Proceedings of the Workshop Net-Works 2008,
Pamplona, 9–11 June, 2008, pp. 1–4.

Information spread in Recommendation Systems

Massimiliano Zanin[1,†], Pedro Cano[1], Javier M. Buldú[2] and Oscar Celma[1]

[1] *Music Technology Group, Universitat Pompeu Fabra*

[2] *Departamento de Física, Universidad Rey Juan Carlos*

Abstract

In this work we introduce a novel algorithm for spreading information across a complex network. Depending on the initial conditions and on the structure of the graph, this algorithm can be applied to a great variety of problems, all of them included in the wide family of *Data Mining* applications. Specifically, when the network is constructed around a buyers dataset and the information represents users tastes, this strategy leads to an improvement of standard Recommendation Systems.

Keywords: Complex networks, information spread, recommendation algorithms, data mining.
MSC 2000: 90C35

1. Introduction

In recent years, recommendation algorithms have been under the focus of the scientific community, because of the challenges they set (e.g., the large amount of information to be parsed and the resulting computational load) and the interest of the market, increased by the growing number of e-stores. Within this framework, recommender systems benefit users by making easier to find items they like and, conversely, this helps the business by generating more sales. This interest can be represented by the acquisition of MyStrands, a young company specialized in recommendation systems, by venture capitals for 31 million Dollars [1].

Many studies have been developed in this field, using a wide variety of theories: from matrix manipulations to sociological approaches (a review of the evolution of recommendation techniques can be found on [2]). Another approach has been proposed by the authors in the past [3], and it focus in the

[†]Corresponding author: massimiliano.zanin@hotmail.com

improvements resulting from applying complex networks theories [4, 5] to the recommendation problem.

In this work, improvements in the recommendation are obtained through the construction of an underlying bipartite complex network of users and items, and by selecting and diffusing some characteristic information about those items. Moreover, this approach allows to construct a family of algorithms, included in the wide topic of *data mining* [6], that can be applied to different problems by changing the information and the diffusion mechanism.

2. Main application: recommending to users

The basic problem covered by this algorithm is a recommendation process: with the information of the past acquisitions of a user, the system should be able to suggest a new item that fulfill with the taste of that user. The data obtained from the users background can be projected into a bipartite network, with users belonging to an upper layer and items in a lower one. Connections between layers are created when a user buys a certain item (see Fig. 1 for a schematic representation of the bipartite network).

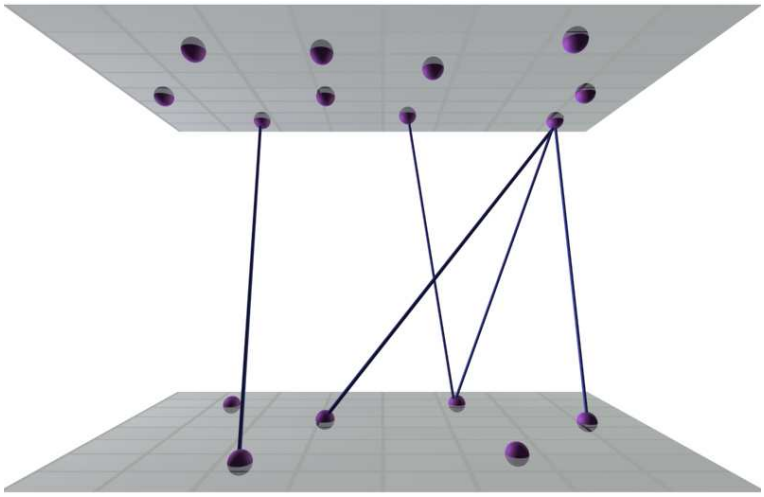

Figure 1: Creation of a bipartite network from users' acquisitions. Nodes in the upper layer represents users while nodes in the bottom layer are items. A connection is created when a user acquires a certain item.

After a first diffusion process, the system outputs a *compatibility value* for each item, in the range $[0, 1]$, which indicates how close is that item to the target user. Items that have been recommended to a given user are ranked

according to their compatibility value, i.e., items in the upper part of this rank should be closer to the target user, at least from the point of view of its previous taste, and therefore they are the first to be recommended. The actual score of the algorithm is calculated by comparing the first n items recommended with the real purchases of the user, and then this score is compared with the one obtained with a standard recommendation system (e.g. item-based collaborative filtering).

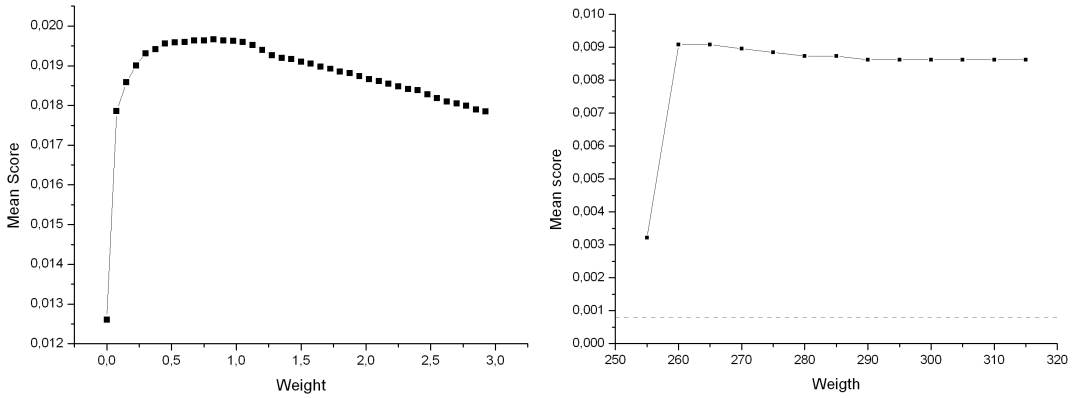

Figure 2: Results obtained with two different datasets, namely the Netflix movies dataset (left) and the Art of the Mix playlists dataset (right).

In Fig. 2 results are shown for two different datasets. The first dataset includes ratings about movies obtained from the Netflix dataset [2]. The basic user-based recommendation algorithm, i.e. when the weight of the reranking algorithm is zero, has a mean score of 0.0127; on the other side, when the reranking is included, we obtain a score close to 0.02, that is an improvement of about 55%. The *Art of the Mix* dataset is slightly different, as it is created by users representing their own musical tastes in personal playlists of songs. Due to the low mean degree of the items (around 2.81 links per song), results obtained with a standard user-based algorithm are disappointing (red horizontal line in Fig. 2 right); when adding the reranking, we can improve the mean score up to 10 times.

3. Conclusions

From the above examples, it is clear that an interesting application of complex network theories can be found in recommendation systems; in this case, we use a bipartite network to construct the layout of an information spreading algorithm, and next we use this information for constructing a *reranking* strategy.

Numerical results obtained over two different datasets show how this approach is useful even in extreme situations, e.g. when the network is too sparse for a standard algorithm.

Other applications include score forecasting (i.e., detecting how a user will score an item, for example a movie) or tags mismatching detections.

4. Acknowledgments

The authors are grateful with all the staff of the Music Technology Group for their support and interesting discussions. This research has been partially funded by the EU Project Pharos IST-2006-045035.

References

[1] `http://blog.strands.com/`207/06/18/0

[2] J. Bennett S. Lanning "Netflix Prize", *Proc. KDD Cup and Workshop 2007*, San Jose, California (2007).

[3] M. Zanin P. Cano J. M. Buldú "Preferential attachment, aging and weights in recommendation systems", *Proc. Net-Works 2007*, Aranjuez (2007).

[4] S. Boccaletti V. Latora Y. Moreno M. Chavez D.-U. Hwang "Complex Networks: Structure and Dynamics", *Physics Reports* **424**, (2006) 175.

[5] M. E. J. Newman "The structure and function of complex networks", *SIAM Review* **45**, (2003) 67–256.

[6] M. H. Dunham, "Data mining", *Prentice Hall*, (2003).