Recommending Friends by Identifying Latent Similarities in Social Environments

Roberto Saia, Luca Piras, and Salvatore Carta

Dipartimento di Matematica e Informatica
Università di Cagliari
{roberto.saia, lucapiras, salvatore}@unica.it

Abstract. When browsing social media, we get in touch with great amounts of content that, if well exploited, can provide valuable knowledge on our preferences. On the one hand, the opportunities that the interaction of the users with social content can offer are great, since they represent implicit feedbacks that the users provide on what they like. On the other hand, the resulting information is very sparse, since the users do not interact in any form with a lot of content (e.g., by liking, commenting, or clicking on an item). Therefore, finding similarities between the users to recommend friends with similar preferences is a challenging but important task. This paper introduces a novel technique able to discover the shared latent-spaces between users by moving the data analysis in the frequency domain, where the spectral patterns of the users are compared. By identifying non-explicit similarities between the users, friend recommendations can be performed.

1 Introduction

With the advent of social media, users have moved from being content consumers to content producers. This led to an overwhelming amount of content available in the existing platforms. When browsing this content, we skip most of it, because it is not of particular interest for us. On the contrary, when something attracts our attention, we interact with it, by means of clicks, likes, and comments. These interactions represent valuable forms of feedbacks that the users provide on what they like (or do not like).

Nowadays, a classic application scenario in which we continuously browse content are social media systems, like Facebook or Twitter. Hence, the feedback left by the users might be exploited to create a user profile with their preferences and provide them recommendation of possibly interesting content or of users with similar preferences.

The problem that arises from collecting implicit feedbacks is the sparsity of the resulting data. Indeed, the content for which we do not provide any feedback is much more than that we interact with. Hence, employing classic user profiles that collect the preferences of the users might not be good to provide recommendations to the users. The idea behind this paper is to move the process of definition of the user profiles from the canonical domain to the frequency one, by performing the spectral analysis through the *Fourier transformation* [3].

As canonical domain we mean the sequence of a user feedbacks in terms of items selection (we consider as an item any piece of content a user can interact with in a social media platform), which we consider as the input *time series* of the *Fast Fourier Transform* (FFT) process.

The new data representation (spectral pattern) offers us an opportunity to discover latent similarities between users, thus allowing us to perform recommendations under a new point of view. Indeed, these latent similarities can be then used to identify users to recommend.

The main contributions given by this research are listed below:

1. definition of the *time series* to be used in the *Fast Fourier Transform* (FFT), made by processing the sequence of choices of each user;
2. definition of the comparison process between users in terms of their *spectral patterns*;
3. definition of the *Spectral Pattern Profiling* (SPP) approach able to generate recommendation for a user on the basis of the *spectral similarity*.

The remainder of the paper is organized as follows: Section 2 discusses the background and related work; Section 3 provides a formal notation and makes some premises; Section 4 describes the implementation of the proposed approach; some concluding remarks and future work are given in the last Section 5.

2 Related work

This section presents related work on user recommendation in social media systems and background on the Fourier Transformation.

2.1 User recommendation in social media systems

In [4], Gupta et al. present Twitter's user recommendation service, which is based on shared interests, common connections, and other related factors and provide some background on the Fourier transformation. The proposed system builds a graph, in which the vertices represent users and the directed edges represent the "follow" relationship; this graph is processed with an open source in-memory graph processing engine called Cassovary. Finally, recommendations are built by means of a user recommendation algorithm for directed graphs, based on SALSA (Stochastic Approach for Link-Structure Analysis). In the next section, we are going to analyze this system, in order to design our proposal.

In [2], Chen et al. describe a people recommender system in an enterprise social network domain. They compare four algorithms, two based on social relationship information and two based on content similarity, and demonstrate that the algorithms that use social information are more capable to find known contacts, while algorithms based on content similarities are better to discover new friends. This approach produces the recommendations by analyzing both the interaction with the content and the interaction with the other users and this characteristic will be analyzed while designing our system.

Guy et al. [5] describe a people recommender system for the IBM Fringe social network. The system uses enterprise information, like org chart relationships, paper and patent co-authorship and project co-membership, which are specific of this social network, so it is hard to take into account this approach when designing our system.

Hannon et al. [6] describe a followee recommender system for Twitter, based on tweets and relationships of their Twitter social graphs. By using this information, they build user profiles and demonstrate how these profiles can be used to produce recommendations. In our proposal, we aim at recommending friendship relationships and not users to follow.

In [13], a recommender system based on collocation (i.e., the position of the user) is presented. It uses short-range technologies of mobile phones, in order to infer the collocation and other correlated information that are the base for the recommendations. In our domain we do not have such a type of information, so we cannot compare with this algorithm.

Zhou et al. [14] propose a framework for users' interest modeling and interest-based user recommendation (it suggests people to follow and not friends), which was tested on the Yahoo! Delicious dataset. Recommendations are produced by analyzing the network and fans properties. Differently from this framework, our proposal aims at producing friend recommendations.

In [1], a study about what cues in a user's profile, behavior, and network might be most effective in recommending people, is presented. So, this approach analyzes both the interaction with the content and the interaction with the other users and this characteristic will be analyzed while designing our system.

Manca et al. [8–12] proposed a friend recommender system that operates in the social bookmarking application domain and is based on *behavioral-data mining*, i.e., on the exploitation of the users activity in a social bookmarking system. The authors show that, thanks to their approach, the impact of the "interaction overload" and the "over-specialization" problems is strongly reduced.

Liben-Nowell and Kleinberg [7] studied the user recommendation problem as a link prediction problem. They develop several approaches, based on metrics that analyze the proximity of nodes in a social network, to infer the probability of new connections among users. Experiments show that the network topology is a good tool to predict future interactions.

2.2 Fourier Transform

The basic idea behind the approach proposed in this paper is to move the process of evaluation of the new instances (*time series*) from their canonical time domain to the frequency one, in order to obtain a representative pattern composed by their frequency components, as shown in Figure 1.

This operation is performed by recurring to the *Discrete Fourier Transform* (DFT), whose formalization is shown in Equation 1, where i is the imaginary unit.

$$F_n \stackrel{\text{def}}{=} \sum_{k=0}^{N-1} f_k \cdot e^{-2\pi ink/N}, \quad n \in \mathcal{Z} \tag{1}$$

The result of the Equation 1 is a set of sinusoidal functions, each corresponding to a particular frequency component (i.e., the *spectrum*[1]).

If it is necessary, we can use the *inverse Fourier transform* shown in Equation 2 to return to the original time domain.

$$f_k = \frac{1}{N} \sum_{n=0}^{N-1} F_n \cdot e^{2\pi ikn/N}, \quad n \in \mathcal{Z} \tag{2}$$

The *Fast Fourier Transform* (FFT) algorithm, used in the context of this paper to perform the Fourier transformations, rapidly computes the DFT by factorizing the input matrix into a product of sparse (mostly zero) factors. It is largely used because it reduces the computational complexity of the process from $O(n^2)$ to $O(n \log n)$, where n denotes the data size.

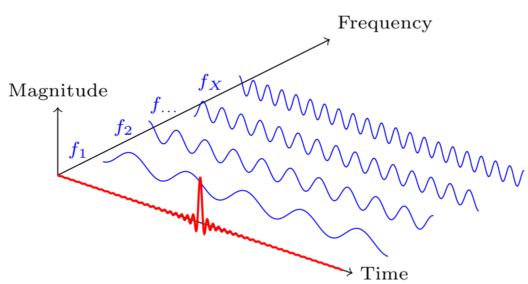

Fig. 1. *Time and Frequency Domains*

3 Preliminaries

Formal notation and premises to this paper are stated in the following.

[1] The *spectrum* of the *frequency components* is the frequency domain representation of a signal.

3.1 Notation

Given a set of users $U = \{u_1, \dots, u_N\}$, a set of items $I = \{i_1, \dots, i_M\}$, we denote as r_{ui} the relevance of an item $i \in I$ for a user $u \in U$, captured from the implicit feedbacks.

We also denote as $I_u \subseteq I$ the set of items $i_1, i_2, \dots, i_K$ in the profile of a user $u \in U$.

Finally, we denote as $F = \{f_1, f_2, \dots, f_X\}$ the frequency components (spectrum) obtained as result of the DFT process.

3.2 Premises

Assuming that the time interval between the collection of the feedbacks is equal, the considered non-periodic wave is given by the sequence of choices $\hat{i}_1, \hat{i}_2, \dots, \hat{i}_K$ with $\hat{i} \in I$, which compose each user profile I_u, which represents the *time series* taken into account.

Their fundamental period T starts with $\hat{i}_1$ and it ends with $\hat{i}_K$, thus we have that $sr = |I_u|$; the sample interval si is instead given by the fundamental period T divided by the number of acquisition, i.e., $si = \frac{T}{|I_u|}$.

Through the FFT algorithm we compute the *Discrete Fourier Transform* of each *time series*, by converting their representation from the time domain to the frequency one. The obtained frequency-domain representation provides information about the signal's magnitude and phase at each frequency. For this reason, the output (denoted as x) of the FFT computation is a series of complex numbers composed by a real part x_r and an imaginary part x_i, thus $x = (x_r + ix_i)$.

We can obtain the x magnitude by using $|x| = \sqrt{(x_r^2 + x_i^2)}$ and the x phase by using $\varphi(x) = \arctan\left(\frac{x_i}{x_r}\right)$, although in the context of this paper we will take into account only the magnitude at each frequency.

4 Proposed Approach

The implementation of our approach is carried out through the following steps:

1. **Time Series Definition**: definition of the *time series* to use in the FFT algorithm, in terms of ordered sequences of items chosen by a user;
2. **Spectral Pattern Comparison**: evaluation of the similarity between users in terms of comparison of their spectral patterns, obtained by processing their *time series* through the FFT algorithm;
3. **Spectral Pattern Recommendation**: formalization of the *Spectral Pattern Profiling* (SPP) algorithm able to generate user recommendation on the basis of the *Spectral Pattern Comparison* process.

In the following, we provide a detailed description of each of these steps.

4.1 Time Series Definition

The first step of our approach is aimed to define the *time series* to use in the *Discrete Fourier Transform* process. Considering that a *time series* is a series of data points stored by following the time order and usually it is a sequence captured at successive equally spaced points in time, thus it can be considered a sequence of discrete-time data.

The *time series* (ts) that we will take into account are defined by using the sequence of items chosen by a user (i.e., the ID of the evaluated items in the user profile I_u), numerically sorted in ascending order after they were divided by the relevance of the item for that user (in order to prioritize those with a higher relevance), as shown in the example of Equation 3. Since we are dealing with implicit feedbacks, the relevance has be inferred by monitoring the activity of the users (e.g., by measuring the time spent while interacting with an item, like the reading or viewing time, or the number of interactions the users had with a piece of content, like the number of comments). In this work we are assuming the relevance scores to be normalized in a fixed range (e.g., [1,5]).

$$\mathrm{ts} = [i_2, i_1, \ldots, i_K] \quad \text{with} \quad \tfrac{i_2}{r(i_2)} \leq \tfrac{i_1}{r(i_1)} \leq \ldots \leq \tfrac{i_K}{r(i_K)} \tag{3}$$

4.2 Spectral Pattern Comparison

In this step, we move the *time series* to the frequency domain by a DFT process performed through the FFT approach introduced in Section 2.2. Basically, we extract the spectral pattern of each user profile by processing the related *time series* defined in the previous step.

The process is aimed at comparing the spectral pattern of a user to the spectral patterns of the other users in the dataset, in order to rank these last ones in terms of spectral similarity. The spectral similarity (σ) between two users u_1 and u_2 (where u_1 is the user to recommend and u_2 is one of the other users in the dataset, with $u_1, u_2 \in U$) is calculated by taking into account the magnitude of all frequency components, as follows:

$$\sigma(u_1, u_2) = \tfrac{1}{|u_1|} \sum_{x=1}^{X} \Theta, \;\; with \;\; \Theta = \begin{cases} 1, & f_x(u_2) \leq f_x(u_1) \\ 0, & otherwise \end{cases} \tag{4}$$

It should be noted that in case of different size of the spectral vectors, the smallest of them is stretched by adding zeros, so that the dimensions of the spectral patterns to compare have the same size (same number of frequency components).

4.3 Spectral Pattern Recommendation

The detected similarities between the user who will receive the recommendation and the other users, are ranked considering the criterion shown in Equation 4.

The produced ranking can then be used to perform friend recommendations.

5 Conclusions and Future Work

This paper proposed a new approach of recommendation based on a novel *latent factor model*.

The proposed *Spectral Pattern Profiling* approach gets such information by comparing the spectral profiles of the users, processing the *time series*, defined on the basis of the item evaluations, through a *Discrete Fourier Transform* process.

Given the similarly to the approaches based on the well-known *Singular Value Decomposition* technique (e.g., SVD and $SVD++$), aimed to discover latent similarities between users, we performed some preliminary study, which are showing that our approach is capable of capturing more effective similarities between the users for friend recommendation purposes.

As future work we will study further characteristics of the spectral profiles, with the objective to exploit them in order to improve the performance of our recommendation algorithm.

Acknowledgments

This research is partially funded by *Regione Sardegna* under project *Next generation Open Mobile Apps Development* (*NOMAD*), through PIA (Pacchetti Integrati di Agevolazione) Industria Artigianato e Servizi (annualit 2013).

References

1. Brzozowski, M.J., Romero, D.M.: Who should i follow? recommending people in directed social networks. In: Adamic, L.A., Baeza-Yates, R.A., Counts, S. (eds.) Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July 17-21, 2011. The AAAI Press (2011)
2. Chen, J., Geyer, W., Dugan, C., Muller, M.J., Guy, I.: Make new friends, but keep the old: recommending people on social networking sites. In: Jr., D.R.O., Arthur, R.B., Hinckley, K., Morris, M.R., Hudson, S.E., Greenberg, S. (eds.) Proceedings of the 27th International Conference on Human Factors in Computing Systems, CHI 2009, Boston, MA, USA, April 4-9, 2009. pp. 201–210. ACM (2009)
3. Duhamel, P., Vetterli, M.: Fast fourier transforms: a tutorial review and a state of the art. Signal processing 19(4), 259–299 (1990)
4. Gupta, P., Goel, A., Lin, J., Sharma, A., Wang, D., Zadeh, R.: Wtf: the who to follow service at twitter. In: Schwabe, D., Almeida, V.A.F., Glaser, H., Baeza-Yates, R.A., Moon, S.B. (eds.) 22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13-17, 2013. pp. 505–514. International World Wide Web Conferences Steering Committee / ACM (2013)
5. Guy, I., Ronen, I., Wilcox, E.: Do you know?: recommending people to invite into your social network. In: Conati, C., Bauer, M., Oliver, N., Weld, D.S. (eds.) Proceedings of the 2009 International Conference on Intelligent User Interfaces, February 8-11, 2009, Sanibel Island, Florida, USA. pp. 77–86. ACM (2009)
6. Hannon, J., Bennett, M., Smyth, B.: Recommending twitter users to follow using content and collaborative filtering approaches. In: Amatriain, X., Torrens, M.,

Resnick, P., Zanker, M. (eds.) Proceedings of the 2010 ACM Conference on Recommender Systems, RecSys 2010, Barcelona, Spain, September 26-30, 2010. pp. 199–206. ACM (2010)
7. Liben-Nowell, D., Kleinberg, J.M.: The link prediction problem for social networks. In: Proceedings of the 2003 ACM CIKM International Conference on Information and Knowledge Management, New Orleans, Louisiana, USA, November 2-8, 2003. pp. 556–559. ACM (2003)
8. Manca, M., Boratto, L., Carta, S.: Design and architecture of a friend recommender system in the social bookmarking domain. In: Science and Information Conference (SAI), 2014. pp. 838–842 (Aug 2014)
9. Manca, M., Boratto, L., Carta, S.: Mining user behavior in a social bookmarking system - A delicious friend recommender system. In: Helfert, M., Holzinger, A., Belo, O., Francalanci, C. (eds.) DATA 2014 - Proceedings of 3rd International Conference on Data Management Technologies and Applications, Vienna, Austria, 29-31 August, 2014. pp. 331–338. SciTePress (2014)
10. Manca, M., Boratto, L., Carta, S.: Behavioral data mining to produce novel and serendipitous friend recommendations in a social bookmarking system. Information Systems Frontiers pp. 1–15 (2015)
11. Manca, M., Boratto, L., Carta, S.: Friend recommendation in a social bookmarking system: Design and architecture guidelines. In: Arai, K., Kapoor, S., Bhatia, R. (eds.) Intelligent Systems in Science and Information 2014, Studies in Computational Intelligence, vol. 591, pp. 227–242. Springer International Publishing (2015)
12. Manca, M., Boratto, L., Carta, S.: Using behavioral data mining to produce friend recommendations in a social bookmarking system. In: Helfert, M., Holzinger, A., Belo, O., Francalanci, C. (eds.) Data Management Technologies and Applications: Third International Conference, DATA 2014, Vienna, Austria, August 29-31, 2014, Revised Selected papers, pp. 99–116. Springer International Publishing, Cham (2015)
13. Quercia, D., Capra, L.: Friendsensing: recommending friends using mobile phones. In: Bergman, L.D., Tuzhilin, A., Burke, R.D., Felfernig, A., Schmidt-Thieme, L. (eds.) Proceedings of the 2009 ACM Conference on Recommender Systems, RecSys 2009, New York, NY, USA, October 23-25, 2009. pp. 273–276. ACM (2009)
14. Zhou, T.C., Ma, H., Lyu, M.R., King, I.: Userrec: A user recommendation framework in social tagging systems. In: Fox, M., Poole, D. (eds.) Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2010, Atlanta, Georgia, USA, July 11-15, 2010. AAAI Press (2010)