Estimating Attributed Central Orders
An Empirical Comparison

Toshihiro Kamishima[1], Hideto Kazawa[2], and Shotaro Akaho[1]

[1] National Institute of Advanced Industrial Science and Technology (AIST),
AIST Tsukuba Central 2, Umezono 1–1–1, Tsukuba, Ibaraki, 305–8568 Japan
mail@kamishima.net
http://www.kamishima.net/

[2] NTT Communication Science Laboratories, NTT Corporation
Hikaridai 2–4, Seika, Soraku, Kyoto, 619–0237 Japan

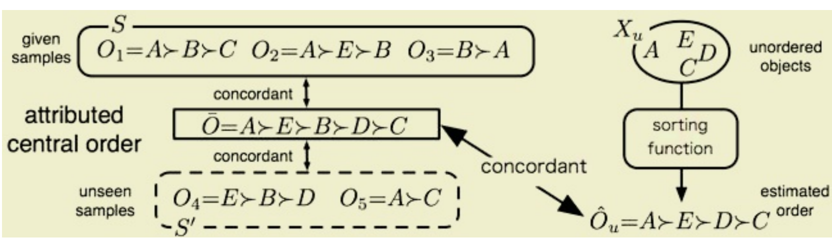
1 Introduction

Lists of ordered objects are widely used as representational forms. Such ordered objects include Web search results or best seller lists. In spite of their importance, the methods of processing orders have received little attention. However, research concerning object ordering is becoming more common. Some researchers have developed various methods to perform almost the same task: a learning function used for sorting objects from examples of ordered sequences. We call this task the estimation of *Attributed Central Orders* (*ACO* for short). The performance of this task is useful for sensory surveys[1], information retrieval, or decision making. We performed a survey of such methods, empirically compared the methods' properties, and discuss their merits and demerits.

Sections 2, 3, and 4, show the specifications of this task, describe the experimental results, and provide a summary, respectively.

2 Attributed Central Orders

This section describes the estimation task of ACOs. An order is a sorted sequence of objects according to a particular property, such as preference, size, or cost. An example of an order is $O_i{=}A{\succ}D{\succ}C$. These sorted objects are the members of the universal object set X^*. In the case of the figure below, $X^* = \{A, B, C, D, E\}$.



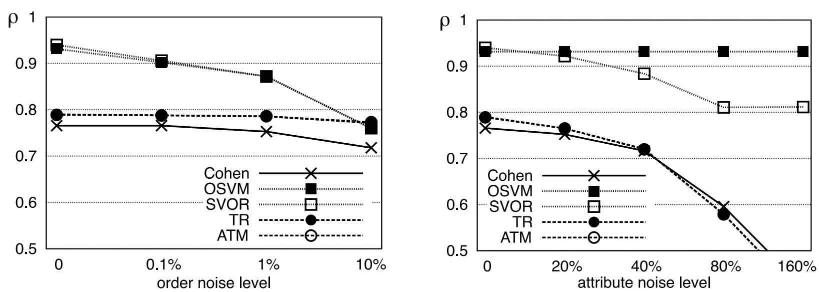
[1] The quantification of respondents' sensation or impression

J.-F. Boulicaut et al. (Eds.): ECML 2004, LNAI 3201, pp. 563–565, 2004.
© Springer-Verlag Berlin Heidelberg 2004

An ACO estimation task can be considered a regression task whose target variable is ordinal. A regression line corresponds to an ACO. Analogous to the case of a regression line, an ACO is estimated so as to be concordant with both given samples S and unseen samples S', to be generated (above figure left). Note that an ACO consists of all the objects in X^*. This task differs from a regression in two ways. First, the target variable is ordinal. Therefore, the discordance is measured by the distance between orders, such as Spearman's or Kendall's distance [1]. Further, an ACO is represented by a sorting function that sorts unordered objects so as to be concordant with the ACO. Second, almost all the samples are incomplete, that is, sample orders consist of subsets of X^*. Hence, there may be objects not observed in given samples (ex., D in the above figure). Such objects should be ranked under the assumption that the neighboring objects in the attribute space would be nearly ranked.

3 Experiments

We tested five methods for estimating ACOs. Cohen et al. proposed a method (Cohen) based on a preference function indicating which of two orders is ranked higher [2]. Kazawa et al. used the Order SVM (OSVM) designed to calculate the cumulative probability distribution [3]. Herbrich et al. proposed the SVM (SVOR) based on the ordered pair of objects [4]. Kamishima and Akaho combined a linear regression method with paired comparison methods [5] (TR), and Akaho and Kamishima [6] extended Thurstone's model to handle attributed objects (ATM).



We applied these methods to artificial data. The robustness against order and attribute noise was tested. Order noise is the permutation in sample orders, while attribute noise is the perturbation of attribute values. The left and right figures above show the depression of estimation accuracies in accordance with the increase of order and attribute noise level, respectively. The ρ represents the Spearman's rank correlation, which measures the concordance between the estimated and the true orders. The larger ρ indicates the more accurate estimation. The two SVM-based methods, OSVM and SVOR, were robust against attribute noise, but not against order noise. Order noise more greatly decreased the level of performance, because the interchanged ordered pairs tend to become support-vectors. The perturbation of attribute values does not affect the support-vectors to as great a degree, so these methods are robust against attribute noise. Conversely, the non-SVM-based methods were robust against permutations of orders, but not against perturbation of attribute values. These methods can learn correctly if correct orders constitute the majority of sample orders, thus these methods are robust against order noise. However, any perturbation in attribute values always affects their performance.

4 Discussion and Summary

We first discuss the time complexities. These methods' sorting complexities are comparable. The learning time of the SVM-based methods are slow, the Thurstone-based TR and ATM are intermediate, and the Cohen method is the fastest. Overall, the slower methods show higher levels of prediction accuracy. In the SVM-based case, the number of pairs in samples are limited to $10^5 \sim 10^6$. Though Thurstone-based methods can deal with a greater amount of data, the number of distinct objects in S is limited to $10^5 \sim 10^6$. The two SVM-based methods and the others are robust against different types of noise. Hence, for the data of which orders are permutated, the non-SVM-based methods are preferable, while for the data whose attributes are disturbed, the SVM-based methods are preferable. Prediction accuracy strongly depends on the fitness of the model bias for the target data. The lower bias model can be incorporated into non-SVM-methods, but the time complexity might increase. The SVM-based methods can use kernels of the higher-biases, but it is not feasible to reduce the time complexities. In our next study, we plan to explore the effects of tuning these model biases. In order to evaluate the methods' generalization abilities, we will perform tests involving another real data set in which $|X^*|$ is large.

Acknowledgments. This work is supported by the grants-in-aid 14658106 and 16700157 of the Japan society for the promotion of science.

References

1. Marden, J.I.: Analyzing and Modeling Rank Data. Volume 64 of Monographs on Statistics and Applied Probability. Chapman & Hall (1995)
2. Cohen, W.W., Schapire, R.E., Singer, Y.: Learning to order things. Journal of Artificial Intelligence Research **10** (1999) 243–270
3. Kazawa, H., Hirao, T., Maeda, E.: Order SVM: A kernel method for order learning based on generalized order statistic. The IEICE Trans. on Information and Systems, pt. 2 **J86-D-II** (2003) 926–933 (in Japanese) [An English version will appear in "Systems and Computers in Japan" Wiley Periodicals, Inc.].
4. Herbrich, R., Graepel, T., Bollmann-Sdorra, P., Obermayer, K.: Learning preference relations for information retrieval. In: ICML-98 Workshop: Text Categorization and Machine Learning. (1998) 80–84
5. Kamishima, T., Akaho, S.: Learning from order examples. In: Proc. of The IEEE Int'l Conf. on Data Mining. (2002) 645–648
6. Akaho, S., Kamishima, T.: A statistical approach for learning from order examples linear models. In: Proc. of the 16th Annual Conference of JSAI. (2002) (in Japanese).