Feature Generation and Selection on the Heterogeneous Graph for Music Recommendation

Chun Guo
School of Informatics and Computing
Indiana University Bloomington
Bloomington, IN, USA, 47405
chunguo@indiana.edu

ABSTRACT

In the past decade, online music streaming services (MSS), e.g. Pandora and Spotify, experienced exponential growth. The sheer volume of music collection makes music recommendation increasingly important and the related algorithms are well-documented. In prior studies, most algorithms employed content-based model (CBM) and/or collaborative filtering (CF) [3]. The former one focuses on acoustic/signal features extracted from audio content, and the latter one investigates music rating and user listening history. Actually, MSS generated user data present significant heterogeneity. Taking user-music relationship as an example, *comment, bookmark*, and *listening history* may potentially contribute to music recommendation in very different ways. Furthermore, user and music can be implicitly related via more complex relationships, e.g., *user-play-artist-perform-music*. From this viewpoint, user-user, music-music or user-music relationship can be much more complex than the classical CF approach assumes.

For these reasons, we model music metadata and MSS generated user data in the form of a heterogeneous graph, where 6 different types of nodes interact through 16 types of relationships. We can propose many recommendation hypotheses based on the ways users and songs are connected on this graph, in the form of meta paths. The recommendation problem, then, becomes a (supervised) random walk problem on the heterogeneous graph [2]. Unlike previous heterogeneous graph mining studies, the constructed heterogeneous graph in our case is more complex, and manually formulated meta-path based hypotheses cannot guarantee good performance. In the pilot study [2], we proposed to automatically extract all the potential meta paths within a given length on the heterogeneous graph scheme, evaluate their recommendation performance on the training data, and build a learning to rank model with the best ones. Results show that the new method can significantly enhance the recommendation performance.

However, there are two problems with this approach: 1. including the individually best performing meta paths in the learning to rank model neglects the dependency between features; 2. it is very time consuming to calculate graph based features. Traditional feature selection methods would only work if all feature values are readily available, which would make this recommendation approach highly inefficient. In this proposal, we attempt to address these two problems by adapting the feature selection for ranking method (FSR) proposed by Geng, Liu, Qin, and Li [1]. This feature selection method developed specifically for learning to rank tasks evaluates features based on their importance when used alone, and their similarity between each other. Applying this method on the whole set of meta-path based features would be very costly. Alternatively, we use it on sub meta paths that are shared components of multiple full meta paths. We start from sub meta paths of length=1 and only the ones selected by FSR have the chance to grow to sub meta paths of length=2. Then we repeat this process until the selected sub meta paths grow to full ones. During each step, we drop some meta paths because they contain unselected sub meta paths. Finally, we will derive a subset of the original meta paths and save time by extracting values for fewer features. In our preliminary experiment, the proposed method outperforms the original FSR algorithm in both efficiency and effectiveness.

Keywords

Music Recommendation, Meta-Path, Feature Selection

1. REFERENCES

[1] X. Geng, T.-Y. Liu, T. Qin, and H. Li. Feature selection for ranking. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '07, pages 407–414, New York, NY, USA, 2007. ACM.

[2] C. Guo and X. Liu. Automatic feature generation on heterogeneous graph for music recommendation. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '15, pages 807–810, New York, NY, USA, 2015. ACM.

[3] Y. Song, S. Dixon, and M. Pearce. A survey of music recommendation systems and future perspectives. In *Proceedings of the 9th International Symposium on Computer Music Modeling and Retrieval*, CMMR '12, pages 395–410, 2012.