Average and Majority Gates: Combining Information by Means of Bayesian Networks

L.M. de Campos, J.M. Fernández-Luna, J.F. Huete, and M.A. Rueda-Morales

Departamento de Ciencias de la Computación e Inteligencia Artificial E.T.S.I. Informática, Universidad de Granada, 18071 - Granada, Spain {lci,jmfluna,jhg,mrueda}@decsai.ugr.es

Abstract. In this paper we focus on the problem of belief aggregation, i.e. the task of forming a group consensus probability distribution by combining the beliefs of the individual members of the group. We propose the use of Bayesian Networks to model the interactions between the individuals of the group and introduce average and majority canonical models and their application to information aggregation. Due to efficiency restrictions imposed by the Group Recommending problem, where our research is framed, we have had to develop specific inference algorithms to compute group recommendations.

1 Introduction

In this paper we investigate the value of using Bayesian Networks (BN) to represent how different individuals in a group interact in order to achieve a final choice or recommendation. Although aggregating information is a common task to a number of disciplines, including statistics, decision theory, economics, political science, psychology, etc., our research is framed in the problem of *Group Recommending*, task included in the more general field of Recommending Systems (RS). The the objective is to obtain the most appropriate recommendations for groups of people where their members may be inter-related in different ways. In this problem it is usual to assume that the individuals do not have observed those items that might be recommended. This kind of RS is appropriate for domains where a group of people participates in a single activity such as watching a movie or going on holiday and also situations where a single person must make a decision about a group of people. This is a relatively novel problem (research started focusing on group recommending at the start of the 21st century [1,2]) and has hardly been researched in the literature.

Since the focus of this paper is on the *combination* of individuals opinions, we will discuss, non the process by which the individuals reach their opinion, neither the relationships between the members of the group. In this case we shall assume that all the individuals use the same set of labels to express their preferences on an item, and that these preferences are represented by means of a probability distribution (probably estimated from a data set). On the other hand, we will not discuss about subjects such as how the groups are formed or how long they have existed.

K. Mellouli (Ed.): ECSQARU 2007, LNAI 4724, pp. 572-584, 2007.

According to [3] "... there is nothing close to a single well-accepted normative basis for group beliefs, group preferences or group decision making." In this paper we show how different common decision rules in the literature could be managed by a proper design of canonical models with the language of BN, giving some new lights into the combination processes, particularly:

- The average strategy, AVG, which obtains the group rate as the average of the members' rates.
- The majority strategy, MAJ, which obtains the group rate as a simply counting of the votes that every group's member gives for an item. The final rate will be that one with more votes in the count.

The second section of this paper presents related work on information aggregation. Section 3 describes how to model the group interaction by means of BN topology, and presents the use of both average and majority gates. Section 4 presents some experimental results obtained when applying the proposed methodology to recommend movies for group of users. Finally, Section 5 includes our conclusions and some comments about further research.

2 Related Work

There are many papers focusing on combination of probabilistic information, ranging from a pure statistical approach (see [4,5] for a review) to more applied problems, as the combination of classifiers [6], prediction markets [7], various sources of information in a single BN [8] or different BNs into a unique model [3].

In general, the methods for combining information are dichotomized [4] into *mathematical* and *behavioural* approaches. Mathematical approaches consist of processes or analytical models that operate on the individual probability distributions to produce a single "combined" probability distribution; whereas behavioural approaches attempt to generate agreement among expert by having them interact in some way. Since this paper is focused on mathematical approaches we are going review those paper relevant to this subject.

Combining Probability Distributions: Mathematical approaches can be further distinguished into axiomatic approaches (considering a set of assumptions that the combination criteria might satisfy) and Bayesian approaches [4]:

- Axiomatic approach: Common functions to deal with belief aggregation are: i) Linear Opinion Pool where the probability of the group, Pr(G), is obtained as the weighted arithmetic average over the probabilities of the individuals, $Pr(V_i), i = 1, ..., n$, i.e. $Pr(G) = \sum_{i=1}^{n} w_i Pr(V_i), w_i$ being weights summing one.
 - ii) Logarithmic Opinion Pool (weighted geometric average) defined as $Pr(G) = \alpha \prod_{i=1}^{n} Pr(V_i)^{w_i}$, α being a normalization constant and the weights w_i (called expert weights) typically are restricted to sum one. If the weights are equal to 1/n, then the combined distribution is proportional to geometric average.

- Bayesian Approach [5,4] has been used to combine expert information. This approach assumes that if there is a decision maker who has prior probability over the group vote Pr_0 , and a likelihood function over individual opinions given the group vote, then, taking the individuals opinions as evidence the group priors over the pattern of vote can be updated according to Bayes rule. Usually, an in order to obtain efficient combinations, it is assumed that individuals opinion are conditionally independent given the group vote.

Group Recommending: Although the problem of group recommending is relatively new, the same dichotomy can be found, depending on whether they use individuals opinions to get to a consensus recommendations [9,10,11] or not [1,2,12]. In general, when focusing on "mathematical" approaches, ad hoc combinations criteria have been used. For instance [2], which selects the music stations to be played at a gym, computes the group preference for each (music) category by summing the squared individual preferences. Then, using a weighted random selection operator, the next music station to be played is selected.

Related to collaborative-based group RS is Polylens [1], which is an extension of the MovieLens [13] system that recommends movies to groups of users. These systems use nearest neighbour algorithms to find those individuals which are similar to group tastes and to obtain recommendations which merge the voting preferences of these individuals according to the principle of least misery (minimum criterion).

3 Modeling Group Decision Networks

As mentioned before, in this paper we shall not consider questions about how groups are formed nor how they are managed. We shall therefore assume that we know the composition of the groups, and our problem is to study how this information can be represented in the BN and also how to predict recommendations for groups, i.e. how the inference processes can be performed.

We propose to identify a group as an entity where recommendations are made by considering the particular recommendations of its members in some way. Figure 1 shows a typical situation where each member of the group has a guess about the probability of relevance of a given item. In order to model group behaviour we propose that the group node (G) has as parents (Pa(G)) the set of nodes representing its individuals (right hand side of Figure 1). In this paper we shall not discuss the case where the combination mechanism might be represented by means of a "naive-Bayes-like" approach, i.e. a root group node having as children the set of individuals. This modelization might be related to the classical Bayesian Approach for combining probability distributions. Following with our model, in Figure 1 we use dashed lines to represent the idea that the individuals opinion would be obtained by considering different pieces of information. It is interesting to note that we do not impose any restriction about the dependences or independences between the members of the group, i.e. the individuals might use information of some common variables when predicting



Fig. 1. Modeling group's interactions

their votes. We will only assume that the group opinion is independent of the information sources, $\mathcal{I}S$, given that we know the opinion of its members, i.e. $I(G|Pa(G)|\mathcal{I}S)$.

With the idea of being general, the predicted rate for the i^{th} group, G_i , will be obtained by considering the individuals opinions, possibly updated after knowing new pieces of evidence, ev. Ideally, the opinion obtained by merging these individuals probabilities should represent the information of the group. Considering the proposed BN representation, the posterior probability of the group voting with rate s is obtained by means of

$$Pr(G_i = s|ev) = \sum_{pa(G_i)} Pr(G_i = s|pa(G_i)) \times Pr(pa(G_i)|ev).$$

being $pa(G_i)$ a particular configuration for the parent set of G_i , i.e an instance of its members. We shall denote by $\mathcal{R} = \{1, \ldots, r\}$ the set of possible rating alternatives. These conditional probability distributions can be considered as a "social value function" describing how the opinions of the members affect the group's recommendation.

One of our objectives is to obtain combining strategies able to be implemented in real-time situations. Therefore the reduction of the computational complexity becomes a key design parameter. Thus, since in real Group Recommending applications, the problem is to select those items, I_k , with higher probability of being liked by the group, i.e. for each pair *item-group* we have to compute the probability $Pr(G_i|I_k)$ and considering that there usually exists a large set of unobserved items in the system (which act as the evidence) this process becomes computationally expensive and therefore it is necessary to look for efficient algorithms. It is also interesting to note that with this requirement in mind, we propose to use canonical models to define the probabilities $P(G_i|pa(G_i))$. In general, since the group might be large, it implies important savings in storage (we do not need to store the exponential number of conditional probability values) and also efficient inference algorithms could be obtained.

3.1 Weighted Sum Gate

With this gate we are modeling that the group rate can be considered as the average vote of its members. This situation can be represented by means of the following additive canonical model, which is an extension of [14]:

$$P(G_i = k | pa(G_i)) = \sum_{j=1}^{|Pa(G_i)|} w(v_{j,s}, g_{i,k}),$$
(1)

where $w(v_{j,s}, g_{i,k})$ can be interpreted as the weight (effect) that the j^{th} group member voting the s^{th} value has in the k^{th} rate of the group G_i . The only restriction that we have to impose is that w are a set of non-negative weights verifying that

$$\sum_{k \in \mathcal{R}} \sum_{V_j \in Pa(G_i)} w(v_{j,t}, g_{i,k}) = 1, \forall \ pa(G_i)$$

It is interesting to note that by the way of defining how to compute the weights w we can control the bias of the individuals (bias is related to the preference of a user to one particular vote and its capability to predict the group judgements) and also the relative quality (importance) of the individuals in the group. For example, given the group in Figure 1 with $\mathcal{R} = \{1, 2, 3\}$, assuming that all the users are equal for prediction purposes and that there is no individual bias, i.e. the weights might be defined as follows

$$w(v_{j,t}, g_{i,k}) = \begin{cases} \frac{1}{|Pa(G_i)|} & \text{if } k = t, \\ 0 & \text{otherwise.} \end{cases}$$
(2)

Then, we have that $Pr(G_i|1, 2, 2, 2) = \{0.25, 0.75, 0.0\}$ and $Pr(G_i|1, 2, 2, 3) = \{0.25, 0.5, 0.25\}.$

Propagating with canonical weighted-sum. Assuming that, given the members opinion, the group's rating is independent of the $\mathcal{I}S$, the *exact* a posteriori probabilities for group nodes, $Pr(G_i = s|ev)$, can be computed efficiently by means of a straight application of the following theorem¹:

Theorem 1. Let G_i be a group node, let V_1, \ldots, V_n be the individuals in the group. If the conditional probability distributions can be expressed with a canonical weighted sum and the evidence, ev, belongs to the information sources, $\mathcal{I}S$, then the exact a posteriori probability distribution for the group can be computed using the following formula:

$$Pr(g_{i,s}|ev) = \sum_{j=1}^{n} \sum_{t=1}^{r} w(v_{j,t}, g_{i,s}) \cdot Pr(v_{j,t}|ev).$$

 $^{^{-1}}$ Due to the lack of space, we do not include the proof of the theorems.

We consider this theorem important because it expresses that exact propagation can be done without imposing any restriction about the dependence or independence among individuals patterns of rating. Moreover, when there is no individual bias, i.e., the weights can be defined as $w(v_{j,t}, g_{i,s}) = w_j$ if t = s, and 0 otherwise, our model coincides with the classical Linear Opinion Pool. Thus Linear Opinion Pool can be considered as a particular case of the average gate.

3.2 Majority Gate

Our objective in this section is to model the Majority criterion where the final decision will depend on a simple counting of the votes received for each rating from the individuals. The rate which receives the largest number of votes is then selected as the consensus (majority) decision. This is the usual combination strategy when, for each individual, we only know the label representing his/her rate.

Definition 1 (Majority Gate). A group node G_i is said that represents a majority combination criterion if given a configuration of its parents $pa(G_i)$ the conditional probability distributions can be represented as

$$Pr(G_i = s | pa(G_i)) = \begin{cases} \frac{1}{m} & \text{if } s = \arg\max_k count(k, pa(G_i)) \\ 0 & \text{otherwise} \end{cases}$$
(3)

where $count(k, pa(G_i))$ is a function returning the number of occurrences of the state k in the configuration $pa(G_i)$, and m is the number of states where $count(k, pa(G_i))$ reaches the maximum value.

For example, considering a node G_i with five parents and with three candidate rates, ranging from 1 to 3, then $Pr(G_i|1, 1, 2, 1, 1) = (1, 0, 0)$ and $Pr(G_i|1, 2, 2, 1, 3) = (0.5, 0.5, 0)$. This representation of the majority gate implies an important saving in storage (we can compute its values when needed). Nevertheless, and in order to combine the individuals opinions, we need to perform the exponential number of computations. We shall see how these computations can be performed efficiently.

Propagating with majority gates: A key idea behind majority criterion is that the order in which the individuals are considered does not matter, and therefore there exist many different configurations collapsing to the same situation. For example, consider that four individual vote with 1 and one individual votes with 2. In this case there are five different configurations representing the same situation, i.e. $pa_1(G_i) = \{2, 1, 1, 1, 1\}, pa_2(G_i) = \{1, 2, 1, 1, 1\}, pa_3(G_i) = \{1, 1, 2, 1, 1\}, pa_4(G_i) = \{1, 1, 1, 2, 1\}$ and $pa_5(G_i) = \{1, 1, 1, 1, 2\}$.

It must be noticed that since the order is not a factor, we might talk about combinations. We will denote by $\Delta(G_i)$ the set of combinations with repetition from the individual votes in $Pa(G_i)$ and we use $\delta(G_i)$ or <> to denote a single combination. Thus, the above situation should be represented by $\delta(G_i) =<$ 1,1,1,1,2>. Considering that the number of parents of G_i is n and that each parent has r different states we find that the number of possible combinations with repetition is $CR_n^r = (n + r - 1)!/(n!(r - 1)!)$

The next theorem shows that in order to combine the different individual rates we only need to take into account the probability distributions associated to the set of combinations with repetition.

Theorem 2. Let G_i a group node in a BN whose conditional probability distributions are represented using a majority gate, let $\Delta(G_i)$ be the set of possible combinations with repetition of the values in its parent set, $Pa(G_i)$, then

$$Pr(G_i = s|ev) = \sum_{\delta(G_i) \in \Delta(G_i)} Pr(G_i = s|\delta(G_i)) Pr(\delta(G_i)|ev)$$
(4)

This theorem shows that if we know $Pr(\delta(G_i)|ev)$, the combination of the information with a majority gate could be done in a time proportional to the size of CR_n^r , i.e. in the order of $O(n^{r-1})$. Taking into account that in many situations $r \ll n$, this implies important savings with respect to considering the number of possible configurations, $O(r^n)$. For instance, if n = 20 and r = 2 we have that $CR_n^r = 21$ whereas the number of configurations (permutations) is more than 1 million.

Nevertheless, to compute $Pr(\delta(G_i)|ev)$ we must sum over all the possible configurations in the combination, i.e.

$$Pr(\delta(G_i)|ev) = \sum_{pa(G_i) \in \delta(G_i)} Pr(pa(G_i)|ev)$$

where $pa(G_i) \in \delta(G_i)$ represents that the combination with repetition $\delta(G_i)$ can be obtained from the configuration $pa(G_i)$ by removing the order constraints. Thus, since we shall need to compute these probability values for all the possible combinations, we find that an exponential number of computations will be required to obtain the group decision.

Assuming independence to approximate $Pr(\delta(G_i)|ev)$. Since we want to compute $Pr(\delta(G_i)|ev)$ efficiently, we propose to approximate this joint distribution by assuming independence between the individuals. Although this assumption might be very restrictive, it has been demonstrated very fruitful in practical purposes when combining information [4,6].

Firstly, and with the idea of being general, we are going to introduce some notation: Let $\pi(x)$ be any configuration of n independent variables X_1, \ldots, X_n . As these variables are independent $Pr(\pi(x)) = \prod_{i=1}^n Pr(x_{i,j}), x_{i,j}$ being the value that variable X_i takes in the configuration $\pi(x)$. Let δ_k be a combination with repetition of a subset of k variables, X_1, \ldots, X_k , and $s \in \delta_k$ represents the fact that there exists at least one variable taking the value s in the combination δ_k . Also, we say that δ_{k-1} is a s-reduction of δ_k , denoted by $\delta_k^{\downarrow s}$, if δ_{k-1} can be



Fig. 2. Recursion graph for computing Pr(<1, 1, 1, 2, 2>)

obtained by removing a value s from the combination δ_k . The following theorem shows how $Pr(\delta_n)$ can be computed recursively:

Theorem 3. Let δ_n be any combination with repetition from the set of X_1, \ldots, X_n . Then, if X_i is independent of $X_j, \forall i \neq j$, the probability associated with the combination δ_n can be computed as

$$Pr(\delta_n) = \begin{cases} Pr(x_{1,k}) & \text{if } n = 1, \text{ with } \delta_1 = < k > \\ \sum_{s \in \delta_n} Pr(\delta_{n-1}^{\downarrow s}) Pr(x_{n,s}) & \text{if } n > 1 \end{cases}$$
(5)

A first idea should be to apply directly this result to compute $Pr(\delta(G_i)|ev)$. For instance, Figure 2 shows the recursion graph for the computation of Pr(< 1, 1, 1, 2, 2 >), where each different combination obtained after a reduction has been displayed only once. The key observation is that the number of (sub)combinations obtained after applying a reduction process is relatively small. Thus, a recursive algorithm may encounter each one of them many times in different branches of its recursion graph. For example, Figure 2 shows that the (sub)combination Pr(< 1, 1, 2 >) should be computed two times and the (sub)combination Pr(< 1, 1 >) three times. Moreover, some of these subproblems might also appear when computing different joint probabilities, like Pr(< 1, 1, 2, 2 >). Therefore, applying directly Theorem 3 does more work than necessary.

We propose to compute every probability for a given subcombination just once and then saves its values in a table, thereby avoiding the work of recomputing this probability every time the subcombination is encountered.

The next algorithm shows how to compute the joint probability distributions for all the possible combinations with replacement in the set Δ . We follow a bottom-up approach where we first compute the probabilities associated with the (sub)combinations with lower size, being the size the number of variables used to form the combinations with repetition, and these probabilities will be used as the base of the calculus for the combinations with greater size. Initially, when considering the first variable X_1 , we have r different combinations with replacement, one for each possible value of the variable X_1 . Then, in a general stage, we found out that the probabilities associated with each combination δ_k of the first k variables are used in the computation of the probabilities of r different combinations with size k + 1, one for each possible value of the variable X_{k+1} . Each one of these combinations will be denoted by $\delta_{k\cup s}$ with $1 \leq s \leq r$.

Computing $Pr(\Delta)$
$Pr(\delta_1) = Pr(X_1)$
for (k = 1; k <n;)<="" k++="" td=""></n;>
for each $\ \delta_k \in CR^r_k$ do // for each combination of size k
for (s = 1; s<=r; s++) //for each value of X_{k+1}
$Pr(\delta_{k\cup s}) + = Pr(\delta_k) \times Pr(x_{k+1,s})$

An inspection of the algorithm yields a running time of $T(n) = \sum_{i=1}^{n} rCR_{i}^{r}$, i.e. $T(n) \in O(rn^{r})$, being much more efficient than applying directly the recursive algorithm from Theorem 3. For example, in case of bivaluated variables, as the usual case in decision problems, we have a quadratic algorithm for combining the output of the different individuals. With respect to the memory needed to store the intermediate results we find out that the values in the stage k are only used in the stage k - 1, therefore the used memory is in the order of $O(CR_n^{r})$.

4 Experimentation: Recommending Movies for Groups

In order to study the performance of the use of AVG or MAJ gates when combining probabilistic information we consider the following problem: The prediction of the rate with which a group of people might score a given movie.

The data sets: With respect to the used datasets, they have been obtained from MovieLens². Since MovieLens does not include group information we have decided to build them from the MovieLens training sets³. We have used two different criteria trying to capture different processes behind the creation of a group: (C1) Implementing the idea of *the group of my colleagues*, we set each user as the group administrator and we look for the 10 most similar users (those which are positively correlated with the administrator in the training dataset). Then, we select those groups of five individuals with the only restriction that at least they have rated (observed) one movie in common. Note that since similarities are not transitive, this criterion does not imply necessarily having groups with highly correlated members. (C2) Second, we have decided to fix a group (also with five individuals) with the only restriction that all the members of the group must rate at least four common movies.

With respect of the group test sets, they are obtained from each one of the MovieLens test sets. Particularly, whenever we find a movie in the test sets which has been rated by all the members of a group we include the tern (group ID, movie ID, group rate) in its group test dataset. Note that the group rate is obtained, by means of a deterministic function $CombineRate(r_1, \ldots, r_n)$, as the

 $^{^2}$ MovieLens was collected by the GroupLens Research Project at the University of Minnesota. The dataset contains 1682 movies and 943 users, containing 100,000 transactions where a user rates a movie using a scale from 1 to 5.

³ With the idea of using 5 fold cross validation, we have used 5 different data subsets, each one obtained by splitting MovieLens into two disjoint sets, the first one for training (with 80% of the data) and the second one for test (with 20% of the data).

average or the majority individual's true rate, r_1, \ldots, r_n . Therefore, combining the decision used by a group to rate a movie and the criterion used to form a group, we obtain four different test datasets, i.e. AVG-C1 and MAJ-C1 (with a mean of 115 different records) and AVG-C2 and MAJ-C2 (with a mean of 17524 records).

Selecting a rate: Given Pr(X) encoding a probability distribution over the candidate rates, the problem is to determine which is the output rate that should be recommended for X. Two basic alternatives might be considered for the RateSelection process:

MP Maximum a posteriori probability, i.e. rate = $\arg \max_s \{Pr(X = s | ev)\}$. AP A posteriori average rate, i.e. rate = $\sum_{k=1}^r k \times Pr(X = k | ev)\}$.

Experimental framework: The objective is to predict which is the rate that a group of people will use to score an unobserved movie, I. We assume that for each member of the group, V_k , we know a probability distribution representing the belief about how this individual should rate this movie, i.e. $Pr(V_k = s|I)$. Particularly, in this experimentation, these probabilities are estimated using a *Collaborative-based RSs* [12], i.e. they have been estimated by considering the ratings that users similar to V_k have given to the movie I.

In this framework it might be possible that, for a given user V_k , none of the users similar to him had rated the movie I. In this situation the probabilities $Pr(V_k|I)$ have been estimated without information. Therefore, and in order to study the bias that the a priori distribution might produce in the predicted rate, we will consider a modification of the approaches used to select a rate: The idea is to use only the new piece of evidence that each candidate rate receives, computed as the difference between the a priori (without evidence) and the a posteriori probability values, i.e. $Pr(G_a = s|ev) - Pr(G_a = s)$. Note that this idea could be used with both AP (denoted by PD+AP) or MP (denoted by PD+MP) in the RateSelection process.

In order to study the performance of the combination methods, we are going to consider two different situations: In the first one, that could be considered the *Baseline* (see left hand side of table below), the predicted rate is defined as the average or majority rate of r_1, \ldots, r_n , being r_k the rate that, individually, each member predicts using **RateSelection**. The second alternative (right hand side below) consists on, firstly using the AVG or MAJ gates, combining the individuals probability distributions into a unique group distribution which ideally represents the group pattern of vote (CombineProb(Pr_1, \ldots, Pr_n)). Then, the group rate is selected.

Baseline	Using Group Layer
For each $V_k \in G$ do	$Pr(G ev) = $ CombineProb(Pr_1, \ldots, Pr_n)
$r_k = RateSelection(Pr_k)$	$G_{rate} = RateSelection(Pr(G ev))$
$G_{rate} = CombineRate(r_1, \ldots, r_n)$	

			(C1) Using Similarity				(C2) Common movies				
Group	Comb.	Rate	Baseline		Group layer		Baseline		Group layer		
Dec.	Gate	Selection	%S	MAE	%S	MAE	%S	MAE	%S	MAE	
AVG	AVG	AP	$47,\!53$	0,566	$45,\!37$	$0,\!590$	$57,\!40$	$0,\!441$	$58,\!90$	$0,\!422$	
		PD+AP	60,86	$0,\!398$	$61,\!42$	$0,\!402$	66,76	0,337	64,76	0,361	
	MAJ	MP	$45,\!63$	$0,\!495$	62, 13	0,394	$44,\!33$	$0,\!486$	62, 12	0,392	
		PD+MP	44,99	$0,\!497$	62, 32	$0,\!392$	$45,\!05$	$0,\!481$	$61,\!86$	$0,\!395$	
MAJ	AVG	AP	$43,\!01$	$0,\!655$	40,72	0,683	$47,\!82$	$0,\!578$	48,23	0,567	
		PD+AP	$54,\!64$	$0,\!476$	$58,\!24$	$0,\!446$	$56,\!30$	$0,\!457$	$58,\!44$	$0,\!438$	
	MAJ	MP	59,34	0,461	60,10	0,424	55,11	$0,\!493$	58,05	0,446	
		PD+MP	59, 19	$0,\!461$	60, 29	$0,\!422$	55,72	$0,\!486$	$58,\!35$	$0,\!443$	

 Table 1. Experimental Results

Two different accuracy measures will be considered [13]: the percentage of success (%S), which measures the frequency with which the system makes correct predictions and the mean absolute error (MAE), which measures the average absolute deviation between a predicted rate and the group's true rate.

Table 1 presents the average results obtained after repeating the experiment with the 5 folds. First column represents the criterion used in the test dataset to decide the group rate. Second column represents the canonical model used in the group layer. In this experimentation, we have used an unbiased uniform weighting scheme in the AVG gate, i.e. $w(v_{j,s}, g_{i,k}) = 1/|Pa(G_i)|$ if k = s and 0 in the other cases. Note that this model corresponds with the classical Linear Opinion Pool when all the users are considered equivalent for prediction purposes. Third column represents the criteria used in the **RateSelection** process. The next two columns represents how the groups have been constructed. In this table, the results for each particular dataset can be found in the cells indexed by the pair (*Group Dec., group construction criterion*).

From this table, as general conclusions, we can say that: i) it is better to use BN (the AVG or MAJ gates) to combine individual preferences, ii) MAJ gate could be preferable in the case it is unknown how the group (true) decisions are obtained (quite good results have been obtained using also MAJ gates when the *"real"* group vote is obtained using AVG), iii) With respect to the AVG gate, it seems preferable to use PD+AP to correct the a priori bias of the AVG gate. Also, better results have been obtained when the groups have been formed without considering similarities (we believe that these results can be improved by using proper weights in AVG gate), iv) With respect to MAJ gate it seems to perform better when considering similar users. No significant differences can be found between the use of MP or PD+MP.

5 Conclusions

A general BN-based model for combine probabilistic information in a group recommending framework has been proposed in this paper. With this model the interaction among the individuals when deciding a group rate are represented intuitively by means of the use of AVG (average) and MAJ (majority) canonical modes. Linear time inference algorithms (assuming independence in the case of MAJ gate) have been developed to compute the a posteriori distribution for the group. These distributions represent the group preferences for a given item. Experimental results show the validity of our proposal.

By way of future work, we are planning to evaluate the model with real data, involving real groups to determine the quality of the recommendations provided and also to apply these methodology to problems as the combination of classifiers.

Acknowledgements

This work has been supported by the Spanish 'Ministerio de Educación y Ciencia' and 'Consejería de Innovación, Ciencia y Empresa de la Junta de Andalucía' under Projects TIN2005-02516 and TIC-276, respectively.

References

- O'Connor, M., Cosley, D., Konstan, J.A., Riedl, J.: Polylens: A recommender system for groups of user. In: Proceedings of the Seventh European Conference on Computer-Supported Cooperative Work, pp. 199–218 (2001)
- McCarthy, J.E., Anagnost, T.D.: Musicfx: an arbiter of group preferences for computer supported collaborative workouts. In: CSCW '00: Proceedings of the 2000 ACM conference on Computer supported cooperative work, p. 348. ACM Press, New York (2000)
- Pennock, D.M., Wellman, M.P.: Graphical models for groups: Belief aggregation and risk sharing. Decision Analysis 2(3), 148–164 (2005)
- Clemen, R.T., Winkler, R.L.: Combining probability distributions from experts in risk analysis. Risk Analysis 19, 187–203 (1999)
- 5. Genest, C., Zidek, J.: Combining probability distributions: A critique and annotated bibliography. Statistical Sciences 1(1), 114–148 (1986)
- Kittler, J., Hatef, M., Duin, R., Matas, J.: On combining classifiers. IEEE Trans. on Pattern Anal. and Machine Intell. 20(3), 226–238 (1998)
- Chen, Y., Chao-Hsien Chu, T.M., Pennock, D.M.: Information markets vs. opinion pools: An empirical comparison. In: EC '05: Proceedings of the 6th ACM conference on Electronic Commerce, pp. 58–67. ACM Press, New York (2005)
- Druzdzel, M., Diez, F.J.: Combining knowledge from different sources in causal probabilistic models. Journal of Machine Learnig Research 4, 295–316 (2003)
- McCarthy, K., Salam, M., Coyle, L., McGinty, L., Smyth, B., Nixon, P.: Group recommender systems: a critiquing based approach. In: IUI '06: Proceedings of the 11th international conference on Intelligent user interfaces, pp. 267–269. ACM Press, New York (2006)
- Masthoff, J., Gatt, A.: In pursuit of satisfaction and the prevention of embarrassment: affective state in group recommender systems. User Model User-Adap Inter. 16, 281–319 (2006)

- Yu, Z., Zhou, X., Hao, Y., Gu, J.: Tv program recommendation for multiple viewers based on user profile merging. User Model. and User-Adap. Inter. 16(1), 63–82 (2006)
- de Campos, L.M., Fernández-Luna, J.M., Huete, J.F., Rueda-Morales, M.A.: Group recommending: A methodological approach based on bayesian networks. In: IEEE-Workshop on Web Personalization, Recommender Systems and Int. User Interface, pp. 835–845. IEEE Computer Society Press, Los Alamitos (2007)
- Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T.: Evaluating collaborative filtering recommender systems. ACM Trans. Inf. Syst. 22(1), 5–53 (2004)
- de Campos, L.M., Fernández-Luna, J.M., Huete, J.F.: The bnr model: foundations and performance of a bayesian network-based retrieval model. Int. J. Approx. Reasoning 34(2-3), 265–285 (2003)