2019 IEEE INTERNATIONAL WORKSHOP ON MACHINE LEARNING FOR SIGNAL PROCESSING, OCT. 13–16, 2019, PITTSBURGH, PA, USA

A WEIGHTED ORDERED PROBIT COLLABORATIVE KALMAN FILTER FOR HOTEL RATING PREDICTION

Myrsini Ntemi, Constantine Kotropoulos, and Emmanouil Gionanidis

Department of Informatics, Aristotle University of Thessaloniki
Thessaloniki 54124, GREECE
Email: nmyrsini@csd.auth.gr, costas@aiia.csd.auth.gr, gionanie@csd.auth.gr

ABSTRACT

A successful recommender system interacts with users and learns their preferences. This is crucial in order to provide accurate recommendations. In this paper, a Weighted Ordered Probit Collaborative Kalman filter is proposed for hotel rating prediction. Since potential changes may occur in hotel services or accommodation conditions, a hotel popularity may be volatile through time. A weighted ordered probit model is introduced to capture this latent trend about each hotel popularity through time. It is demonstrated by experiments that such model of hotel popularity trends reinforces the performance of Collaborative Kalman filter, yielding more accurate potential recommendations.

Index Terms— recommender systems, ordered probit model, hotel rating prediction

1. INTRODUCTION

The goal of a recommender system is to provide efficient personalized recommendations to its users. Over the past years, collaborative filters (CFs) have been widely used in recommender systems. The input to CF is a dyad, i.e., a user-item pair. The objective is to anticipate the ratings that will be given to items from users with similar taste [1]. This procedure corresponds to information filtering by utilizing the assessments of other individuals. The key idea is that users, who concurred in their assessment of specific items in the past, are probably going to concur again in future assessments.

One of the major challenges in CFs is the missing value problem, which results in the construction of incomplete data matrices. Matrix factorization (MF) techniques have been proposed to deal with matrix completion [2], [3], [4]. Probabilistic versions of these methods have been also proposed in order to estimate the distribution of missing entries [5], [6]. According to these methods, the ratings are approximated via the dot products of the latent feature vectors of users and items. Other methods model a matrix as the product of two latent matrices, one for all users and one for all items [7]. Many recent state-of-the-art methods utilize deep learning frameworks in MF CFs, such as convolutional networks to infer user and item latent features [8], [9].

However, the majority of the aforementioned methods treat the user - item interactions as stationary processes. On the contrary, the collaborative Kalman filter (CKF) assumes that the user and item latent vectors evolve through time and models the dynamic information lying in their interactions [10]. A pair of user and item latent vectors forms a dyad. According to CKF, the prediction of integer valued ratings is conducted by dividing the real line into as many regions as the number of rating stars. The rating prediction is conducted through the calculation of the probability each dyad belongs to a specific region. This assumption is related to ordered probit regression, which suggests that the dependent variable (i.e., the dyad) takes potential values that have a natural ordering [11].

Here, a CKF with a weighted ordered probit assumption (coined as WCKF) is introduced for hotel rating prediction. Throughout the paper, the hotel latent vectors are treated as item latent vectors. The magnitude of the regions in the ordered probit model is assumed to evolve through time with respect to each hotel popularity evolution. Specifically, a time window over the past observations is defined and the mean value of the ratings for a specific hotel is computed. The region of the ordered probit model, where the mean value falls in, increases. This is done by controlling the magnitude assigned to each region. The magnitude corresponds to the standard deviation of a truncated normal distribution utilized for the inference of the posterior distribution of user and hotel latent vectors. This assumption enriches CKF with the ability to capture the evolution of each hotel popularity through time, yielding more accurate predictions.

The outline of the paper is as follows. In Section 2, CKF is briefly reviewed. In Section 3, WCKF is introduced. Experimental results are demonstrated in Section 4, and conclusions are drawn in Section 5.

Acknowledgement: This research has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project code:T1EDK-02474).

2. COLLABORATIVE KALMAN FILTER

2.1. Matrix Factorization

CKF is inspired from MF techniques [10]. According to MF techniques, a set of M users and a set of N hotels are organized into matrices $\mathbf{U} \in \mathbb{R}^{K\times M}$ and $\mathbf{H} \in \mathbb{R}^{K\times N}$, respectively. Here, K is the dimension of the latent space providing a balance between prediction performance and overfitting. The ith column of matrix $\mathbf{U}$, $\mathbf{u}_i$, represents the location of user i in a latent space and the jth column of matrix $\mathbf{H}$, $\mathbf{h}_j$, represents the location of hotel j in the same latent space. An incomplete matrix $\mathbf{Z} \in \mathbb{R}^{M\times N}$ contains all the ratings the users have assigned to hotels. That is, z_{ij} is the rating user i has given to hotel j. The ratings are modeled through

$$\mathbf{Z} \approx \mathbf{U}^T\mathbf{H}. \tag{1}$$

Matrices $\mathbf{U}$ and $\mathbf{H}$ have to be learned. The probabilistic extension of MF suggests a probabilistic approximation of z_{ij}. It is assumed that the probability density function (pdf) of observations z_{ij} is Gaussian, i.e.,

$$p(z_{ij}|\mathbf{u}_i, \mathbf{h}_j, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{1}{2\sigma^2}\big(z_{ij} - \mathbf{u}_i^T\mathbf{h}_j\big)^2\right\} \tag{2}$$

where σ denotes the standard deviation. The negative log likelihood of the observations, assumed as independent identically distributed (i.i.d.) random variables, is given by

$$\begin{aligned} -\log\mathcal{L} &= |\mathcal{S}_{\mathbf{u}_i}|\left\{\log\sigma + \frac{1}{2}\log(2\pi)\right\} \\ &\quad + \frac{1}{2\sigma^2}\sum_{j\in\mathcal{S}_{\mathbf{u}_i}} \big(z_{ij} - \mathbf{u}_i^T\mathbf{h}_j\big)^2 \end{aligned} \tag{3}$$

where $\mathcal{S}_{\mathbf{u}_i}$ denotes the set containing all hotels user i has rated. To minimize Eq.(3), the gradient of log likelihood with respect to $\mathbf{u}_i$ is set equal to zero, i.e.,

$$\begin{aligned} &\frac{\partial \log\mathcal{L}}{\partial \mathbf{u}_i} = 0 \Rightarrow \sum_{j\in\mathcal{S}_{\mathbf{u}_i}} (z_{ij} - \mathbf{u}_i^T\mathbf{h}_j)(\mathbf{h}_j) = 0 \Leftrightarrow \\ &\sum_{j\in\mathcal{S}_{\mathbf{u}_i}} z_{ij}\mathbf{h}_j = \sum_{j\in\mathcal{S}_{\mathbf{u}_i}} \mathbf{h}_j\mathbf{h}_j^T\mathbf{u}_i. \end{aligned} \tag{4}$$

Solving (4) for a user latent vector $\mathbf{u}_i$, we obtain

$$\mathbf{u}_i = \Bigg(\sum_{j\in\mathcal{S}_{\mathbf{u}_i}} \mathbf{h}_j\mathbf{h}_j^T\Bigg)^{-1}\Bigg(\sum_{j\in\mathcal{S}_{\mathbf{u}_i}} z_{ij}\mathbf{h}_j\Bigg). \tag{5}$$

A similarly update can be obtained for hotel latent vector $\mathbf{h}_j$. The major problem in MF based models is that latent vectors do not evolve through time and they are unable to capture temporal information. A solution to this problem is provided by CKF, which assumes that latent vectors and their respective probability distributions are functions of time.

2.2. Collaborative Kalman Filter

Let $\mathbf{u}_i[t] \in \mathbb{R}^K$ and $\mathbf{h}_j[t] \in \mathbb{R}^K$ denote the latent vectors of user i and hotel j at time t, respectively, which are assumed to be Gaussian random vectors. Since CKF is a probabilistic model, at every time step t, the prior distributions of users and hotels latent vectors are approximated as

$$\mathbf{u}_i[t] \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{u}_i}[t], \boldsymbol{\Sigma}_{\mathbf{u}_i}[t]), \tag{6}$$

$$\mathbf{h}_j[t] \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{h}_j}[t], \boldsymbol{\Sigma}_{\mathbf{h}_j}[t]) \tag{7}$$

where the prior mean vector and the prior covariance matrix are given by

$$\boldsymbol{\mu}[t] = \boldsymbol{\mu}'[t-1] \tag{8}$$

$$\boldsymbol{\Sigma}[t] = \boldsymbol{\Sigma}'[t-1] + \mathbf{I}\Delta^{[t]} \tag{9}$$

with $\boldsymbol{\mu}'[t-1] \in \mathbb{R}^K$ and $\boldsymbol{\Sigma}'[t-1] \in \mathbb{R}^{K\times K}$ denoting the posterior mean vector and posterior covariance matrix at $t-1$. In (8), $\Delta^{[t]}$ is the time elapsed since the previous observation, and $\mathbf{I}$ is the identity matrix. After having measured the rating $z_{ij}[t]$, the posterior distributions of the latent vectors are

$$\mathbf{u}_i'[t] \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{u}_i}'[t], \boldsymbol{\Sigma}_{\mathbf{u}_i}'[t]), \tag{10}$$

$$\mathbf{h}_j'[t] \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{h}_j}'[t], \boldsymbol{\Sigma}_{\mathbf{h}_j}'[t]). \tag{11}$$

The posterior distributions are estimated via variational inference [10] [12]. Specifically, given that the true posterior distribution is $p(\cdot)$, an approximate posterior distribution $q(\cdot)$ should be defined. This is accomplished through the minimization of the Kullback-Leibler (KL) divergence

$$KL(q||p) = \mathbb{E}_q\left[\log\frac{q}{p}\right] \tag{12}$$

where $\mathbb{E}_q[]$ is the expectation with respect to pdf q. In order to calculate the approximate distributions, the optimal parameters of $q(\cdot)$ should be found through a coordinate ascent update [13].

3. WEIGHTED ORDERED PROBIT CKF

Let us assume that the ratings are real numbers. Then, the line of real numbers $\mathbb{R}$ is partitioned into as many regions as the number of rating stars. Let us consider the case of tripadvisor. Since tripadvisor utilizes a five-star rating system, the real line is partitioned into five regions. Let the kth region be the interval

$$\mathcal{I}_k = (l_k, r_k] \tag{13}$$

where $l_k < r_k$, $l_k = r_{k-1}$, $r_k = l_{k+1}$, and $k = 1, 2, \ldots, 5$. These regions are ordered and their width is prefixed to σ [10]. This assumption is related to ordered probit regression, where the values of the dependent variable, (i.e., rating), have a natural ordering. The goal is to estimate the probability of the dependent variable falling into a specific region.

The weighted version of the model, suggests that instead of assuming the width of these regions to be the same and fixed, a dynamically changing magnitude is defined with respect to a specific criterion. That is, a sliding window $\mathcal{W}$ is defined over the time series of star ratings with length $|\mathcal{W}|$. Let

$$\mu_{z_j}[t] = \frac{\sum_{t\in\mathcal{W}} z_j[t]}{|\mathcal{W}|} \tag{14}$$

be the average rating that all users in the time window $\mathcal{W}$ have assigned to hotel j. Accordingly, the subscript corresponding to users has been suppressed. If $\mu_{z_j}[t]$ falls in region $\mathcal{I}_k$, i.e. $\mu_{z_j}[t] \in (l_k, r_k], k = 1, 2, \ldots, 5$, the magnitude of this region defined as the updated standard deviation of the weighted ordered probit model, becomes

$$\sigma^* = \begin{cases} \sigma\ c, & \text{if } \mu_{z_j}[t] \in \mathcal{I}_k \\ \sigma & \text{otherwise} \end{cases} \tag{15}$$

where c is a fixed constant weight, whose optimal value is estimated heuristically. The sliding window $\mathcal{W}$ tracks the evolution of each hotel popularity. This information is critical, since the services and operating principles of a hotel might change through time. The mapping $\mu_{z_j}[t] \mapsto \mathcal{I}_k$ encloses the just described information and results in weighting the respective potential rating. The magnitude defined in Eq.(15) corresponds to the updated standard deviation of the whole model utilized in variational inference and posterior distribution updates.

Let us now introduce an auxiliary normally distributed latent variable $s_{ij}[t]$ related to the output estimation at time t

$$s_{ij}[t] \mid \mathbf{u}_i, \mathbf{h}_j \sim \mathcal{N}(\mathbf{u}_i[t]^T \mathbf{h}_j[t], \sigma^{*2}). \tag{16}$$

Its mean is equal to the inner product of the latent vectors $\mathbf{u}_i$ and $\mathbf{h}_j$, and its standard deviation equals to σ^*, given by Eq.(15).

An approximate distribution should also be defined for s_{ij}, which can be accomplished through the exponentiation of the expected log joint likelihood

$$\begin{aligned} q(s_{ij}[t]) \quad \propto \quad & \exp\{\ln p(z_{ij}[t] \mid s_{ij}[t]) \\ & + \mathbb{E}_q[\ln p(s_{ij}[t] \mid \mathbf{u}_i[t], \mathbf{h}_j[t])\}. \end{aligned} \tag{17}$$

After having introduced the auxiliary latent variable $s_{ij}[t]$, the distribution of rating $z_{ij}[t]$ reads

$$p(z_{ij}[t] \mid s_{ij}[t]) = \mathbb{I}(s_{ij}[t] \in \mathcal{I}_{z_{ij}[t]}) \tag{18}$$

where $\mathbb{I}(\cdot)$ denotes an indicator function. Eq.(18) suggests that if the rating $z_{ij}[t]$ belongs to the region $\mathcal{I}_{z_{ij}[t]}$, the approximate distribution $q(s_{ij}[t])$ is defined in the same interval. This approximate distribution becomes

$$q(s_{ij}[t]) = \mathcal{TN}_{\mathcal{I}_{z_{ij}[t]}}(s_{ij}[t] \mid \mathbb{E}_q[\mathbf{u}_i^T[t]]\mathbb{E}_q[\mathbf{h}_j[t]], \sigma^{*2}) \tag{19}$$

where $\mathcal{TN}_{\mathcal{I}_{z_{ij}[t]}}(s_{ij}[t] \mid \cdot, \cdot)$ denotes the truncated normal distribution with mean $\mathbb{E}_q\left[\mathbf{u}_i^T[t]\right]\mathbb{E}_q\left[\mathbf{h}_j[t]\right]$ and variance σ^{*2}. Let the mean value of $q(s_{ij}[t])$ be

$$\hat{z}_{ij}[t] = \mathbb{E}_q[\mathbf{u}_i^T[t]]\mathbb{E}_q[\mathbf{h}_j[t]]. \tag{20}$$

Eq.(20) defines the prediction for z_{ij}. The estimation of the auxiliary latent variable $s_{ij}[t]$ relies on the region $\mathcal{I}_k$ of the truncated normal distribution to which its expected value belongs to. Let us assume that the aforementioned region is constructed by two boundaries with respect to the rating $z_{ij}[t]$, i.e., the left one $l_{z_{ij}[t]}$ and right one $r_{z_{ij}[t]}$. The following intervals are defined

$$\begin{aligned} \zeta_{ij}[t] \quad &= \quad \frac{l_{z_{ij}[t]} - \hat{\mu}_{ij}[t]}{\sigma^*} \\ \xi_{ij}[t] \quad &= \quad \frac{r_{z_{ij}[t]} - \hat{\mu}_{ij}[t]}{\sigma^*} \end{aligned} \tag{21}$$

where $\hat{\mu}_{ij}[t] = \mathbb{E}_q[\mathbf{u}_i'^T[t]]\mathbb{E}_q[\mathbf{h}_j'[t]]$, which is the inner product of the expectations of the posterior latent vectors. Finally, the expected value of $s_{ij}[t]$ is [10]

$$\mathbb{E}_q[s_{ij}[t]] = \hat{\mu}_{ij}[t] + \sigma^* \frac{\phi(\zeta_{ij}[t]) - \phi(\xi_{ij}[t])}{\Phi(\xi_{ij}[t]) - \Phi(\zeta_{ij}[t])} \tag{22}$$

where $\phi(\cdot)$ stands for pdf and $\Phi(\cdot)$ for cumulative distribution function (cdf) of a standard normal distribution $\mathcal{N}(0, 1)$.

The last step is the calculation of the optimal approximate distributions of latent vectors of user $\mathbf{u}_i$ and hotel $\mathbf{h}_j$, which are also multivariate Gaussians. The optimal parameters of the approximate distribution of user $\mathbf{u}_i$ at time step t are [10]

$$\begin{aligned} \mathbf{\Sigma}'_{\mathbf{u}_i}[t] \quad &= \quad \left(\mathbf{\Sigma}^{-1}_{\mathbf{u}_i}[t] + \frac{\boldsymbol{\mu}'_{\mathbf{h}_j}[t]\ \boldsymbol{\mu}'^T_{\mathbf{h}_j}[t] + \mathbf{\Sigma}'_{\mathbf{h}_j}[t]}{\sigma^{*2}}\right)^{-1} \\ \boldsymbol{\mu}'_{\mathbf{u}_i}[t] \quad &= \quad \mathbf{\Sigma}'_{\mathbf{u}_i}[t]\left(\frac{\mathbb{E}_q[s_{ij}[t]]\boldsymbol{\mu}'_{\mathbf{h}_j}[t]}{\sigma^{*2}} + \mathbf{\Sigma}^{-1}_{\mathbf{u}_i}[t]\ \boldsymbol{\mu}_{\mathbf{u}_i}[t]\right) \end{aligned} \tag{23}$$

Similarly, the optimal parameters of the approximate distribution of hotel $\mathbf{h}_j$ at time step t are

$$\begin{aligned} \mathbf{\Sigma}'_{\mathbf{h}_j}[t] \quad &= \quad \left(\mathbf{\Sigma}^{-1}_{\mathbf{h}_j}[t] + \frac{\boldsymbol{\mu}'_{\mathbf{u}_i}[t]\ \boldsymbol{\mu}'^T_{\mathbf{u}_i}[t] + \mathbf{\Sigma}'_{\mathbf{u}_i}[t]}{\sigma^{*2}}\right)^{-1} \\ \boldsymbol{\mu}'_{\mathbf{h}_j}[t] \quad &= \quad \mathbf{\Sigma}'_{\mathbf{h}_j}[t]\left(\frac{\mathbb{E}_q[s_{ij}[t]]\boldsymbol{\mu}'_{\mathbf{u}_i}[t]}{\sigma^{*2}} + \mathbf{\Sigma}^{-1}_{\mathbf{h}_j}[t]\ \boldsymbol{\mu}_{\mathbf{h}_j}[t]\right) \end{aligned} \tag{24}$$

These parameters will constitute the prior parameters of the next time step $t + 1$.

4. EXPERIMENTAL RESULTS

WCKF was applied to tripadvisor hotel rating prediction. A dataset was created from scratch by employing a Python scrapper to collect hotel data from tripadvisor. 16 Greek destinations were monitored. The data were ordered with respect to time, i.e., from the older assignments to the more recent ones. The following parameters are recorded: hotel id, user id, rating stars (1-5), and date. The dimensionality of the latent state vectors of users $\mathbf{u}_i$ and hotels $\mathbf{h}_j$ was set to $K = 5$, since performance did not improve further for $K = 10, 15, 40$. The initial standard deviation was set to $\sigma = 1.76$ [10] in order to compare WCKF and CKF on the same basis. After multiple trials, the best performing window length was $|\mathcal{W}| = 10$. The best performing weight c was found through leave-one destination-out validation. Specifically, 15 experiments were conducted for Preveza destination with $c = 1, 2, \ldots, 14$ and $c = 20$ and the root-mean-square error (RMSE) in number of stars was utilized as figure of merit. The best performing parameter was $c = 10$ and the minimum respective RMSE $= 0.7865$. Then, the remaining destinations were tested using the same parameters. Since WCKF is an online method, at every time step t, predictions are conducted for the next time step $t + 1$.

The performance of the WCKF is summarized in detail in Table 1. The first column shows 15 Greek destinations.

Table 1: Hotel rating prediction performance.

Destination	**RMSE WCKF**	**RMSE CKF**	**number of users**	**number of hotels**	**number of ratings**
Anatoliki Thraki	**0.6479**	1.4122	5,593	110	6,762
Athens	**0.6712**	1.4691	5,1476	382	56,056
Argolida	**0.7529**	1.598	24,276	327	26,687
Crete	**0.7447**	1.5782	326,153	3,204	386,634
Ioannina	**0.855**	1.7677	15,819	286	18,541
Kefalonia	**0.7895**	1.7611	44,221	575	50,706
Magnisia	**0.791**	1.637	13,598	439	15,420
Messinia	**0.7341**	1.5638	5,694	59	6,218
Mykonos	**0.8051**	1.6956	80,216	513	89,668
Naxos	**0.8191**	1.7005	31,583	494	34,374
Paros	**0.8004**	1.7007	32,087	462	35,034
Rhodes	**0.6875**	1.5558	9,058	50	9,700
Santorini	**0.8204**	1.7134	170,187	1,097	196,536
Skiathos	**0.7518**	1.5747	6,177	137	6,696
Thessaloniki	**0.6714**	1.4595	26,119	254	30,517

The second column shows the performance of the proposed WCKF, the third column summarizes the performance of the Collaborative Kalman Filter [10], the fourth column refers to the number of users who rated the hotels at each destination, the fifth column shows the respective number of hotels, while the sixth column summarizes the total number of ratings for each destination. The experimental results demonstrate that the proposed WCKF outperformed the method in [10] in all cases. The integration of weighted ordered probit model within CKF, reinforced strongly the prediction performance of the algorithm.

Next, F-tests were applied in order to evaluate if the differences in RMSE are statistically significant. The test statistic for WCKF - CKF is

$$\mathcal{F}_{1,2} = \frac{w_1^2}{w_2^2} \tag{25}$$

where the subscripts 1 and 2 refer to the proposed WCKF, and CKF respectively. For each model

$$w^2 = \frac{1}{N-1} \sum_{l=1}^{N} (\hat{z}_{ij}[l] - \bar{z})^2 \tag{26}$$

where $\hat{z}_{ij}^{[l]}$ denotes the l^{th} predicted value for z_{ij}, and $\bar{z} = \frac{1}{N} \sum_{l=1}^{N} \hat{z}_{ij}$ is the mean of predicted ratings, and N denotes the number of observations, i.e, number of ratings appearing in the sixth column in Table 1.

The significance level of the F test was set at $\beta = 5\%$. The null hypothesis for the WCKF - CKF RMSE performance comparison is $H_0 : w_1^2 = w_2^2$ which is rejected if $\mathcal{F}_{1,2} < F_{1-\beta/2}$ or $\mathcal{F}_{1,2} > F_{\beta/2}$, where $F_{1-\beta/2} = F(1-\beta/2, N-1, N-1)$ and $F_{\beta/2} = F(\beta/2, N-1, N-1)$ are the critical values of the F distribution with $N-1$ degrees of freedom and significance level equal to the subscript. Table 2 summarizes the critical values and the F statistic.

Table 2: F-test.

Destination	$F_{1-\beta/2}$	$F_{\beta/2}$	$\mathcal{F}_{1,2}$
Anatoliki Thraki	0.9608	1.0108	**1.0534**
Athens	0.9862	1.0140	**1.1122**
Argolida	0.9801	1.0203	**1.0534**
Crete	0.9947	1.0053	**1.0551**
Ioannina	0.9761	1.0245	**1.1683**
Kefalonia	0.9855	1.0147	**1.2176**
Magnisia	0.9739	1.0268	**1.1061**
Messinia	0.9591	1.0426	**1.2024**
Mykonos	0.9891	1.0110	**0.9410**
Naxos	0.9824	1.0179	**1.1313**
Paros	0.9826	1.0177	**1.0184**
Rhodes	0.9671	1.0340	**0.8596**
Santorini	0.9926	1.0074	**1.0083**
Skiathos	0.9606	1.0410	**0.8954**
Thessaloniki	0.9813	1.0190	**0.8692**

In Table 2, the $\mathcal{F}_{1,2}$ measure indicates that there is evidence to reject the null hypothesis of equal variances for the proposed WCKF, and CKF at the 0.05 level of significance for all destinations. As a consequence, the differences in RMSE are statistical significant.

Table 3 summarizes the RMSE for all destinations, where experiments were conducted for different values of c. It is

Table 3: RMSE of WCKF for each destination and for different values of weight c.

Destination	RMSE c=1	RMSE c=2	RMSE c=3	RMSE c=4	RMSE c=5	RMSE c=6	RMSE c=7	RMSE c=8	RMSE c=9	RMSE c=10	RMSE c=11	RMSE c=12	RMSE c=13	RMSE c=14	RMSE c=20
Anatoliki Thraki	1.341	0.7212	0.7039	0.6937	0.6881	0.6863	0.685	0.6722	0.6559	0.6479	0.6478	0.6459	0.649	0.679	0.7301
Athens	1.4376	0.9694	0.6884	0.6802	0.6741	0.666	0.6744	0.6731	0.6721	0.6712	0.6731	0.6771	0.7099	0.7281	0.8993
Argolida	1.5214	0.8714	0.8663	0.8421	0.7518	0.7509	0.7513	0.7519	0.7522	0.7529	0.7542	0.7718	0.792	0.8102	0.8419
Crete	1.4982	0.743	0.7439	0.7442	0.7446	0.7452	0.7458	0.7451	0.7449	0.7447	0.7459	0.7481	0.7509	0.7531	0.9165
Ioannina	1.5546	0.8571	0.8547	0.8564	0.8578	0.8631	0.8644	0.8601	0.8541	0.855	0.8583	0.8631	0.8685	0.8721	0.9001
Kefalonia	1.5222	0.7937	0.7912	0.7901	0.7851	0.7882	0.7889	0.7891	0.7893	0.7895	0.7937	0.7963	0.798	0.8012	0.8214
Magnisia	1.3421	0.7844	0.7867	0.7908	0.7915	0.7928	0.7931	0.7926	0.7919	0.791	0.7924	0.7931	0.7948	0.7961	0.9211
Messinia	1.428	0.7129	0.7071	0.7361	0.7368	0.7375	0.7381	0.7369	0.735	0.7341	0.7352	0.7361	0.7371	0.7499	0.9082
Mykonos	1.5562	0.8024	0.7991	0.7928	0.7958	0.7964	0.7982	0.7989	0.8032	0.8051	0.8109	0.8123	0.8179	0.8298	1.1099
Naxos	1.3392	0.7641	0.7839	0.7991	0.7999	0.8021	0.8099	0.8199	0.8195	0.8191	0.8201	0.8214	0.8219	0.8241	0.9993
Paros	1.4586	0.7947	0.8137	0.8168	0.8201	0.8214	0.8223	0.8109	0.8013	0.8004	0.8035	0.8145	0.8201	0.8239	1.0081
Rhodes	1.2987	0.8114	0.7032	0.7846	0.7881	0.7901	0.7914	0.7801	0.7611	0.6875	0.7012	0.7069	0.7123	0.7889	0.9044
Santorini	1.4018	0.8097	0.8161	0.8188	0.819	0.8195	0.8208	0.8213	0.8209	0.8204	0.8221	0.8229	0.8235	0.8247	0.9002
Skiathos	1.1793	0.6814	0.6919	0.7416	0.7442	0.7449	0.7492	0.7521	0.7522	0.7518	0.752	0.7529	0.7541	0.7566	0.8002
Thessaloniki	1.2241	0.6246	0.6232	0.6503	0.6541	0.6599	0.6623	0.6682	0.6721	0.6714	0.6708	0.6892	0.6901	0.6945	0.8149

seen that a much smaller RMSE is obtained when c is set to a smaller value than 10 for destinations associated to large volumes of ratings, such as Santorini or Crete. More elaborate validation techniques, such as those in [14], could be helpful in choosing the value of c.

5. CONCLUSION

A weighted ordered probit CKF, namely WCKF has been proposed, for tripadvisor hotel rating prediction. The weighted ordered probit model has enhanced the prediction performance of CKF. This model has been demonstrated that captures the evolution of each hotel popularity through time. The dynamically changing magnitude of the regions of the weighted ordered probit model with respect to each hotel trend ensured the efficacy of rating predictions. WCKF takes into consideration possible changes in hotel services or accommodation conditions and readjusts the rating predictions with respect to this temporal information.

6. REFERENCES

[1] Y. Koren and R. Bell, "Advances in collaborative filtering," in *Recommender Systems Handbook*. Springer, 2015, pp. 77–118.

[2] D. Agarwal and B.-C. Chen, "fLDA: matrix factorization through latent dirichlet allocation," in *Proc. 3rd ACM Int. Conf. Web Search and Data Mining*. ACM, 2010, pp. 91–100.

[3] H. Shan and A. Banerjee, "Generalized probabilistic matrix factorizations for collaborative filtering," in *Proc. 2010 IEEE Int. Conf. Data Mining*. IEEE, 2010, pp. 1025–1030.

[4] I. Fernández-Tobías, I. Cantador, P. Tomeo, V. W. Anelli, and T. Di Noia, "Addressing the user cold start with cross-domain collaborative filtering: exploiting item metadata in matrix factorization," *User Modeling and User-Adapted Interaction*, pp. 1–44, 2019.

[5] A. Mnih and R. R. Salakhutdinov, "Probabilistic matrix factorization," in *Advances in Neural Information Processing Systems*, 2008, pp. 1257–1264.

[6] R. R. Rastogi, V. Agnihotri, R. Bhatt, and S. Merugu, "Probabilistic matrix factorization system based on personas," Oct. 2 2018, US Patent App. 10/089,675.

[7] B. Pal and M. Jenamani, "Kernelized probabilistic matrix factorization for collaborative filtering: Exploiting projected user and item graph," in *Proc. 12th ACM Conf. Recommender Systems*. ACM, 2018, pp. 437–440.

[8] K. Li, X. Zhou, F. Lin, W. Zeng, and G. Alterovitz, "Deep probabilistic matrix factorization framework for online collaborative filtering," *IEEE Access*, 2019.

[9] H. Wang, N. Wang, and D.-Y. Yeung, "Collaborative deep learning for recommender systems," in *Proc.21th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*. ACM, 2015, pp. 1235–1244.

[10] S. Gultekin and J. Paisley, "A collaborative Kalman filter for time-evolving dyadic processes," in *Proc. 2014 IEEE Int. Conf. Data Mining*, 2014, pp. 30–37.

[11] V. K. Borooah, *Logit and Probit: Ordered and Multinomial Models*. Sage, 2002, no. 138.

[12] M. Ntemi and C. Kotropoulos, "Prediction methods for time evolving dyadic processes," in *Proc. 26th European Signal Processing Conf.*, 2018, pp. 2588–2592.

[13] D. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *Journal American Statistical Association*, vol. 112, no. 518, pp. 859–877, 2017.

[14] J. Ding, V. Tarokh, and Y. Yang, "Model selection techniques: An overview," *IEEE Signal Processing Magazine*, vol. 35, no. 6, pp. 16–34, November 2018.