On the Use of BOWA Operators in Cluster Analysis for Collaborative Filtering

Hanna Łącka

Institute of Computer Science, Polish Academy of Sciences
ul. Jana Kazimierza 5, 01-248 Warsaw, Poland

Abstract. In this paper a construction of a similarity measure between groups of rankings, based on a Bipolar OWA (BOWA) function is discussed. The measure possesses some interesting properties that make it useful in cluster analysis performed as a part of collaborative filtering process. An extended data representation model for consumer preferences and objects of their preferences is assumed. Practical issues of conducting and evaluating the clustering procedure are discussed.

Keywords: Aggregation function, association measure, bipolarity, cluster analysis, collaborative filtering, ordering, OWA operator, preferences, ranks, rating, recommender system.

1 Introduction

The most common web-based recommender systems predict what movies, books or other goods a user would prefer, based on the historical ratings, views or purchases of the user [1, 2]. Explicit user feedback once given continues to be useful. The most popular setting in which preferences are represented in such systems is a matrix (sometimes called a *utility matrix*) with rows corresponding to users, columns corresponding to items and cells containing values of ratings given to items by the users. In [3] a more complicated setting of representing user preferences was proposed. It is especially suited for recommending services, e.g. vacation trips, cultural events, conferences, for which no explicit feedback exists. The setting has a form of a matrix where the entries of the cells for each user-attribute pair contain rankings. Each ranking expresses user preference for items belonging to the domain of a given attribute.

Collaborative filtering approach is the most common technique successfully applied in recommender systems [4, 5]. It creates item recommendations based on similarity measures between users and/or items. An application of cluster analysis and grouping the users based on their similarity was considered a natural and interesting direction of inference about preferences.

In order to apply this approach for the assumed data representation, a similarity measure between groups of rankings [3] coming from a pair of users was

defined. The measure is based on a function that is a member of the family of BOWA operators (Bipolar Ordered Weighted Averaging function) proposed in [6], which are used to aggregate bipolar data.

In this paper a work on the similarity measure between groups of rankings is summarized and the application of the measure in cluster analysis is proposed. The clustering is assumed to be a part of a collaborative filtering process whose purpose is to detect natural similarity groups of consumers, as opposed to a possible segmentation approach [7], and to generate object recommendations for the consumers.

The paper is organized as follows. An introductory example explaining the difference between the classical situation considered in recommender systems and the suggested setting is given in Section 2. In Section 3 a data representation model for consumers is recalled. We also propose a data representation for objects of consumer preferences, based on groups of vectors and referring to consumer data representation model. A notion of a *history of choices* of a consumer is introduced. Section 4 recalls definition of the chosen similarity measure between groups of rankings and of the family of operators the measure is based on. In Section 5 we propose to apply the defined similarity measure in cluster analysis and describe example choice of existing methods to conduct and evaluate it.

2 Introductory example

Let us consider a travel agency, that gathers a history of vacation trips of its clients. For every client a number of times he or she attended a certain trip is known.

The agency stores data about trips in the following way. Every trip is described by the same set of attributes. For each separate attribute the availability of its all possible variants (which depend on a domain of the attribute) is marked. An exemplary output of trips (or trip types) data set is given in Table 1, where 1 means a certain variant is available and 0 it is not.

Data about clients are stored in a form of rankings of choices made by the clients. For each attribute of a trip, all available variants concerning this aspect of a trip have the ranks assigned, according to historical preferences of the client. In other words, for each client a ranking of variants of a given attribute from the most preferred to the least preferred variant is obtained. Table 2 presents an exemplary output of the clients data set, where numbers indicate the ranks assigned.

Such data sets as shown in Table 1 and Table 2 are collected because the agency plans to prepare new trip offers for each client. To maximize the possibility of accepting a new offer it should be prepared in a way that guarantees client's satisfaction when chosen. To achieve this, the agency plans to create recommendations based on similarity between clients and/or trip offers.

Table 1. Exemplary data set of trips.

	Accommodation	Means of transport	Activities
Trip T	tent - 0 guesthouse - 1 hotel - 0	car - 1 bus - 0 train - 1 airplane - 0	sunbathing - 1 sightseeing - 0
Trip U	tent - 1 guesthouse - 0 hotel - 0	car - 1 bus - 0 train - 0 airplane - 1	sunbathing - 1 sightseeing - 1
...	...	...	...

Table 2. Exemplary data set of clients' preferences.

	Accommodation	Means of transport	Activities
Client A	tent - 2 guesthouse - 1 hotel - 3	car - 1 bus - 4 train - 2 airplane - 3	sunbathing - 1 sightseeing - 2
Client B	tent - 3 guesthouse - 2 hotel - 1	car - 3 bus - 4 train - 1 airplane - 2	sunbathing - 2 sightseeing - 1
...	...	...	...

3 Data representation

Let $\mathcal{Y}$ be a finite set of attributes of size n. Moreover, assume that $\mathcal{U}_j$ is a domain of the attribute $Y_j \in \mathcal{Y}$ which consists of l_j variants, i.e. $\mathcal{U}_j$ is a finite set of size l_j, where $j = 1, \ldots n$.

Let $\mathcal{X}$ denote a set of consumers. For each attribute consumer preferences are expressed with respect to all available variants of that attribute. Hence, for any consumer $A \in \mathcal{X}$ we get n rankings corresponding to successive attributes, so the observation related to A might be perceived as a vector

$$\mathbb{R}_A = [R_{A1}, R_{A2}, \ldots, R_{An}], \tag{1}$$

where R_{Aj} is a ranking of variants belonging to the domain of the j-th attribute.

Consider now a ranking R_{Aj}. Since it reflects the consumer's preferences on variants belonging to the domain $\mathcal{U}_j$ of the attribute $Y_j \in \mathcal{Y}$, it is also a vector. Namely,

$$R_{Aj} = (r_{Aj}^{(1)}, r_{Aj}^{(2)}, \ldots, r_{Aj}^{(l_j)}), \tag{2}$$

where $r_{Aj}^{(k)}$, $k = 1, \ldots, l_j$ is a rank assigned to k-th variant belonging to $\mathcal{U}_j$ and where l_j stands for the size of the domain $\mathcal{U}_j$.

Next, let $\mathcal{Z}$ be a set of objects of preference (products, offers, events etc.). Each object $T \in \mathcal{Z}$ is characterized by available variants of the attributes $\mathcal{Y}$ already considered by consumers. Hence, object $T \in \mathcal{Z}$ is described by a vector

$$\mathbb{V}_T = [V_{T1}, V_{T2} \ldots, V_{Tn}] \tag{3}$$

where V_{Tj} is an l_j-element vector indicating available variants, i.e.

$$V_{Tj} = (v_{Tj}^{(1)}, v_{Tj}^{(2)}, \ldots, v_{Tj}^{(l_j)}), \tag{4}$$

where $v_{Tj}^{(k)} \in \{0, 1\}$, $k = 1, \ldots, l_j$, and $v_{Tj}^{(k)} = 1$ denotes that the k-th variant in $\mathcal{U}_j$ is available (in offer T), while $p_{Aj}^{(k)} = 0$ means that it is not available. In general there is no restriction on the number of simultaneously available variants.

Given a finite set of considered objects $\{T^{(1)}, T^{(2)}, \ldots, T^{(t)}\} \in \mathcal{Z}$ and a consumer $A \in \mathcal{X}$, let

$$H_A = [h_{A1}, h_{A2}, \ldots, h_{At}] \tag{5}$$

denote a consumer A's *history of choices*, where $h_{Ai} \in \{0, 1, \ldots, t\}$ for $i = 1, \ldots, k$, and $h_{Ai} = 0$ means the i-th object from the set $\{T^{(1)}, T^{(2)}, \ldots, T^{(t)}\}$ was never chosen by A, while $h_{Ai} \in \{1, \ldots, t\}$ denotes a rank assigned to the i-th object. For the most often chosen object we obtain $h_{A\cdot} = 1$, the second most often chosen object has rank 2, and so on till the least often chosen object.

We assume the client A's representation (1) is linked to the history of choices H_A in the following way. Observation $\mathbb{R}_A = [R_{A1}, R_{A2} \ldots, R_{An}]$ is generated on the basis of: a history of choices H_A and an additional information about exact number of times each object from the history was chosen. For a given j-th attribute we obtain a ranking R_{Aj} by summing up the number of times each of the l_j variants, if available, was chosen in the history and assigning ranks to each of the l_j obtained sums in the non-increasing order.

Example 1. Let $\{T, U, V\}$ be a considered set of objects of preference, such that

$$\begin{aligned}
\mathbb{V}_T &= [(0, 1, 0), (1, 0, 1, 0), (1, 0)] \\
\mathbb{V}_U &= [(1, 0, 0), (1, 0, 0, 1), (1, 1)] \\
\mathbb{V}_V &= [(1, 0, 0), (1, 1, 1, 1), (0, 1)].
\end{aligned}$$

Given a client A, his or her history of choices $H_A = [1, 2, 0]$ and additional information that the first object was chosen 13 times and the second object 4 times by the client A, we obtain the following vector of sums for each variant of each attribute:

$$\begin{aligned}
&[(1 \cdot 4, 1 \cdot 13, 0), (1 \cdot 4 + 1 \cdot 13, 0, 1 \cdot 13, 1 \cdot 4), (1 \cdot 13 + 1 \cdot 4, 1 \cdot 4)] = \\
&= [(4, 13, 0), (17, 0, 13, 4), (17, 4)].
\end{aligned}$$

After assigning ranks to the sums in the non-increasing order, we obtain client A's representation:

$$\mathbb{R}_A = [(2,1,3),(1,4,2,3),(1,2)].$$

□

4 Measure of association between groups of rankings

In order to group the consumers based on their similarity, for the assumed consumer data representation (1), a measure of association between two groups of rankings was searched for. A set of requirements which a measure should satisfy was specified [3] and the form of the desired measure between two groups of rankings corresponding to consumers A and B, $A, B \in \mathcal{X}$, was stated as

$$S(A,B) = F(s^1_{AB}, s^2_{AB}, \ldots, s^n_{AB}), \tag{6}$$

where $(s^1_{AB}, s^2_{AB}, \ldots, s^n_{AB})$ is a vector of pairwise correlations obtained for all attributes under study for two consumers $A, B \in \mathcal{X}$, i.e. $s^j_{AB} = s(R_{Aj}, R_{Bj})$, $j = 1, \ldots, n$, s denotes any pairwise correlation measure between two rankings, taking values in $[-1,1]$ (e.g. Kendall's τ or Spearman's r_S [8]) and $F : [-1,1]^n \to [-1,1]$ is a suitable function.

Since the goal of F is to aggregate several correlations to a single value, one may expect that it should be an appropriate aggregation function. The preservation of bounds property of any aggregation function coincides with the specified requirement that the measure should take its maximal (minimal) value when all rankings are pairwise perfectly concordant (discordant). One of the other requirements was to reward higher correlations, hence an OWA operator [9] might seem a good choice. However, the reward was postulated to be given regardless of the correlation signs. Hence, F cannot be monotone on the whole interval $[-1,1]$ and cannot fulfill the monotonicity condition (see, e.g., [10–12]) of any aggregation function.

A new family of semi-aggregation operators was therefore proposed [6]. It is a generalization of OWA operators for the case of bipolar data and, most importantly, was shown to be monotone for absolute values of arguments while still keeping track of signs.

Definition 1. *Let* $\boldsymbol{w} = [w_1, \ldots, w_n]$ *be a vector of weights such that* $w_j \geq 0$ *for* $j = 1, \ldots, n$ *and* $\sum_{j=1}^n w_j = 1$. *Suppose that* $x_1, \ldots, x_n$ *are realizations of the continuous random variable defined on the interval* $[-1,1]$. *A function* $F : [-1,1]^n \to [-1,1]$ *defined as*

$$F(x_1, \ldots, x_n) = \sum_{j=1}^n w_j \cdot x^*_{(j)} \tag{7}$$

is called the Bipolar OWA function (BOWA), where $x^*_{(j)}$ *denotes the* j*-th largest absolute value of element in the collection of aggregated objects* $x_1, \ldots, x_n$ *multiplied by the original sign of that element.*

Alternative notation to express BOWA operator is

$$F(x_1, \ldots, x_n) = \langle \boldsymbol{w}, \boldsymbol{x}^* \searrow_B \rangle \,, \tag{8}$$

where $\langle \cdot, \cdot \rangle$ is the scalar product of vectors and the symbol $\searrow_B$ indicates a non-increasing ordering (proposed to be called *bipolar*) of elements obtained for their absolute values and thus ignoring their signs.

The BOWA operator definition in the presence of ties in the *bipolar* ordering of arguments was separately defined in [6]. Arguments having the same absolute values are given identical weights, which are computed as the average of weights that would be gathered if the arguments had not been tied.

The basic BOWA operator properties [3] include idempotence, symmetry and homogeneity. Moreover, each BOWA function for absolute values of its arguments is an OWA operator. Similarly as OWA functions, BOWA operators do not have neutral or absorbing elements, except for the special cases. They are however not shift-invariant. BOWA operator with adequately chosen weights, such that higher correlations are rewarded whatever are their signs, is a suitable function F that satisfies all postulates required by the measure of association (6) searched for. An example of such vector of weights was suggested in [3] and is also recalled below.

Example 2. Consider two consumers A and B. Assume that the pairise correlation between their preferences for each of the three attributes under study was calculated using Spearman's coefficient. As a result we received the following three numbers: $s^1_{AB} = 0.5$, $s^2_{AB} = 0.4$ and $s^3_{AB} = -1$.

To aggregate these three coefficients the following operator $F_{LG} : [-1, 1]^n \to [-1, 1]$ was suggested in [3]

$$F_{LG}(x_i, \ldots, x_n) = \frac{2}{n(n+1)} \sum_{j=1}^{n} r(|x_j|) \cdot x_j, \tag{9}$$

where $r : [0, 1] \to \mathbb{R}^+$ is a function such that

$$r(z) = \frac{1}{2} + \sum_{i=1}^{n} c(z - |x_i|) \tag{10}$$

and where c is defined as

$$c(u) = \begin{cases} 0 & \text{if} \quad u < 0 \\ \frac{1}{2} & \text{if} \quad u = 0 \\ 1 & \text{if} \quad u > 0. \end{cases} \tag{11}$$

The suggested operator is a member of a family of BOWA operators (7). Let us consider given correlation coefficients as a vector $\boldsymbol{x} = [0.5, 0.4, -1]$. Hence we get a vector of argument values $\boldsymbol{x}^* \searrow_B = [-1, 0.5, 0.4]$ in the bipolar order. Using

ranks given by (10), we may compute a vector of weights $w = [0.5, 0.(3), 0.1(6)]$ and therefore, by (8) we get

$$F(0.5, 0.4, -1) = 0.5 \cdot (-1) + 0.(3) \cdot 0.5 + 0.1(6) \cdot 0.4 = -0.2(6).$$

□

5 Using BOWA-based similarity measure in cluster analysis for collaborative filtering

The constructed similarity measure (6) between a pair of consumers based on BOWA operator, allows us to conduct cluster analysis for a set of consumers represented as in (1). We assume the goal of such clustering proccess is to find natural similarity groups among consumers and characterize the groups.

Consider a finite set of consumers $X \subset \mathcal{X}$ and the K-medoids [7] as a cluster analysis method to be conducted on X. K-medoids is a combinatorial cluster analysis [7], a generalization of a popular K-means algorithm for observations with arbitrary attributes. It admits arbitrary dissimilarity measure instead of a squared Euclidean distance. The center of each cluster, the medoid, is the cluster member that minimizes a total dissimilarity to all other members of the cluster. The BOWA based S similarity measure (6) can be easily adopted to be used as a dissimilarity S' in K-medoids method, i.e. $S' = -S$.

Dissimilarity S' can then also be used for several distance-based clustering quality measures, as Silhouette coefficient [13], Gamma index [14], C-index [15] or Caliński and Harabasz index generalised for dissimilairites [16]. Another proposed way to assess clustering quality is to measure the averaged agreement among consumers belonging to the same cluster in relation to the agreement among medoids. To compute an overall agreement of a group of consumers we can use the analogy to how the BOWA based similarity measure between two groups of rankings is constructed (6). First, we measure the agreement for each attribute separately, i.e. concordance of a set of rankings, using e.g. the Kendall's coefficient of concordance [8] which ranges between 0 (no agreement) to 1 (perfect agreement). Then using OWA operator with a proper weight vector that rewards higher correlations, e.g. obtained by (10), we aggregate the coefficients to obtain single value agreement indicator.

Now, consider a certain resulting cluster and assume that the history of object choices (5) of each cluster member is known. Let $\{T^{(1)}, T^{(2)}, \ldots, T^{(t)}\}$ be a set of all considered objects that the histories are based on. The following procedures of obtaining a meaningful cluster description are suggested:

P1. Picking or creating a representative consumer. Obvious way of representing a cluster is to pick the consumer that serves as the medoid. However, an equivalent of a centroid [7] known from the K-means procedure could also be computed, i.e. the averaged member of the cluster having the form of a vector (1), such that its each element is obtained as a result of aggregating corresponding elements of vectors representing all cluster members.

P2. Creating a representative object. A single object $T^{(rep)} \in \mathcal{Z}$ of the form $\mathbb{V}_{T^{(rep)}} = [V_{T_1^{(rep)}}, V_{T_2^{(rep)}} \ldots, V_{T_n^{(rep)}}]$ is created such that for the j-th attribute each element of the vector $V_{T_j^{(rep)}}$ is obtained as a result of aggregating corresponding elements of vectors $V_{T_j^{(i)}}$, $i = 1, \ldots, t$, if the i-th object is among the ones most often chosen by the cluster members.

P3. Creating a representative history of object choices. A vector having the form of the history of choices (5) is created, such that its each element is obtained as a result of aggregating corresponding elements of vectors representing histories of choices of all cluster members.

Keeping in mind the introductory example discussed in Section 2, we observe that cluster analysis can be especially beneficial also for recommendation creation purpose, for the assumed data representation model.

Firstly, we notice that history of consumer choices (5) can serve as a ground truth for algorithms that learn to predict a preferred order of a given set of objects for a given consumer, i.e. object ranking [17] or preference-based [4] algorithms. In practice, histories of choices can vary a lot between consumers regarding the number and the types of objects chosen. Notice that it applies even to very similar consumers (where similarity is understood as defined in Section 4). Big differences in the types of objects chosen result in sparse history vectors. Here, the cluster analysis can be helpful in dealing with the sparsity. Any consumer A to be used in the learning or testing phase of the object ranking procedure can be replaced with his or her cluster's representative consumer (see P1.). Ground truth history of choices of the cluster's representant is enriched with the history of A, reducing the sparsity problem.

On the other hand, the fact that a given consumer is assigned to a certain cluster, can be used as additional information (a feature in the input vector) about the consumer, possibly improving the prediction quality of object ranking procedure.

6 Conclusions

In this paper an extended data representation model of consumer preferences and objects of their preferences was proposed. A construction of the similarity measure between groups of rankings coming from a pair of consumers, based on a family of semi-aggregation BOWA operators, was summarized. It was shown how the measure can be applied in cluster analysis performed as a part of collaborative filtering process and what are the motivations behind it. Practical issues of conducting and evaluating the clustering process were discussed. Further work assumes experimental verification of the proposed consumer clustering procedure, including the defined similarity measure, performed on real and generated data.

Acknowledgements

Study was supported by research fellowship within "Information technologies: research and their interdisciplinary applications" project co-financed by European Social Fund (agreement no. POKL.04.01.01-00-051/10-00).

References

1. Bennett, J., Lanning, S.: The Netflix Prize. In: Proceedings of KDD Cup and Workshop 2007. (2007)
2. Dror, G., Koenigstein, N., Koren, Y., Weimer, M.: The Yahoo! Music Dataset and KDD-Cup'11. Journal of Machine Learning: Workshop and Conference Proceedings **18** (2012) 3–18
3. Łącka, H., Grzegorzewski, P.: On Measuring Association between Groups of Rankings in Recommender Systems. In Rutkowski, L.e.a., ed.: Proceedings of the 13th International Conference on Artificial Intelligence and Soft Computing (ICAISC 2014), Part II, Lecture Notes in Artificial Intelligence 8468, Springer (2014) 423–432
4. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering **17**(6) (2005) 734–749
5. Shi, Y., Karatzoglou, A., Baltrunas, L., Larson, M., Hanjalic, A.: GAPfm: optimal top-n recommendations for graded relevance domains. In: Proceedings of the 22nd ACM International Conference on Information and Knowledge Management. (2013) 2261–2266
6. Grzegorzewski, P., Łącka, H.: Recommender systems and BOWA operators. In Angelov, P.e.a., ed.: Proceedings of the 7th International Conference on Intelligent Systems IEEE IS 2014, Part I, Advances in Intelligent Systems and Computing 322, Springer (2015) 11–21
7. Hastie, T., Tibshirani, R., Friedman, J.: Cluster analysis. Practical issues. In: The Elements of Statistical Learning. Springer (2001) 518–520
8. Gibbons, J., Chakraborti, S.: Nonparametric Statistical Inference. Marcel Dekker Inc. (2003)
9. Yager, R.: On ordered weighted averaging aggregation operators in multicriteria decisionmaking. IEEE Transactions and Systems, Man and Cybernetics **18** (1988) 183–190
10. Beliakov, G., Pradera, A., Calvo, T.: Aggregation Functions: A Guide for Practitioners. Springer (2007)
11. Calvo, T., Kolesarova, A., Komornikova, M., Mesiar, R.: Aggregation operators: Properties, classes and construction methods. In: Aggregation Operators. New Trends and Applications. Volume 97 of Studies in Fuzziness and Soft Computing. Springer (2002) 3–104
12. Grabisch, M., Pap, E., Marichal, J., Mesiar, R.: Aggregation Functions. Cambridge (2009)
13. Rousseeuw, P.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics **20**(1) (1987) 53–65
14. Baker, F., Hubert, L.: Measuring the power of hierarchical cluster analysis. Journal of the American Statistical Association **70** (1975) 31–38

15. Hubert, L., Levin, J.: A general statistical framework for assessing categorical clustering in free recall. Psychological Bulletin **83** (1976) 1072–1080
16. Hennig, C., Liao, T.: How to find an appropriate clustering for mixed-type variables with application to socio-economic stratification. Journal of the Royal Statistical Society, Series C Applied Statistics **62** (2013) 309–369
17. Fürnkranz, J., Hüllermeier, E.: Preference learning: An introduction. In Fürnkranz, J., Hüllermeier, E., eds.: Preference Learning. Springer-Verlag (2010) 1–17