Lyubchyk L.M., Galuza A.A., Grinberg G.L.
National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine

Ranking Model Real-Time Adaptation via Preference Learning Based on Dynamic Clustering

Learning to rank has become an important research topic in many information retrieval problems, such document retrieval, collaborative filtering, search engine modeling, e-commerce and other internet-related applications [1]. A typical example of learning to rank problem is a learning model adaptation for an unknown ranking mechanism of search engine based on the information about the search results of the ranked set of documents that match a particular sequence of queries [2]. So the training data consists of queries and ranked sequence of documents characterized by their features. One of the most effective approaches to solving these problems is preference learning, based on the recovery of a scalar continuous function in the objects feature space, known as a function of preferences or utility function, which is used to rank objects considered. It is assumed that a certain object is preferable to another if its feature vector corresponds to a larger value of preference function. Preference function can be considered as a model of unknown ranking function, describes, for example, search engine operation mechanism or user preferences. The problem of preference function recovery is extensively studied in relation to the development of multi-criteria decision-making methods [3]. In practice frequently used a simple model of preference functions as a linear combination of features with weights that determine their relative importance, but linear model not always adequately reflects the complex nature of preferences, arising in real ranking systems. Various approaches to the preference function nonlinear model construction are considered, at that the most difficult is the problem of the choice of its structure ensuring the possibility of preference high accuracy of approximation with restricted complexity of the model. A perspective method of preferences function identification used kernel-based machine learning technique with regularization based on expert information about the features weights [4]. At that the identified model may be represented in a nonparametric form as linear combination of kernel function, though the weighting coefficients may be computed without making direct reference to high-dimensional vectors of model coordinate functions. The difficulty of solving the problem of real time ranking learning upon receipt of the current information in data stream scheme is that the relative rank of objects with similar feature vectors in different data series may differ significantly. In this regard, a new approach to preference ranking learning is proposed, based on pre-clustering of the objects followed by preference function identification using the information of average ranks of objects belonging to each of the clusters [5]. Thus the problem is reduced to preference learning to rank on clusters in the feature space of ranked objects, while aggregated training dataset is formed from the centers of clusters and estimates of the average rank of the objects from the cluster. The proposed preference learning on clusters method allows to fully realizing the advantages of the kernel-based approach. While the dimension of the model is determined by a pre-selected number of clusters and its complexity do not grow with increasing number of observations. Thus real-time preference function identification algorithm based on training data stream includes successive estimates of cluster parameter as well as average cluster ranks updating and recurrent kernel-based nonparametric estimation of preference model [5].

References. **1.** T.Y. Liu. Learning to Rank for Information Retrieval. Foundations and Trends in Information Retrieval, Vol. 3, No.3, pp. 225–331, 2009. **2.** L. Hang, A. Short. Introduction to Learning to Rank. IEICE Trans. Inf. and Syst., Vol. E94-D, No.10, 2011. **3.** J. Figueira, S. Greco, M. Ehrgott. Multiple Criteria Decision Analysis: State of the Art Surveys. Springer, Switzerland, 2005. **4.** L. Lyubchyk, G. Grinberg. Preference Function Reconstruction for Multiple Criteria Decision Making Based on Machine Learning Approach. In "Recent Developments and New Directions in Soft Computing", L.A. Zadeh et al. (eds), Springer, Switzerland, 2014, pp. 53–63. **5.** L. Lyubchyk, G. Grinberg. Real Time Recursive Preference Learning to Rank from Data Stream. The 1th IEEE Int. Conf. on Data Stream Mining and Processing, 23–27 August 2016, Lviv, Ukraine.