Expert Systems with Applications 40 (2013) 3378–3388

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

# A multi-stage learning framework for intelligent system

Chang-Yi Kao *, Chin-Shyurng Fahn

*National Taiwan University of Science and Technology, Institute for Information Industry, Taiwan*

ARTICLE INFO

*Keywords:*
Human computer interaction
Boosting algorithm
Pair-wise ranking
Recommendation system
Speech recognition

ABSTRACT

As information technologies advance and user-friendly interfaces develop, the interaction between humans and computers, information devices, and new consumer electronics is increasingly gaining attention. One example that most people can relate to is Apple's innovation in human–computer interaction which has been used on many products such as iPod and iPhone. Siri, the intelligent personal assistant, is a typical application of machine-learning human–computer interaction.

Algorithms in machine learning have been employed in many disciplines, including gesture recognition, speaker recognition, and product recommendation systems. While the existing learning algorithms compute and learn from a large quantity of data, this study proposes an improved learning to rank algorithm named MultiStageBoost. In addition to ranking data through multiple stages, the MultiStageBoost algorithm significantly improves the existing algorithms in two ways. Firstly, it classifies and filters data to small quantities and applies the Boosting algorithm to achieve faster ranking performance. Secondly, it enhances the original binary classification by using the reciprocal of fuzzily weighted membership as the ranking distance.

The importance of data is revealed in their ranked positions. Usually data ranked in the front are given more attention than those ranked in the middle. For example, after ranking 10,000 pieces of data, the top 10, or at most 100, are the most important and relevant. Whether the data after the top ones are ranked precisely does not really matter. Due to this reason, this study has made improvement on the conventional methods of the pair-wise ranking approach. Not only are data classified and ranked binarily, they are also given different weights depending on whether they are concordant or discordant. Incorporating the concept of weighting into the ranking distance allows us to increase the precision of ranking. Results from experiments demonstrate that our proposed algorithm outperforms the conventional methods in three evaluation measures: P@n, MAP, and NDCG. MultiStageBoost was then applied to speech recognition. However, we do not aim to improve the technology of speech recognition, but simply hope to provide evidences that MultiStageBoost can be used in the classification and ranking in speech recognition. Experiments show that the recognition optimization procedures established by this study are able to increase the recognition rate to over 95% in the personal computing device and industrial personal computer. It is expected that in the future this voice management system will accurately and effectively identify speakers answering the voice response questionnaire and will successfully carry out the functions in the choice of answers, paving the way for the formation of a virtual customer service person.



## 1. Introduction

As more data are computed and stored in the clouds, the data we collect are getting larger in quantity, more complex in feature, and more diverse in type. Being overwhelmed by flood of information, a common goal of many researchers is to provide methods and strategies to simplify data. If the most important data can be extracted from the boundary less data universe, decision makers will be able to exploit the hidden knowledge and wisdom for their references. One of the common strategies in data simplification is to choose features that can be used to classify data, especially critical when data amount is enormous. After the features are determined, data can then be selected, classified, and computed. Machine learning is a discipline which focuses on processing raw data in order to retrieve useful information. It has been applied to human–computer interaction and information systems, making computers, information devices, and new consumer electronics smarter and people's lives more convenient.

Learning to rank is a research area which combines information retrieval and machine learning. The purpose of learning to rank is to let the computer automatically generate a ranking function based on the training data which consists of the perfect ranking lists of documents for each query. The performance of the ranking

* Corresponding author. Tel.: +886 986125010.
*E-mail addresses:* D9515011@mail.ntust.edu.tw, edenkao@iii.org.tw (C.-Y. Kao).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.12.046

function is evaluated by comparing it to the correct ranking results. However, the evaluation method used here is somewhat different from regular methods. Since the ranking function compares data in two lists, not only the data at the matched position in each list are compared (for example, pair-wise comparison of data at position n in list A and position n in list B), the relationships of similarity of the two lists must also be taken into consideration while the evaluation is performed.

The learning to rank algorithm proposed in this paper is developed from AdaBoost. AdaBoost pools together a series of rough weighting rules to create highly precise rules. Its advantages include easy execution and highly precise classification. However, it also has drawbacks. For instance, AdaBoost is sensitive to noise, which is a disadvantage shared by most ranking and classification algorithms. Further, its execution performance relies on the weak classifiers used. AdaBoost boasts high efficiency and detection rate, hence is suitable for handling high-dimensional feature data. Such Boosting-based ranking algorithm is very effective in human–computer interaction and is even applicable to recognition technologies.

## 2. Related works

Existing methods for learning to rank fall into three categories: point-wise, pair-wise, and list-wise. The point-wise approach allows the ranking function to learn the scores of the perfect ranking. Its purpose is to minimize each data's regression error between the ranking function and the perfect ranking. For instance, McRank (Li, Burges, & Wu, 2007) uses multi-class classification (Mc) to minimize the distance between the ranking function and the perfect ranking. The pair-wise approach transforms ranking into classification on data pairs. One of the data point is each pair is either better or worse than the other data point in the given pair, turning the comparison into a binary classification. As the pair-wise approach puts more emphasis on the correctness of ranking results, it is usually considered better than the point-wise approach. Ranking SVM (Joachims, 2002) is a good example which uses SVM (Support Vector Machines) to efficiently performing binary classification for data ranking. Other examples include RankBoost (Freund et al., 1998), an algorithm which uses the boosting approach to minimize ranking errors, and Frank (Tsai et al., 2007), which minimizes fidelity loss. The list-wise approach, which emerged quite recently, minimizes the loss function defined on the perfect ranked list and the list generated by the ranking function. Representative algorithms include ListNet (Cao et al., 2007), which computes based on probability models, and AdaRank (Xu & Li, 2007), which uses the boosting approach to minimize the loss function. Learning to rank is a supervised machine learning problem, especially in a vector space (Liu, 2008).

Unsupervised learning refers to the problem of trying to find hidden structure in training data that are not artificially labeled. In other words, an unsupervised learning model is not given labeled data during training. Therefore, it can only depend on the existing methods without reward signals to evaluate a potential solution. Among the applications of unsupervised learning algorithms, most summarization models can be divided into two categories based on matching strategies. One of them is to measure the relevance between each sentence and the document by literal term matching, as exampled by the Vector Space Model (VSM). Higher relevance correlates to higher importance. As opposed to unsupervised learning, supervised learning is a machine learning task that is given labeled data during training so that it infers a function from the labeled training data. Text summarization can be made into a binary classification problem. Each sentence s is determined whether it belongs to the summary set S or not. As a result, many lexical features and sentence structural information can be used in the training of supervised summarization models. In recent years, the application of learning to rank in information retrieval can be categorized into three approaches: point-wise training, pair-wise training, and list-wise training.

The rationales of the three training approaches are described below.

### 2.1. Point-wise

The point-wise approach approximates the ranking problem by a regression or views it as a classification problem.

By applying the point-wise approach to automatic summarization, each sentence in the document is a unit in the training data. Through function computation, each sentence is given a score as a label indicating its probability of belonging to the summary. The figure below shows how the point-wise approach is applied to automatic voice summarization. Each sentence in the document is given various features after computation. Each feature as to each document can be represented with a vector. Each vector, based on the known matched relationships in the training data, assigns a label to each sentence. If a sentence belongs to the summary, it is labeled 1 (positive); if not, then it is labeled $-1$ (negative). After the learning process, the testing data is ranked based on the trained ranking model. The final ranking results indicate the level of importance of each sentence in the summary.

### 2.2. Pair-wise

The pair-wise approach can be seen as a classification, regression, or ordinal Minimum ranking function regression problem.

The application of the pair-wise approach in information retrieval first appeared in 1995 and has been applied to experiments in medical diagnosis. In 2002, Joachims proposed Ranking SVM based on ordinal regression and it marked the first time in which ordinal regression is used for information retrieval.

### 2.3. List-wise

The list-wise approach appeared later than point-wise and pair-wise approaches. In 2007, two papers about the application of the list-wise approach in information retrieval were presented in the SIGIR conference. Yue et al. proposed SVM-MAP, an algorithm derived from Ranking SVM. Xu et al. proposed another algorithm AdaRank, a novel learning method derived from RankBoost, allowing retrieval models to optimize performance measures such as MAP (Mean Average Precision) and NDCG (Normalized Discount Cumulative Gain).

Literally speaking, the pair-wise approach gives labels to the ranked paired sentences in the training data. In the list-wise approach, lists of ranked sentences in the training data are labeled and used as 'instances' in learning. Both approaches use the minimum ranking error for approximation.

They are boosting, learning to rank, and direct optimization of performance measures. Typical methods of the approach include Ranking SVM (Herbrich, Graepel, & Obermayer, 2000), RankBoost (Freund, Iyer, Schapire, & Singer, 2003), RankNet (Burges et al., 2005), and some other approaches to learning to rank (Yang, Kriegman, & Ahuja, 2002). AdaRank is also one that tries to directly optimize multivariate performance measures. AdaRank is unique in that it employs an exponential loss function boosting technique. AdaRank is a simple yet powerful method. More importantly, it is a method that can be justified from the theoretical viewpoint. The following description was the algorithm (Xu & Li, 2007).

Connector manufacturers have for years resorted to a closer supply chain link with downstream electronics makers to enhance