1C2-4

Beyond Similarity: Serendipitous Music Recommender Systems

Sirawit Sopchoke*1 Ken-ichi Fukui*2 Masayuki Numao*2

Graduate School of Information Science and Technology, Osaka University*1
The Institute of Scientific and Industrial Research, Osaka University*2

Serendipitous music recommenders (SMRs), unlike traditional recommender system, recommend the items that user will not expect, but may find them interesting which result in a better performance. However, current available SMRs need user interaction to generate an unexpected recommendation list. Hence, we propose an automatic SMR algorithm inspired by bisociations concept, which is implemented using probabilistic logic framework.

1. Introduction

People are increasingly overwhelmed by information available online channels which seem to provide ample choices. It has become more challenging for people to find the choice that will perfectly satisfy their needs. In addition, some may have insufficient experience to make right decisions. Hence, it becomes necessary to have a tool which can scope or screen choices for the users or suggest other alternatives that users do not even know that they exist. With this tool, user will have more chances to make a better decision or select a better choice. Such tool is known as "Recommender Systems".

In developing a recommender system, the traditional techniques including content-based and collaborative filtering are commonly used. Both techniques assume that users want to see the content that is similar to what they already rate highly (content-based) or similar to the content their friends rate highly (collaborative filtering). With these techniques, users are locked into the clusters of similarity with no or low chances of discovering things that are truly novel to them. The recommendations might be boring since they are too predictable, especially for the music recommender system. The recommended music is basically similar to the previous set of music in users' activities either based on music contents (i.e., rhythm pattern, melodiousness) or music context (i.e. artist, tags).

To broaden the choices, increase the chances for users to discover music that are unexpected and truly novel and to open up a whole new musical universe to users, the notion of "serendipity" has been recently proposed. However, to generate an unexpected recommendation list in most existing algorithms require user interaction.

In this paper, we, therefore, propose an alternative automatic serendipitous music recommendations (SMR) algorithm inspired by bisociative knowledge discovery concept and is implemented using probabilistic logic framework.

2. Related Works

In this section, some of the recent works on serendipitous music recommender systems and bisociative knowledge discovery are discussed.

2.1 Serendipity

Serendipity is currently a hot topic in recommendation systems. A number of different approaches to serendipitous recommender systems have been proposed including [Shani 11], [Iaquinta 08], [Kawamae 10]. However, there are two problems commonly found when introducing serendipity into recommender systems. One problem is that the quality of predictions become progressively worse, so there is a need to find the right balance between similarity and novelty, and between the immediate surrounding and the periphery [Adamopoulos 14]. The other problem is that the proposed algorithms mostly require users' interaction to generate unexpected recommendation lists, e.g., [Stober 12] propose a music recommendation using MusicGalaxy, an adaptive user interface application for exploring music collections.

2.2 Serendipity in Music

The methodologies currently used for developing music recommender system can be classified into three different categories [Schedl 12]: music content-based, music context-based, and collaborative filtering-based. All of them share the common concept which is to create the recommendation lists based on the musical similarity.

In content-based, the similarity may be computed on some sort of acoustic features extracted from the audio signal via signal processing techniques. Alternatively, context-based approach or collaborative filtering-based approach derives the musical similarity from listening co-occurrences among users.

However, each approach has its own disadvantages. For example, the recommendation list by music content-based approach is too perfect or homogeneous, and boring since the recommendations should be similar, but too similar. In music context-based, most data is on artist level, thus too similar and no novelty. Similarly, collaborative filtering-based typically analyzes all users' model globally only one aspect which is, e.g., listening or buying. Both music context-based and collaborative filtering-based are also faced with the lack of data problem or cold-start problem, not unlike the other researchers in the field. Cold-start problem has a big impact on the quality of recommendation.

Contact: Sopchoke Sirawit, 〒 567-0047 大阪府茨木市美穂ヶ丘 8-1, 06-6879-8426, sirawit@ai.sanken.osaka-u.ac.jp

Improving recommendation performance no longer depends only on the efficiency of the similarity algorithms, the serendipitous music recommendation becomes an option for users to increase the chances to discover music that are truly novel and unexpectedly useful.

[Zhang 12] introduced the Auralist recommendation framework which attempts to inject serendipity into recommendations whilst limiting the impact on accuracy. The two novel serendipity-enhancing techniques, Community-Aware Auralist and Bubble-Aware Auralist were presented. Both can be combined with the existing ranking methods through hybrid recommendation approach.

Community-Aware Auralist interpolates a new item-based collaborative filtering algorithm called Artist-based LDA. Artist-based LDA is based on Latent Dirichlet Allocation and is combined with Listener Diversity. Bubble-Aware Auralist is the combination of Artist-based LDA and Declustering algorithm. Community-Aware Auralist and Bubble-Aware Auralist effectively boost novelty, diversity and serendipity scores, with the latter offering a better trade-off with regards to accuracy. However, these serendipity-enhancing methods still neither include the explicit user feedback to shape the algorithm interpolation for individual users nor allow the system to adapt to the adventurousness and mood of different personalities.

In this paper, we present an alternative solution to the serendipitous music recommendation problem using probabilistic logic framework to model and to generate the recommendations. This framework also provides a mechanism to extend the system by incorporating and reasoning over currently unspecified types and similarity measures of additional information collected from several sources. We also define the relations or links crossing different context domains. Thus, it leads to the use of bisociative knowledge discovery concept to find yet unexpected interesting links to generate the serendipitous recommendation lists. The details of our methodology will be discussed in the third section.

2.3 Bisociative Knowledge Discovery

Bisociative knowledge discovery focuses on the discovery of surprising relations in the repositories coming from diverse origins or heterogeneous data sources, forming different domain [Nagel 11]. The term "bisociation" was coined by Koestler in 1964[*1]. According to Koestler, the two concepts are bisociated if and only if there is no direct, obvious evidence linking them. Besides, one concept has to cross contexts to find the link and the new link provided provides some novel insight. The promising approach for bisociations data exploration is the graph structure [Berthold 12].

As mention above, [Stober 12] proposed a serendipitous music recommendation using MusicGalaxy application. They turned their music discovery application into an environment that supported bisociative music discovery. The idea is to combine two distinct domain views into one visualization using the secondary focus to highlight connections to the nearest neighbors in a different domain than the one used for projection. The "primary domain" is directly visualized by the projection and contains the displayed tracks connected by neighborhood relations that are implicitly induced between each track and its neighbors in the projection while the "secondary domain" is used to identify nearest neighbors for the secondary focus distortion and not directly visible to the user. A bisociation occurs in this setting if two tracks are not neighbors in the projection domain but are connected in the secondary domain.

However, [Stober 12] focused only on the field of bisociative music collection exploration using a user interface which integrated the music graph information for each user's interaction and led to the highlighted secondary domain focusing for serendipitous music discovery.

3. Method

To reach our goal in designing and building an automatic serendipitous music recommender system inspired by bisociative knowledge discovery concept, the designed algorithm was implemented using probabilistic logic framework.

The success of Inductive Logic Programing and Statistical Relational Learning has been shown in the number of researches. For example, the recent work of Hoxha and Rettinger in 2013 [Hoxha 13] and Kouki et al. in 2015 [Kouki 15] which applied Inductive Logic Programing and Statistical Relational Learning to the hybrid recommendations proved that incorporating additional information for users and items is beneficial to the cold-start settings in particular [Gantner 10]. The recent progress in relational learning was presented in [Hoxha 13], a probabilistic graphical modeling representation using Markov Logic Networks to combine content-based with collaborative filtering.

In the first-order logic rules, constants represent the objects in a domain of interest. Variable symbols range over the objects. Predicate symbols represent the relationship between objects or features of objects. Variables and constants might be typed, in which case variables only range over objects of the given type. The recommendation task begins with predicting the existence probability of a relation between particular user and particular item. Then, returning the recommendation for the query predicate, e.g., *rates("Bob", "book1", r)*, by choosing the items with the high probability according to the specified threshold.

In 2015, the study by [Kouki 15] showed that Statistical Relational Learning framework could be used to develop a general and extensible hybrid recommendation system framework called Hybrid Probabilistic Extensible Recommender. This framework also provided a mechanism to extend the system by incorporating and reasoning over currently unspecified types and similarity measures of additional information collected from several sources. A learning method used to appropriately balance the different input signals from many information sources was also discussed.

The evidences from the researches by [Hoxha 13] and [Kouki 15] discussed above indicate that relational learning provides a better recommendation performance by in-

∗1 Koestler, A.: *The Art of Creation*, Macmillan (1964)

corporating additional information compared to traditional method with a single dyadic relationship between the objects, i.e. users and items. Hence, the relational learning became our interest to model and provide a potential solution for serendipitous music recommendations.

We use probabilistic logic framework, a kind of relational learning in which some of the facts are annotated with probabilities, since it is suitable for representing and solving bisociative knowledge discovery problem for music recommendations. ProbLog [De Raedt 07], a probabilistic logic tool, is chosen for our implementation. ProbLog is a tool that allows us to intuitively build programs that encode complex interactions between large sets of heterogeneous component and Inherent uncertainties that are presented in real-life situations. With ProbLog, it is also easier to describe the relations between the items in dataset than other representation method such as matrix.

The ProbLog syntax is used to describe the model of item-item similarities, according to bisociations concept, as the "primary" interesting domain. This model matches the music purchased or rated by a target user to the similar music in the dataset and then adds all similar music into a recommendation list like traditional item-based collaborative filtering algorithms.

For the "secondary" domain in bisociations, we tested on both the text analysis and semantics. The lyric content of all songs was analyzed and compared in terms of semantics. Our serendipitous recommendation is the matched songs in the secondary domain which have the same semantic lyric content with the active song in primary domain and are excluded in recommendation list generated by primary domain only (Fig. 1 modified from [Berthold 12]). The initial experiment for proof of concept will be discussed in next section.

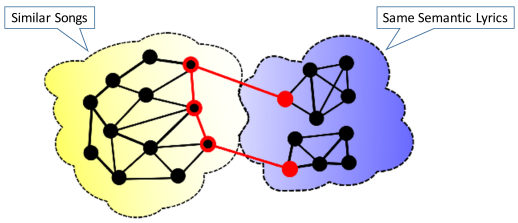

Figure 1: Find interesting bridging nodes at the intersection of the two domains

4. Experiment for Concept Proof

The goal of our study is to build a more flexible model for serendipitous music recommendations by combining the incorporated data from different aspects or different levels. We used ProbLog, the probabilistic logic framework, to implement and test a serendipitous music recommendation inspired by bisociative knowledge discovery.

4.1 Datasets

Our initial experiments were mainly conducted over a 1k Last.fm user dataset*2 which was collected by Òscar Celma in 2008. It contained the ⟨user-id, timestamp, artist-id, artist-name, song-id, song-title⟩ tuples of 1,000 users which were retrieved via Last.fm 's web service. For music description dataset which contained music content and context, the subset of One Million Songs Dataset*3 consisting of 10,000 HDF5 song files and all music features was used in addition to SQLite database files. For the lyrics, musicXmatch dataset*4 which provided the official lyrics collection of One Million Songs Dataset was used.

These three datasets were aggregated and converted into Prolog file format to form an input for our implementation.

4.2 Evaluation of Serendipity

[Adomavicius 05] suggested that recommender systems should be evaluated not only by their accuracy, but also by other important metrics such as serendipity, unexpectedness, and usefulness. In order to accurately and precisely measure the unexpectedness of candidate items and generate recommendation lists, the metric introduced by [Ge 10] was used to evaluate our method. The unexpected set of recommendations can be calculated as follows:

$$UNEXP = RS \setminus PM \tag{1}$$

where PM is a set of recommendations generated by a primitive prediction model and RS is the recommendations generated by our serendipitous recommendation. Any RS element which does not appear in PM is considered as an unexpected.

As [Ge 10] argues, unexpected recommendations may not always be useful and, hence, they also propose a serendipity measure as:

$$SRDP = \frac{|UNEXP \cap USEFUL|}{|N|} \tag{2}$$

where USEFUL is the set of useful items and N is the length of the recommendation list. For instance, the usefulness of an item can be approximated by the users' feedbacks.

4.3 Initial Experiment Result

We focus only on SRDP measure to evaluate our model which recommends the music that user does not expect, but may find it them interesting. The average SRDP result for initial experiment with similar lyric contents as the secondary domain is 35.62% (recommendation list size = 10). The initial experiment shows a positive result for serendipitous music recommendation. Additional experiment with another secondary domain definitions and results analysis is currently on going.

5. Discussion

This paper proposed a promising solution to serendipitous music recommendation problem. To develop a fully

∗2 http://www.dtic.upf.edu/∼ocelma/MusicRecommendationDataset/lastfm-1K.html
∗3 http://labrosa.ee.columbia.edu/millionsong/
∗4 http://labrosa.ee.columbia.edu/millionsong/musixmatch/

automatic serendipitous music recommender system, an SMR algorithm inspired by bisociative knowledge discovery concept was proposed. The designed algorithm was implemented using ProbLog, a probabilistic logic framework. The result of our initial experiment on Last.fm dataset with semantic lyric contents as a secondary domain in bisociations is promising and encouraging to make further experiment on this approach.

More work is currently under way to test on alternative secondary domain definitions. Music emotion is one of our interests since music could well expresses feelings and lead to serendipitous music discovery by finding the similar songs with same emotion in secondary domain in bisociations concept. Another interesting point is the pairs of people who start listening to the same song within the same time. This "same song within same time" for bisociative secondary domain is inspired by Kyle McDonald's serendipity project.[*5] According to his study, in every second, there are at least 10 pairs of people who start listening to the same song within a tenth of a second of each other. It would be interesting if we can find the links between the pairs of users who share the same music tastes. For instance, if one is listening to a popular song, there is a good chance that someone else is listening to it in sync with the person. This will be a subject for further study in details.

Acknowledgement

This work was supported by JSPS Core-to-Core Program, A. Advance Research Networks. In particular, we would like to express our deepest thanks and appreciation to Professor Dr. Luc De Raedt for his valuable guidance and advice and to Professor Dr. Hendrik Blockeel for his support during the program.

References

[Adamopoulos 14] Adamopoulos, P. and Tuzhilin, A.: On Unexpectedness in Recommender Systems: Or How to Better Expect the Unexpected, *ACM Trans. Intell. Syst. Technol.*, Vol. 5, No. 4, pp. 54:1–54:32 (2014)

[Adomavicius 05] Adomavicius, G. and Tuzhilin, A.: Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions, *IEEE Trans. on Knowl. and Data Eng.*, Vol. 17, No. 6, pp. 734–749 (2005)

[Berthold 12] Berthold, M. R. ed.: *Bisociative Knowledge Discovery: An Introduction to Concept, Algorithms, Tools, and Applications*, Springer-Verlag (2012)

[De Raedt 07] De Raedt, L., Kimmig, A., and Toivonen, H.: ProbLog: A Probabilistic Prolog and Its Application in Link Discovery, in *Proceedings of the 20th International Joint Conference on Artifical Intelligence*, IJCAI'07, pp. 2468–2473, Morgan Kaufmann Publishers Inc. (2007)

∗5 https://www.spotify.com/int/arts/serendipity/

[Gantner 10] Gantner, Z., Drumond, L., Freudenthaler, C., Rendle, S., and Schmidt-Thieme, L.: Learning Attribute-to-Feature Mappings for Cold-Start Recommendations, in *IEEE 10th International Conference on Data Mining (ICDM)*, pp. 176–185 (2010)

[Ge 10] Ge, M., Delgado-Battenfeld, C., and Jannach, D.: Beyond Accuracy: Evaluating Recommender Systems by Coverage and Serendipity, in *Proceedings of the Fourth ACM Conference on Recommender Systems*, RecSys '10, pp. 257–260, ACM (2010)

[Hoxha 13] Hoxha, J. and Rettinger, A.: First-Order Probabilistic Model for Hybrid Recommendations, in *12th International Conference on Machine Learning and Applications (ICMLA)*, Vol. 2 (2013)

[Iaquinta 08] Iaquinta, L., Gemmis, d. M., Lops, P., Semeraro, G., Filannino, M., and Molino, P.: Introducing Serendipity in a Content-Based Recommender System, in *HIS '08 Eighth International Conference on Hybrid Intelligent Systems*, pp. 168–173 (2008)

[Kawamae 10] Kawamae, N.: Serendipitous Recommendations via Innovators, in *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '10, ACM (2010)

[Kouki 15] Kouki, P., Fakhraei, S., Foulds, J., Eirinaki, M., and Getoor, L.: HyPER: A Flexible and Extensible Probabilistic Framework for Hybrid Recommender Systems, in *Proceedings of the 9th ACM Conference on Recommender Systems*, RecSys '15, pp. 99–106, ACM (2015)

[Nagel 11] Nagel, U., Thiel, K., Kötter, T., Piatek, D., and Berthold, M. R.: *Proceedings of the Advances in Intelligent Data Analysis X: 10th International Symposium (IDA)*, chapter Bisociative Discovery of Interesting Relations between Domains, pp. 306–317, Springer Berlin Heidelberg (2011)

[Schedl 12] Schedl, M., Hauger, D., and Schnitzer, D.: A Model for Serendipitous Music Retrieval, in *Proceedings of the 2Nd Workshop on Context-awareness in Retrieval and Recommendation*, CaRR '12, pp. 10–13, ACM (2012)

[Shani 11] Shani, G. and Gunawardana, A.: *Evaluating Recommendation Systems*, chapter 8, Springer (2011)

[Stober 12] Stober, S., Haun, S., and Nürnberger, A.: *Bisociative Knowledge Discovery: An Introduction to Concept, Algorithms, Tools, and Applications*, chapter Bisociative Music Discovery and Recommendation, pp. 472–483, Springer Berlin Heidelberg (2012)

[Zhang 12] Zhang, Y. C., Séaghdha, D. Ó., Quercia, D., and Jambor, T.: Auralist: Introducing Serendipity into Music Recommendation, in *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, WSDM '12, pp. 13–22, ACM (2012)