of classification is to build a model (called classifier) to predict future data objects for which the class label is unknown. Now, by focusing on a limited subset of association rules, i.e. those rules where the consequent of the rule is restricted to the classification class attribute, it is possible to build classifiers. Several publications (for instance Wang *et al.* 2000; Chen *et al.* 1999; Dong *et al.* 1999 and others) have shown that associative classification in general generates at least equal accuracy than state-of-the-art classification algorithms such as C4.5. (Quinlan 1993) The reasons for the good performance are obvious. Association rules will search globally for all rules that satisfy minimum support and minimum confidence norms. They will therefore contain the full set of rules, which may incorporate important information. The richness of the rules gives this technique the potential of reflecting the true classification structure in the data (Wang *et al.* 2000). However, the comprehensiveness and complexity of dealing with the often-large number of association rules have led to difficulties and (accuracy versus generality) trade-off questions. Contributions to tackle a number of these difficulties can be found in Dong *et al.* (1999), Li *et al.* (2001) and in Wang *et al.* (2000). Liu, Ma and Wong (2001) proposed an improvement of their original CBA (classification based on associations)-system (Liu *et al.* 1998) in (Liu *et al.* 2001) to cope with the weaknesses in the system. In spite of the fact that the presented adaptations of CBA are valuable, some important issues still remain unsolved. Some unsolved issues were already tested and implemented with success within the context of the 16 UCI Repository of Machine Learning binary class databases (Janssens *et al.* 2004). In this work, it was found that the adapted CBA algorithm has the lowest average error rate in comparison with the four other classifiers under comparison (i.e. original CBA system, C4.5., C4.5. with discretized values and Naïve Bayes). The aim of this chapter is to extend this adaptation towards multi-class problems and to test the performance of the adapted CBA algorithm within the context of the datasets that are used in Albatross.

The remainder of this chapter is organized as follows. First, the conceptual framework underlying the Albatross-system is briefly discussed in order to provide some background information with respect to this transportation model. Section 3 explains the original CBA-algorithm (Liu *et al.* 1998).

Weaknesses and adaptations of the original system are suggested in section 4. Section 5 then describes the design of the experiments that were carried out to validate the new approach and gives an overview of the data that were used. A discussion of the empirical results is also given in this section. Finally, conclusions and recommendations for further research are presented in section 6.

2 The Albatross System

The Albatross system (Arentze and Timmermans 2000) is a computational process model that relies on a set of decision rules to predict activity-travel patterns. Rules are typically extracted from activity diary data. The activity scheduling agent of Albatross is the core of the system which controls the scheduling processes in terms of a sequence of steps. These steps are based on an assumed sequential execution of decision tables to predict activity-travel patterns (see Figure 1). The first step in the figure involves for each person decisions about which activities to select, with whom the activity is conducted and the duration of the activity. The order in which (the non-work) activities are evaluated is pre-defined as: daily shopping, services, non-daily shopping, social and leisure activities.

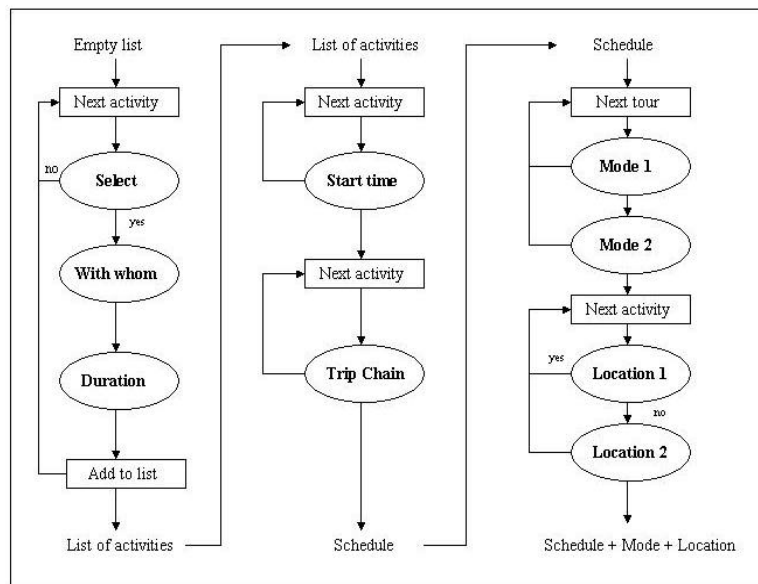


Figure 1: Albatross' scheduling engine (Arentze, Timmermans, 2000)

The assignment of a scheduling position to each selected activity is the result of the next two steps. After a start time interval is selected for an activity, trip-chaining decisions determine for each activity whether the activity has to be connected with a previous and/or next activity. Those trip chaining decisions are not only important for timing activities but also for organizing trips into tours. The next steps involve the choice of transport mode for work (referred to as mode1), the choice of transport mode for other purposes (referred to as mode2) and the choice of location. Possible interactions between different facets of the transportation model, such as for instance between mode and location choices are taken into account by using location information as conditions of mode selection rules. As mentioned before, each decision in the Albatross system (see oval boxes of Figure 1) is extracted from activity travel diary data using the CHAID-based (Kass 1980) algorithm.

3 Classification Based on Associations

Before elaborating on the improvements that can be made to CBA, an overview of the original algorithm will be provided. First, a definition of association rules is given. Hereafter, class association rules (CARs) are introduced.

3.1 Association Rules

Let $I = \{i_1, i_2, \dots, i_k\}$ be a set of literals, called items. Let D be a set of transactions, where each transaction T is a set of items such that $T \subseteq I$. We say that a transaction T contains X , a set of items in I , if $X \subseteq T$. An association rule is an implication of the form $X \Rightarrow Y$, where $X \subset I$, $Y \subset I$ and $X \cap Y = \emptyset$. The rule $X \Rightarrow Y$ holds in the transaction set D with confidence c if $c\%$ of transactions in D that contain X also contain Y . The rule $X \Rightarrow Y$ has support s in the transaction set D if $s\%$ of transactions in D contain $X \cup Y$. Given a set of transactions D , the problem of mining association rules is to generate all association rules that have support and confidence greater than a user-specified minimum support (minsup) and minimum confidence (minconf) (Agrawal *et al.* 1993).

3.2 Class Association Rules

To make association rules suitable for the classification task, the CBA method focuses on a special subset of association rules, i.e. those rules with a consequent limited to class label values only; or so-called class association rules (CARs). Thus, only rules of the form $A \Rightarrow c_i$ where c_i is a possible class, need to be generated by the CBA-algorithm. Therefore, the Apriori algorithm which is widely used for generating association rules, was modified to build the CARs. Details about these modifications can be found in (Liu *et al.* 1998).

To reduce the number of rules generated, the algorithm performs two types of pruning. The first type is the pessimistic error rate used in (Quinlan *et al.* 1993). The second type of pruning is known as database coverage pruning (Li *et al.* 2001). Building a classifier in CBA is therefore also largely based on this coverage pruning method, which is applied after all the CARs have been generated. The original algorithm, which is used in CBA, is shown in figure 2. Before the pruning, the algorithm will first rank all the CARs and sort them in the descending order. As it will be shown in the next section, this rank will be subject to one of the modifications that were implemented. The ranking is as follows: given two rules r_i and r_j , $r_i > r_j$ (or r_i is said having higher rank than r_j), if (1) $\text{conf}(r_i) > \text{conf}(r_j)$; or (2) $\text{conf}(r_i) = \text{conf}(r_j)$, but $\text{sup}(r_i) > \text{sup}(r_j)$; or (3) $\text{conf}(r_i) = \text{conf}(r_j)$ and $\text{sup}(r_i) = \text{sup}(r_j)$, but r_i is generated before r_j .

```
R=sort(R);
for each rule r in R in sequence do
    temp =  $\emptyset$ ;
    for each case d in D do
        if d satisfies the conditions of r then store d.id in temp and mark r if it
        correctly classifies d;
    end
    if r is marked then
        insert r at the end of C;
        delete all the cases with the ids in temp from D;
        selecting a default class for the current C;
        compute the total number of errors of C;
    end
end
```

Figure 2: Building a classifier in CBA (Liu *et al.*, 1998)

If at least one case among all the cases covered by the rule is classified correctly by the rule, the rule is inserted into the classifier by following this sorted descending sequence order and all the cases it covers are removed from the database. The rule insertion stops when either all of the rules are used or no cases are left in the database. The majority class among all cases left in the database is selected as the default class. The default class is used in case when there are no covering rules. Then, the algorithm computes the total number of errors, which is the sum of the number of errors that have been made by the selected rules in the current classifier and the number of errors made by the default class in the training data.

After this process, the first rule that has the least number of errors is identified as the cutoff rule. All the rules after this rule are not included in the final classifier since they will only produce more errors (Liu *et al.* 1998).

4 Identifying Weaknesses and Proposing Adaptations to CBA

4.1 Limits of conditional probability (confidence)

A profound examination of the algorithm identified a potential weakness in the way the rules are sorted. Since rules are inserted in the classifier following the sorted confidence order, this will determine to a large extent the accuracy of our final classifier. Confidence is a good measure for the quality of (class) association rules but it also suffers from certain weaknesses. The aim of this section is to elaborate on them.

The first weakness is that the conditional probability of a rule $X \Rightarrow Y$ is invariable when the size of $s(Y)$ or D varies. The subset of the cases which are covered by the consequent of the rule is given by $s(Y)$, while D is the total number of observations in the dataset. The confidence property is also insensitive to cardinal dilatation (i.e. the size of the subsets increases in the same proportion). Figure 3 graphically displays the problem. It can be seen from the figure that the three cases have the same confidence (indicated by the intersections between the ovals). Nevertheless, the rule $X \Rightarrow Y$ is more likely to happen when the size of $s(Y)$ increases or when the size of D

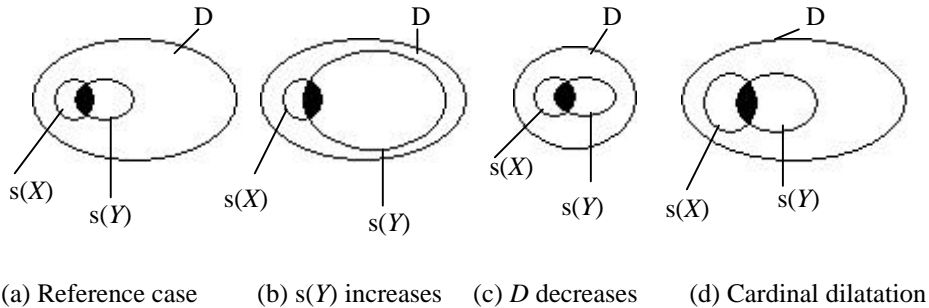


Figure 3: Three cases with constant conditional probability

decreases; and furthermore, the implication will be more meaningful when the size of all the sets grows in the same proportion. Figure 3(b) shows that it is not surprising that, when $s(Y)$ is close to the size of D , the observations which are covered by the antecedent X of the rule, are also included in $s(Y)$. In figure 3(c) it is evident that when the size of $s(X)$ and the size of $s(Y)$ are close to that of D , $s(X)$ and $s(Y)$ share many objects. And also in the final case (figure 3(d)), a more important sample will allow to be more confident in the statistical implication $X \Rightarrow Y$.

The second drawback of the use of conditional probability is that when for a particular class, the minsup parameter is set to 1% or even lower, it might very well happen that some rules have a high confidence parameter but on the other hand they might be confirmed by a very limited number of instances, and that those rules stem from noise only. This is why it is always dangerous to look for implications with small support even though these rules might look very “interesting”. This danger seems to exist all the more in CBA because the application which implements the algorithm even offers a possibility to include rules with high confidence that do not satisfy the minimum support threshold in the final classifier. As a result, choosing the most confident rules may not always be the best selection criterion. Therefore, a suitable measure which takes both drawbacks that were identified above into account, was used to perform the sorting. The next section elaborates on this.

4.2 Intensity of implication

Intensity of implication, introduced by Gras & Lahrer (1993) measures the distance to random choices of small, even non statistically significant, subsets. In other words, it measures the statistical surprise of having so few examples on a rule as compared with a random draw (Guillaume *et al.* 1998). Consider a database D , where $|D|$ is the total number of observations in the database, and an association rule $X \Rightarrow Y$. Now, let U and V be two sets randomly chosen from D with the same cardinality as X and Y respectively, i.e., $s(X)=s(U)$ and $s(Y)=s(V)$, and let $\neg Y$ mean ‘not Y ’ as shown in figure 4.

Let $s(U \wedge \neg V)$ be the random variable that measures the expected number of random negative examples under the assumption that U and V are independent, and $s(X \wedge \neg Y)$ the number of negative examples observed on the rule. Now, if $s(X \wedge \neg Y)$ is unusually small compared with $s(U \wedge \neg V)$, the one we would expect at random, then we say that the rule $X \Rightarrow Y$ has a strong statistical implication. In other words, the intensity of implication for a rule $X \Rightarrow Y$ is stronger, if the quantity $P[s(U \wedge \neg V) \leq s(X \wedge \neg Y)]$ is smaller. Intensity of implication is then defined as $1 - P[s(U \wedge \neg V) \leq s(X \wedge \neg Y)]$. Since the random variable $s(U \wedge \neg V)$ follows the hypergeometric law, which means $\Pr[s(U \wedge \neg V) = k] = \Pr[\text{of } s(U) \text{ examples selected at random, exactly } k \text{ are not in } V] = \frac{C_{s(\neg V)}^k \times C_{s(V)}^{s(U)-k}}{C_{|D|}^{s(U)}}$ and taking into account

that $s(U) = s(X)$ and that $s(V) = s(Y)$, the intensity of implication can be written as:

$$1 - \sum_{k=\max(0, s(X)-s(Y))}^{s(X \wedge \neg Y)} \frac{C_{s(\neg Y)}^k \times C_{s(Y)}^{s(X)-k}}{C_{|D|}^{s(X)}} \quad (1)$$

This formula for intensity of implication is suitable as long as the number of cases in the database, i.e. $|D|$, is reasonably small. Otherwise, the combination numbers in the above formula explode quite soon. Therefore, Suzuki *et al.* (1998) came up with an approximation of this formula for big datasets. They argue that if $s(U \wedge \neg V)$ is small, which is often the case in rule discovery, then Poisson approximations can be applied. In that case, the above formula for intensity of implication reduces to a much simpler version that is easier to compute, i.e.:

$$\begin{aligned}
& 1 - \sum_{k=\max(0, s(X)-s(Y))}^{s(X \wedge \neg Y)} \frac{C_{s(\neg Y)}^k \times C_{s(Y)}^{s(X)-k}}{C_{|D|}^{s(X)}} \\
& \approx 1 - \sum_{k=0}^{s(X \wedge \neg Y)} \frac{I^k}{k!} e^{-I} \tag{2}
\end{aligned}$$

with

$$I \equiv \frac{s(X) \times (|D| - s(Y))}{|D|}$$

where $|D|$ is the number of cases, $s(X)$ is the number of cases covered by the antecedent and $s(Y)$ is the number of cases covered by the consequent of the rule. The coefficient $s(X \wedge Y)$ represents the number of cases that are covered by the antecedent and the consequent of the rule, while $s(X \wedge \neg Y)$ stands for the number of cases which are covered by the antecedent but not by the consequent of the rule. Since confidence and support are standard measures for determining the quality of association rules, it would be suitable if those could be incorporated in (2). This procedure is quite straightforward.

Rewriting $s(X \wedge \neg Y)$ gives:

$$\begin{aligned}
& s(X \wedge \neg Y) = s(X) - s(X \wedge Y) \\
& = s(X \wedge Y) \times \left(\frac{s(X)}{s(X \wedge Y)} - 1 \right) \\
& = \frac{s(X \wedge Y)}{|D|} \times |D| \times \left(\frac{1}{\frac{s(X \wedge Y)}{s(X)}} - 1 \right) \\
& = \text{support} \times \text{cases} \times \left(\frac{1}{\text{confidence}} - 1 \right)
\end{aligned}$$

Rewriting I gives:

$$\begin{aligned}
 I &= \frac{s(X) \times (|D| - s(Y))}{|D|} \\
 &= \frac{\frac{s(X \wedge Y)}{|D|}}{\frac{s(X \wedge Y)}{s(X)}} \times (|D| - s(Y)) \\
 &= \frac{\text{support}}{\text{confidence}} \times (\text{cases} - \text{abssupcons})
 \end{aligned}$$

Substituting both derivations in (2) gives

$$\begin{aligned}
 1 - \sum_{k=0}^{\infty} \left(\frac{1}{\text{confidence}} - 1 \right)^k \left(\frac{\text{support}}{\text{confidence}} \times (\text{cases} - \text{abssupcons}) \right)^k \times \\
 e^{-\left(\frac{\text{support}}{\text{confidence}} \times (\text{cases} - \text{abssupcons}) \right)} \quad (3)
 \end{aligned}$$

By means of formula (3), we are now ready to adapt the CBA algorithm. This is done by using intensity of implication as the primary criteria when doing the sorting work in the first rule of figure 2. Rule r_i has a higher rank than rule r_j if it has a larger value of intensity of implication. When two rules have the same values of intensity of implication, they are ranked according to the sorting mechanism of the original CBA, which is mentioned in section 4.1. Guillaume *et al.* (1998) claim that the relevance of the discovered association rules can be significantly improved by using intensity of implication. The following section examines whether the same conclusion can be found when the adaptation is evaluated within the context of the multi-class datasets that are used for the Albatross modelling system.

5. Empirical Section

5.1 Description of the data

The activity diary data used in this study were collected in 1997 in the municipalities of Hendrik-Ido-Ambacht and Zwijndrecht in the Netherlands (South Rotterdam region) to develop the *Albatross* system (Arentze, Timmermans 2000). The data involve a full activity diary, implying that both in-home and out-of-home activities were reported. The sample covered all seven days of the week, but individual respondents were requested to complete the diaries for two designated consecutive days. Respondents were asked, for each successive activity, to provide information about the nature of the activity, the day, start and end time, the location where the activity took place, the transport mode, the travel time, accompanying individuals and whether the activity was planned or not. A pre-coded scheme was used for activity reporting.

There are some general variables that are used for each choice facet of the *Albatross* model (i.e. each oval box). These include (among others) household and person characteristics that might be relevant for the segmentation of the sample. Each dimension also has its own extensive list of more specific variables, which are not described here in detail.

5.2 Empirical Results

In this section, the CBA algorithm is evaluated on the datasets that are used within the context of the *Albatross* modelling system. To be able to test the validity of the presented models on a holdout sample, only a subset of the cases is used to build the models (i.e., “training set”). The decline in goodness-of-fit between this “training” set and the validation set (“test set”) is taken as an indicator of the degree of overfitting. The purpose of the test set is also to evaluate the predictive ability of the three techniques for a new set of cases. For each decision step, we used a random sample of 75% of the cases to build and optimise the models. The other subset of 25% of the cases were presented as “unseen” data to the models; this part of the data was used as the test set. The accuracy percentages that indicate the predictive performance of the three models on the training and test sets

are presented in Table 1. The nine datasets that are used in *Albatross* are multiple class datasets. The number of classes has been indicated between brackets in Table 1.

Table 1: Accuracy Results

Dataset	CHAID		CBA		Adapted CBA	
	Train (%)	Test (%)	Train (%)	Test (%)	Train (%)	Test (%)
Duration (3)	41.30	38.80	44.71	39.21	40.67	40.88
Location1 (7)	57.50	58.90	66.34	62.72	64.54	68.13
Location2 (6)	35.40	32.60	52.58	41.14	26.78	26.29
Mode for work (4)	64.80	66.70	83.57	73.68	74.71	76.84
Mode other (4)	52.80	49.50	66.53	60.88	54.92	54.82
Selection (2)	72.40	71.60	79.56	78.74	79.07	79.19
Start time (6)	39.80	35.40	34.48	33.68	33.33	33.06
Trip chain (4)	83.30	80.90	83.86	80.40	82.72	81.95
With whom (3)	50.90	48.40	61.14	56.20	54.65	48.07
Average	55.36	53.64	63.64	58.52	56.82	56.58

In order to get a more comprehensive evaluation, results are also compared with the original CHAID algorithm that is used in *Albatross* besides the obvious comparison with original CBA,. When the average results of the three algorithms are compared, it is clear that the CBA classifier produces better results than CHAID. Only with respect to the “start time”, and “trip chain”-datasets, the CBA algorithm performs somewhat worse. Unfortunately, while the adapted CBA algorithm does better than the CHAID decision tree approach, the average (original) CBA accuracy could not be improved. However, in this case, it remains important to notice that the adapted CBA algorithm contains less rules and that also the degree of overfitting is significantly lower. While a significant improvement could be seen on the “location 1”-dataset, the performance on the “location 2”-dataset is significantly worse with respect to the original CBA algorithm. It can be seen from table 1 that the “location 2”-dataset deteriorates to a large extent the general average performance. The reason for this is not clear and is likely attributable towards the nature of the dataset.

Generally speaking, adapted CBA was able to achieve an improvement in performance on the test set for 5 out of the 9 datasets. As mentioned before, in previous work (Janssens *et al.* 2004), it was shown on 16 binary UCI datasets that the performance of adapted CBA achieved an important

improvement with respect to the other classification algorithms under evaluation (CBA, C4.5 and Naïve Bayes). For this reason, the reason for the disagreement in results is not immediately obvious. One plausible assumption is that intensity of implication performs worse when dealing with multi-class problems. In order to evaluate this assumption, 10 additional experiments were carried out on multi-class UCI data. The datasets that were used in these additional tests are Annealing, Automobile, Glass Identification, Iris Plant, Led Display, Lymphography, Vehicle, Waveform, Wine recognition and Zoo (Blake and Merz 1998). The average error rate over the 10 datasets was 18.05% for adapted CBA, while it was 15.54% for original CBA. This finding is of the same magnitude as the result that was found in table 1. Based on this additional external validation, it is fair to say that intensity of implication in its current form performs somewhat worse when dealing with multi-class problems. The fact that Albatross is dealing with real-life (and not artificial) data does not seem to be a differentiator in results.

Finally, it is important to stress that the accuracy versus generality trade-off remarks that were already made in the introductory section of this chapter, needs to be kept in mind in the interpretation of the results. Indeed, if accuracy is the main objective, the original CBA algorithm (with huge number of rules) is likely the best option to choose, while CHAID or adapted CBA will be more favourable choices when the complexity of the classifier is an issue.

6 Conclusions

The research presented in this chapter focused on the integration of supervised and unsupervised learning. In doing so, a modified version of the CBA algorithm, which can be used to build classifiers based on association rules, has been proposed. The modified CBA algorithm was tested within the context of the datasets that are used in Albatross, which is a multi-agent rule-based system that predicts which activities are conducted where, when, for how long, with whom and the transport mode involved. For this reason, CBA needed to be extended for application on multi-class datasets. CBA was adapted by coupling it with intensity of implication, a measure to calculate the distance to random choices of small, even non statistically significant, subsets.

Experiments proved that CBA achieved better average performance than the CHAID decision tree algorithm that is used in Albatross. While adapted CBA still generated better average results than CHAID, the performance with respect to original CBA was slightly worse. However, adapted CBA generated less decision rules than original CBA and also the degree of overfitting was significantly lower. Also, the average result was mainly deteriorated by a significantly worse performance of adapted CBA on one particular dataset. In spite of both arguments, the result was somewhat surprising in comparison with previous good performances of the adapted algorithm (on binary class problems).

For explanatory purposes, tests on 10 additional multi-class datasets were therefore carried out. Due to the fact that a similar finding was found on these datasets, the authors conclude that the implementation of intensity of implication in its current form performs somewhat worse on multi-class problems.

Finally, it needs to be stated that in its current form, it is still possible that rules with a relatively low confidence occasionally may have a high intensity of implication. Therefore, the evaluation of a more advanced sorting mechanism needs to be the subject of additional and future research.

References

Arentze, T.A. and H.J.P. Timmermans (2000), *Albatross: A Learning-Based Transportation Oriented Simulation System*. European Institute of Retailing and Services Studies. Eindhoven, The Netherlands.

Agrawal, R., T. Imielinski and A. Swami (1993), Mining Association Rules between Sets of Items in Large Databases, in: Proc. of the ACM SIGMOD Conference on Management of Data, Washington D.C., pp. 207-216.

Blake C.L. and Merz, C.J. (1998), UCI Repository of machine learning databases[<http://www.ics.uci.edu/~mlearn/MLRepository.html>]. Irvine, CA: University of California, Department of Information and Computer Science.

Chen, J., H.Y.Liu and G.Q. Chen (1999), Mining insightful classification rules directly and efficiently IEEE/SMC'99

Dong, G., Zhang, X., Wong, L., & Li, J. (1999), CAEP: Classification by aggregating emerging patterns, in Proc. of the Second International Conference on Discovery Science, Tokyo, Japan, pp. 30-42

Gras, R. and A. Lahrer (1993), L'implication statistique: une nouvelle méthode d'analyse des données, Mathématiques, Informatique et Sciences Humaines n°120.

Guillaume, S., F. Guillet and J. Philippé (1998), Improving the discovery of association rules with intensity of implication, in Principles of Data Mining and Knowledge Discovery, vol. 1510 of Lecture Notes in Artificial Intelligence, pp. 318-327.

Janssens, D., Wets, G., Brijs, T., Vanhoof, K., Timmermans, H.J.P., Arentze, T.A. (2004), Improving the Performance of a Multi-Agent Rule-Based Model for Activity Pattern Decisions Using Bayesian Networks Forthcoming in Journal of the Transportation Research board, also in: Electronic conference proceedings of the 83rd Annual Meeting of the Transportation Research Board (CD-ROM)

Janssens, D., Lan, Y., Wets G., Chen G. and T. Brijs (2004), Empirically validating an adapted classification based on associations algorithm on UCI data, Applied Computational Intelligence (D. Ruan, P. D'hondt, M. De Cock, M. Nachtegael, E. E. Kerre, eds.), World Scientific, ISBN 981-238-873-7.

Kass, G.V. (1980), An Exploratory Technique for Investigating Large Quantities of Categorical Data, Applied Statistics, Vol.29, 119-127

Li, W., J. Han and J. Pei (2001), CMAR: Accurate and Efficient Classification Based on Multiple Class-Association Rules, in ICDM 2001, San Jose, California, pp. 369-376

Liu, B., W. Hsu and Y. Ma (1998), Integrating Classification and Association Rule Mining, in Proc. of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98), New York, pp. 80-86

Liu, B., Y. Ma and C. Wong (2001), Classification using Association Rules: Weaknesses and Enhancements, To appear in Vipin Kumar, et al, (eds.), Data mining for scientific and engineering applications

Moons, E., G. Wets, M. Aerts, T.A. Arentze and H.J.P. Timmermans, (2004) The Impact of Irrelevant Attributes on the Accuracy of Classifier Systems in Generating Activity Schedules. Conditionally accepted for publication in Environment and Planning A.

Quinlan J.R., (1993), C4.5: Programs for Machine Learning, Morgan Kaufmann, Los Altos.

Suzuki, E. and Y. Kodratoff (1998), Discovery of surprising exception rules based on intensity of implication, in PKDD'98, pp. 10-18.

Wang, K., S. Zhou and Y. He (2000), Growing decision tree on support-less association rules, in Proc. of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'00), Boston, pp. 265-269.

Wets, G., K. Vanhoof, T.A Arentze and H.J.P. Timmermans (2000), Identifying Decision Structures Underlying Activity Patterns: An Exploration of Data Mining Algorithms. Transportation Research Record: Journal of the Transportation Research Board, No. 1718, TRB, National Research Council, Washington, D.C., 1-9.