Computational Statistics and Data Analysis 121 (2018) 113–136

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Angle-based models for ranking data

Hang Xu [a], Mayer Alvo [b], Philip L.H. Yu [a,*]

[a] *Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong*
[b] *Department of Mathematics and Statistics, University of Ottawa, Canada*

ARTICLE INFO

Article history:
Received 30 June 2017
Received in revised form 6 December 2017
Accepted 9 December 2017
Available online 27 December 2017

Keywords:
Ranking data
Bayesian variational inference
Incomplete ranking

ABSTRACT

A new class of general exponential ranking models is introduced which we label angle-based models for ranking data. A consensus score vector is assumed, which assigns scores to a set of items, where the scores reflect a consensus view of the relative preference of the items. The probability of observing a ranking is modeled to be proportional to its cosine of the angle from the consensus vector. Bayesian variational inference is employed to determine the corresponding predictive density. It can be seen from simulation experiments that the Bayesian variational inference approach not only has great computational advantage compared to the traditional MCMC, but also avoids the problem of overfitting inherent when using maximum likelihood methods. The model also works when a large number of items are ranked which is usually an NP-hard problem to find the estimate of parameters for other classes of ranking models. Model extensions to incomplete rankings and mixture models are also developed. Real data applications demonstrate that the model and extensions can handle different tasks for the analysis of ranking data.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Ranking data are often encountered in practice when judges (or individuals) are asked to rank a set of t items, which may be political goals, candidates in an election, types of food, etc. We see examples in voting and elections, market research and food preference just to name a few.

Alvo and Cabilio (1991) considered tests of hypotheses related to problems of trend and independence using only the ranks of the data. In another direction, the interest may be in modeling the ranking data. Some of these models are: (i) order-statistics models (Thurstone, 1927; Yu, 2000), (ii) distance-based models (Critchlow et al., 1991; Lee and Yu, 2012), (iii) paired-comparison models (Mallows, 1957), and (iv) multistage models (Fligner and Verducci, 1988). A more comprehensive discussion on these probability ranking models can be found in the book by Alvo and Yu (2014). However, some of these models cannot handle the situation in which the number of items being ranked is large, nor when incomplete rankings exist in the data. For distance-based models: (i) there is no closed-form for the normalizing constants for Spearman distances and (ii) the modal ranking is discrete over a finite space of $t!$ dimensions and searching for it will be time consuming when the number of items, t, becomes large.

In this article, we first propose a new class of general exponential ranking models called angle-based models for the distribution of rankings. We assume a consensus score vector $\boldsymbol{\theta}$ which assigns scores to the items, where the scores reflect a consensus view of the relative preference of the items. The probability of observing a ranking is proportional to the cosine of the angle from the consensus score vector. The distance-based model with Spearman distance can be seen as a special case of our model. Unlike the Spearman distance-based model, we obtain a very good approximation of the normalizing constant

* Corresponding author.
E-mail address: plhyu@hku.hk (P.L.H. Yu).

https://doi.org/10.1016/j.csda.2017.12.004
0167-9473/© 2017 Elsevier B.V. All rights reserved.

of our angle-based model. Note that this approximation allows us to have the explicit form in the first or second derivative of normalizing constant which can facilitate the computation of the ranking probabilities under the model.

For the parameter estimation of the model, we first place a joint Gamma-von Mises–Fisher prior distribution on the parameter. We describe several mathematical difficulties incurred in determining the resulting posterior distribution and propose to make use of the variational inference method. From the simulation experiments, it can be seen that the Bayesian variational inference approach not only has great computational advantage compared to the traditional Markov Chain Monte Carlo (MCMC), but also avoids the over-fitting problem in maximum likelihood estimation (MLE). Our model also works when the number of items being ranked is large, while it is usually an NP-hard problem to obtain the parameter estimates for other classes of ranking models. Model extensions to the incomplete rankings and mixture model are also discussed. From the simulations and applications, it can be seen that our extensions can handle well incomplete rankings as well as the clustering and classification tasks for ranking data.

The article is organized as follows. Section 2 introduces the angle-based model as well as the Bayesian MCMC approach. In Section 3, we describe the method of variational inference for our model and derive the predictive density of a new ranking. In Section 4, we consider model extensions to incomplete rankings and mixture models for clustering and classification. In Section 5, we describe several simulation experiments whereas in Section 6, the methodology is then applied to real data sets including a sushi data set, ranking data from the American Psychological Association (APA) presidential election of 1980 and a breast cancer gene expressions data set. We conclude with a discussion in Section 7.

2. Angle-based models

2.1. Model setup

A ranking $\boldsymbol{R}$ represents the order of preference with respect to a set of items. In ranking t items, labeled $1, \ldots, t$, a ranking $\boldsymbol{R} = (R(1), \ldots, R(t))^T$ is a mapping function from $1, \ldots, t$ to ranks $1, \ldots, t$, where $R(2) = 3$ means that item 2 is ranked third and $R^{-1}(3) = 2$ means that the item ranked third is item 2. It will be more convenient to standardize the rankings as:

$$\boldsymbol{y} = \frac{\boldsymbol{R} - \frac{t+1}{2}}{\sqrt{\frac{t(t^2-1)}{12}}},$$

where $\boldsymbol{y}$ is the $t \times 1$ vector with $\|\boldsymbol{y}\| = 1$.

We consider the following ranking model:

$$p(\boldsymbol{y}|\kappa, \boldsymbol{\theta}) = C(\kappa, \boldsymbol{\theta}) \exp\left\{\kappa \boldsymbol{\theta}^T \boldsymbol{y}\right\},$$

where the parameter $\boldsymbol{\theta}$ is a $t \times 1$ vector with $\|\boldsymbol{\theta}\| = 1$, parameter $\kappa \geq 0$, and $C(\kappa, \boldsymbol{\theta})$ is the normalizing constant. In the case of the distance-based models (Alvo and Yu, 2014), the parameter $\boldsymbol{\theta}$ can be viewed as if a modal ranking vector. In fact, if $\boldsymbol{R}$ and $\boldsymbol{\pi}_0$ represent an observed ranking and the modal ranking of t items respectively, then the probability of observing $\boldsymbol{R}$ under the Spearman distance-based model is proportional to

$$\begin{aligned} \exp\left\{-\lambda\left(\frac{1}{2}\sum_{i=1}^{t}\left(R\left(i\right)-\boldsymbol{\pi}_0\left(i\right)\right)^2\right)\right\} &= \exp\left\{-\lambda\left(\frac{t\left(t+1\right)\left(2t+1\right)}{12}-\boldsymbol{\pi}_0^T\boldsymbol{R}\right)\right\} \\ &\propto \exp\left\{\kappa\boldsymbol{\theta}^T\boldsymbol{y}\right\}, \end{aligned}$$

where $\kappa = \lambda\frac{t(t^2-1)}{12}$, and $\boldsymbol{y}$ and $\boldsymbol{\theta}$ are the standardized rankings of $\boldsymbol{R}$ and $\boldsymbol{\pi}_0$ respectively. However, the $\boldsymbol{\pi}_0$ in the distance-based model is a discrete permutation vector of integers $\{1, 2, \ldots, t\}$ but the $\boldsymbol{\theta}$ in our model is a real-valued vector, representing a consensus view of the relative preference of the items from the individuals. Since both $\|\boldsymbol{\theta}\| = 1$ and $\|\boldsymbol{y}\| = 1$, the term $\boldsymbol{\theta}^T\boldsymbol{y}$ can be seen as $\cos\phi$ where ϕ is the angle between the consensus score vector $\boldsymbol{\theta}$ and the observation $\boldsymbol{y}$. Fig. 1 illustrates an example of the angle between the consensus score vector $\boldsymbol{\theta} = (0, 1, 0)^T$ and the standardized observation of $\boldsymbol{R} = (1, 2, 3)^T$ on the sphere for $t = 3$. The probability of observing a ranking is proportional to the cosine of the angle from the consensus score vector. The parameter κ can be viewed as a concentration parameter. For small κ, the distribution of rankings will appear close to a uniform whereas for larger values of κ, the distribution of rankings will be more concentrated around the consensus score vector.

To compute the normalizing constant $C(\kappa, \boldsymbol{\theta})$, let P_t be the set of all possible permutations of the integers $1, \ldots, t$. Then

$$(C(\kappa, \boldsymbol{\theta}))^{-1} = \sum_{\boldsymbol{y}\in \mathrm{P}_t} \exp\left\{\kappa\boldsymbol{\theta}^T\boldsymbol{y}\right\}. \tag{1}$$

Notice that the summation is over $t!$ elements in P_t. When t is large, say greater than 15, the exact calculation of the normalizing constant is prohibitive. Using the fact that the set of $t!$ permutations lie on a sphere in $(t-1)$-space, our model resembles the continuous von Mises–Fisher distribution, abbreviated as $vMF(\boldsymbol{x}|\boldsymbol{m}, \kappa)$, which is defined on a $(p-1)$ unit sphere with mean direction $\boldsymbol{m}$ and concentration parameter κ:

$$p(\boldsymbol{x}|\kappa, \boldsymbol{m}) = V_p(\kappa)\exp(\kappa\boldsymbol{m}^T\boldsymbol{x}),$$

Download English Version:

https://daneshyari.com/en/article/6868803

Download Persian Version:

https://daneshyari.com/article/6868803

Daneshyari.com