Target Tracking in the Recommender Space

Toward a new recommender system based on Kalman filtering

Samuel Nowakowski, Anne Boyer
LORIA-KIWI
Campus scientifique – BP239
F-54506 Vandœuvre-lès-Nancy Cedex
Samuel.nowakowski@loria.fr
Anne.boyer@loria.fr

Cédric Bernier
Alcatel-Lucent Bell Labs France – LORIA-KIWI
Route de Villarceaux
F-91600 Nozay
Cedric.bernier@loria.fr

Abstract— **We assume that users and their consumptions of television programs are vectors in the multidimensional space of the categories of the resources. Knowing this space, we propose an algorithm based on a Kalman filter to track the user's profile and to foresee the best prediction of their future position in the recommendation space. The approach is tested on data coming from TV consumptions.**

Keywords-recommender system; user profile; group profile; Kalman filter; target tracking

I. INTRODUCTION

In Web-based services of dynamic content, recommender systems face the difficulty of identifying new pertinent items and providing pertinent and personalized recommendations for users.

Personalized recommendation has become a mandatory feature of Web sites to improve customer satisfaction and customer retention. Recommendation involves a process of gathering information about site visitors, managing the content assets, analyzing current and past user interactive behavior, and, based on the analysis, delivering the right content to each visitor.

Recommendation methods can be distinguished into two main approaches: content based filtering [9] and collaborative filtering [10]. Collaborative filtering (CF) is one of the most successful and widely used technology to design recommender systems. CF analyzes user ratings to recognize similarities between users on the basis of their past ratings, and then generates new recommendations based on like-minded users' preferences. This approach suffers from several drawbacks, such as cold start, latency, sparsity [11], even if it gives interesting results.

The main idea of this paper is to propose an alternative way for recommender systems. Our work is based on the following assumption: we consider Users and Web resources as a dynamic system described in a state space. This dynamic system can be modeled by techniques coming from control system methods. The obtained state space is defined by state variables that are related to the users. We consider that the states of the users (by states, we understand « what are the resources they want to see in the next step ») are measured by the grades given to one resource by the users.

In this paper, we are going to present the effectiveness of Kalman filtering based approach for recommendation. We will detail the backgrounds of this approach, i.e., state space description and Kalman filter. Then, we expose the applied methodology. Our conclusion will give some guidelines for future works.

II. PRINCIPLES

Kalman filter is an optimal state estimator of a linear system. It can estimate the state of the system using a priori knowledge of the evolution of the state and the measurements.

A. Target tracking in the cyberspace

Hypothesis: the user is a target which is moving along an a priori unknown trajectory in the multidimensional space of the categories. Figure 1 shows the principle of our approach.

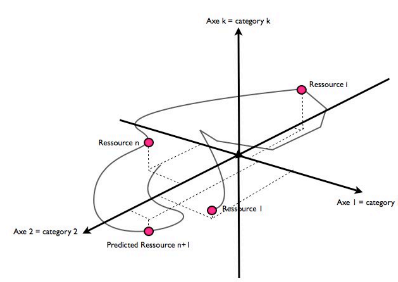

Figure 1. Trajectory in the recommender space

B. Kalman filter: equations

How can we know about a target moving in the recommender space?

Using its position, speed and acceleration, we choose as the state vector the following form:

$$X_k = \begin{bmatrix} x \\ \dot{x} \\ \ddot{x} \end{bmatrix}_k \tag{1}$$

where :
x is the vector containing the position vector
$\dot{x}$ is the vector containing the speed vector,
$\ddot{x}$ is the vector containing the acceleration vector.

The estimation of this state vector will give the necessary knowledge of the trajectory in the recommender space. We use the following state space model:

$$\begin{cases} X_{k+1} = AX_k + w_k \\ Z_k = HX_k + v_k \end{cases} \quad (2)$$

where matrix A includes the relationship between the position, speed and acceleration where T represents the time period. In our case, we consider T = 1. w_k and v_k are random noises which takes into account unexpected variations in the trajectories.

$$A = \begin{bmatrix} \alpha & T & \frac{1}{2}T^2 \\ 0 & \alpha & T \\ 0 & 0 & \alpha \end{bmatrix} \quad (3)$$

Matrix H will have the following structure, as shown in the Figure 2.

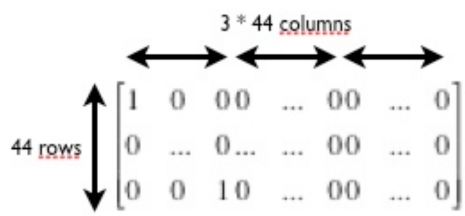

Figure 2. Structure of Matrix H

The Kalman filter equations are then given [6]:

Prediction at time k knowing k+1 ($\hat{X}_{k/k-1}$)

$$\begin{cases} \hat{X}_{k+1/k} = \hat{X}_{k/k-1} + K_k\left(Z_k - H\hat{X}_{k/k-1}\right) \\ \quad\quad = \left(A - K_k C\right)\hat{X}_{k/k-1} + K_k Z_k \end{cases} \quad (4)$$

Kalman gain:

$$K_k = AP_{k/k-1}H^T\left(HP_{k/k-1}H^T + R\right)^{-1} \quad (5)$$

The evolution of the uncertainty on the estimation is then given by:

$$P_{k+1/k} = AP_{k/k-\&}A^T - AP_{k/k-1}H^T\left(HP_{k/k-1}H^T + R\right)^{-1}HP_{k/k-1}A^T \quad (6)$$

where the initial conditions are given by:

$$\hat{X}_{0/-1} = X_0 \,, P_{0/-1} = P_0$$

and the state prediction by:

$$\hat{X}_{k+1/k}$$

III. Application

A. Description of the experiment

This experiment is based on TV consumption. The dataset is the TV consumption of 6423 English households over a period of 6 months (from 1st September 2008 to 1st March 2009) (Broadcaster Audience Research Board, [7]), [8]. This dataset contains information about the user, the household and about television program. Each TV program is labelled by one or several categories. In the experiment, a user profile build for each person. The user profile is the set of categories associated to the value of interest of the user. This user profile is elaborated in function of the quality of a user's TV consumption: if a TV program is watched entirely, the genre associated to this TV program increases in the user profile. Several logical rules are applied to estimate the interest of a user for a TV program.

The methodology of the experimentation is the following:

- Each user profile is computed at different instants (35) from the TV viewing data.
- The Kalman filter is applied iteratively to estimate the following positions of the user profile in the recommender space.

The entire consumption is described by 44 types which will define the 44 dimensions space where users are "moving".

B. Numerical results

The obtained results can be exposed as follows:

- Kalman filter predicts the interest of a specific user for one gender knowing his past.

Using this prediction, we can propose a new recommendation strategy:

- If the Quantity of Interest (QoI) of the user is predicted to be in one specific region of the space, we can recommend something inside this specific region:
- For example, if the specific region is defined by dimensions Documentary and Drama, we can recommend contents related to these two dimensions
- If the predicted quantity of interest (QoI) changes to another dimension of the space, we can automatically recommend content from this new region of the space.

C. Results

The results can be analyzed as follows: Kalman filter predicts the specific interest for a category of contents of one user.

Figures 3 and 4 show Estimation / Prediction computed by Kalman filtering. Doted-lines show the evolution of the

 ISBN: 978-1-61208-141-0

real values. Continuous lines show the obtained predictions. We can see that the prediction curve given by the filter fits very well the real data.

Figure 5 shows the results of the cosine distance which has been computed between the true values and the prediction by the filter. It shows that the prediction will quickly converge with the true values.

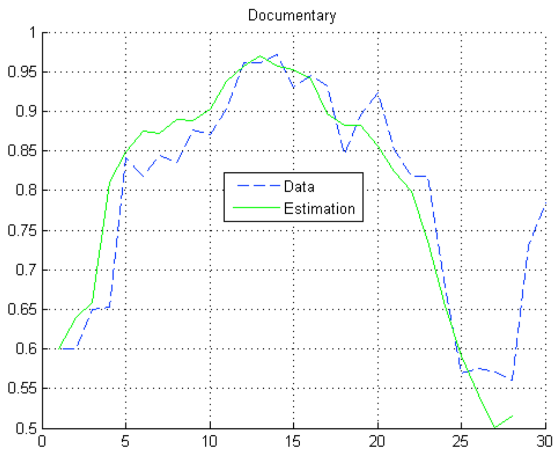

Figure 3. Prediction for Drama

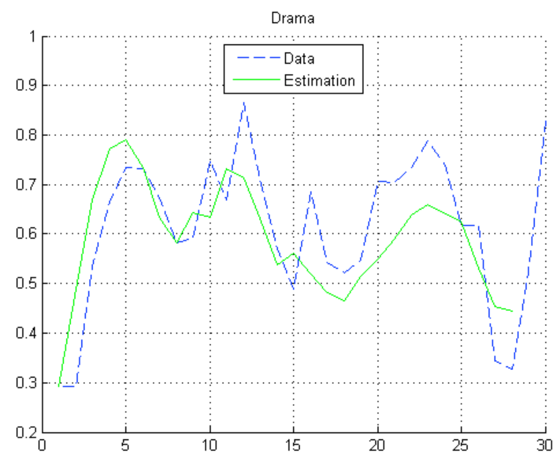

Figure 4. Prediction for Documentary

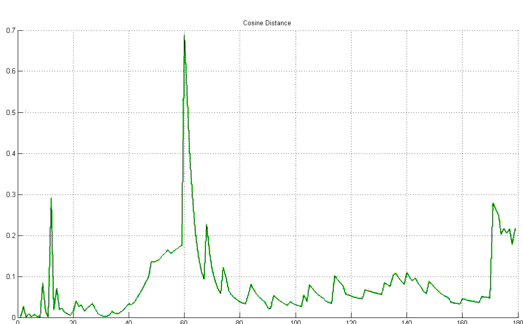

Figure 5. Cosine distance

IV. RECOMMENDATION STRATEGY

In this approach, we can build a recommendation by analyzing the estimation provided by Kalman filter.

The profile is built from the consumption of TV programs. Each TV program is defined by categories such as entertainment, science fiction, talk show, etc. The analysis of the way different TV programs are watched allows us the possibility to estimate the interest of a user for each category. Hence, the user profile is calculated from the TV consumption and it is represented by a vector of valuated concepts.

The user profile is considered as a point in the 44 dimensions-space. This point moves at each different time in the space along a trajectory. With the Kalman filter, we predict the future position of the user profile. The prediction shows the evolution of the trajectory in subspaces restricted to specific dimensions.

For our new recommending strategy (see Figure 6), we observe the difference between the predicted category and the computed one. The rule can be derived as follows:

- If the computed QoI for one category is superior to the estimated QoI (noted negative difference in figure 6), then the user's interest for this category is decreasing.
- If the predicted QoI is superior to the computed one (noted positive difference in Figure 6), then the user's interest for this category is increasing.

Our strategy will could be :

- QoI for specific categories with an important positive difference will influence the recommendation towards these categories
- QoI with an important negative difference discourage the recommendation towards these categories.

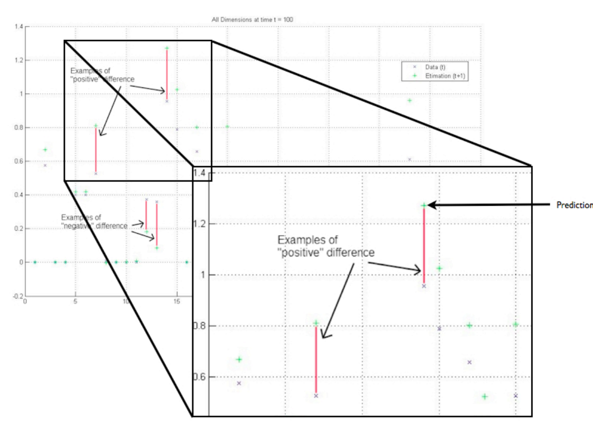

Figure 6. Analysis of the evolution of the prediction for recommendation

Conversely to existing methods which recommend specific contents for a given user, this method takes into account the user's state of mind and will recommend a set of categories of movies inside a subspace of the whole recommender space. Our method performs on the macroscopic level. We find out the type of content the user appreciates and can determine some dimensions that can deliberately be closed out.

The recommendation is based on the two preceding arguments.

- the user's actual state of mind
- the subset of identified dimensions.

From the analysis of these "positive" or "negative" dimensions and from the TV program, we will define the recommendation strategy for a set of TV programs. According to what the user watched during the day, we can refine our recommendation:

- if the user is interested in contents of types x, y and z and if he has already watched content of type x and y, the recommendation would essentially concentrate on content of type z.

The recommender strategy will recommend contents belonging to the categories corresponding to the selected dimensions of the recommender space.

V. CONCLUSION

In this paper, the main idea is to consider that the one who chooses films as a target which moves along a trajectory in the recommender space. The recommender space is seen as a 44 dimensions space based on the main categories describing the movies. The position of the target is measured by the Quantity of Interest (QoI) for certain categories of movies. Then, the Kalman filter applied using a tracking state space model predicts the "positions" in the recommender space. Knowing the past positions of the user in this space along the different axis of the 44 dimensions space, our Kalman filter based recommender system will suggest:

- if the user is interested in contents of types x, y and z and if he has already watched content of type x and y that day, the recommendation would essentially concentrate on content of type z
- knowing the position in the space, the best prediction for his future positions in the recommender space, i.e., his best index of interest related to the favorite contents.

The strength of our approach is in its capability to make recommendations at a higher level which fit users habits, i.e., given main directions to follow knowing the trajectory in the space and not to suggest specific resources.

Future works will be focused on tracking groups of users and on the definition of the topology of the recommendation space as a space including specific mathematical operators.

REFERENCES

[1] Anderson, B., and Moore, J. B., 1977. Optimal filtering. Prentice Hall – Information and System Sciences Series

[2] Gibson, W, 1988. Neuromancien. Collection J'ai Lu, La découverte, ISBN 2-7071-1562-2

[3] Söderström, T., 1994. Discrete-time stochastic systems : estimation and control. Prentice Hall International.

[4] Box, G.E.P., and Jenkins, G.M., 1970. Time series analysis : forecasting and control. Holden Day.

[5] Bernier C., Brun A., Aghasaryan A., Bouzid M., Picault J., Senot C., and Boyer A., 2010. Topology of communities for the collaborative recommendations to groups, SIIE 2010 conference, (Sousse, Tunisia, February 17 – 19, 2010).

[6] Gevers, M., and Vandendorpe, L. Processus stochastiques, estimation et prediction. DOI=http://www.tele.ucl.ac.be/EDU/INMA2731/].

[7] BARB: Broadcaster Audience Research Board, DOI=http://www.barb.co.uk/.

[8] Senot, C., Kostadinov D., Bouzid M., Picault J., Aghasaryan A., and Bernier C., 2010. Analysis of strategies for building group profiles. User Modeling, Adaptation and Personalization.

[9] M Pazzani, D. Billsus, 2007. Content-Based Recommendation Systems. In Brusilovsky, P. Kobsa, A. Nejdl, W. (réds) The Adaptive Web : Methods and Strategies of Web Personalization, pp. 325–341.

[10] D. Goldberg, D. Nichols, B. Oki, and D. Terry, 1992. Using Collaborative Filtering to Weave an Information Tapestry. Communications of the ACM, 35(12), pp. 61–70

[11] M. Grcar, D. Mladenic, B. Fortuna, and M. Groblenik, 2006. Data Sparsity Issues in the Collaborative Filtering Framework. Advances in Web Mining and Web Usage Analysis, pp. 58–76

[12] Smith, J., 1998. The book, The publishing company. London, 2nd edition