2013 Ninth International Conference on Intelligent Information Hiding and Multimedia Signal Processing

Recommending Music Based on Probabilistic Latent Semantic Analysis on Korean Radio Episodes

Ziwon Hyung, Kyogu Lee
Department of Digital Contents Convergence
Seoul National University
Seoul, South Korea
ziotoss@snu.ac.kr, kglee@snu.ac.kr

***Abstract*—Recommending music that satisfies the user's taste has been a challenging problem. Previous works on music recommendation system focused on the user's purchase history or the content of the music. In this paper, we propose a music recommendation system purely based on analyzing textual input of the users. We first mine a large corpus of Korean radio episodes, which is written by the listener. Each episode is composed of a personal story and a song request which we assume to be somehow related to the story. We then performing probabilistic Latent Semantic Analysis (pLSA) to find similar documents and recommend music that are associated to those documents. We evaluate our system by computing the mean reciprocal rank and mean average precision, which are both conventional metrics in evaluating information retrieval systems. The result shows that music similarity and document similarity are closely correlated, and thus it is possible to recommend music purely based on text analysis.**

***Keywords*-Music Recommendation, Text Mining, Probabilistic Latent Semantic Analysis**

I. Introduction

Due to significant growth in computing power and enormous expansion of digital music libraries, the need to recommend proper digital music to users has become very important. Conventionally, recommendation systems can be categorized into Collaborative filtering approach and Content-based approach. Most of the commercialized recommendation systems are based on Collaborative filtering approach. Collaborative filtering method utilizes purchase histories and item ratings to recommend music to the users. Therefore, the items must have sufficient information otherwise the system will confront a major problem, called the Cold Start problem, in which the recommender cannot process the items due to lack of sufficient information [1]. Another problem with this approach is that the diversity of the recommended music is poor [2]. This stems from the Long Tail phenomenon [3], which shows that the majority of the users consume very few popular items while very few users consume less popular items. Celma showed that the Long Tail phenomenon applies to the music industry, and hence the diversity of recommended songs based on the Collaborative filtering approach is poor [4].

Content-based approach analyzes the acoustic features [5], [6], or metadata such as genre, artist, and lyrics [7], [8], [9] to recommend music. This approach supplements the Cold Start problem and poor diversity problem of the Collaborative filtering approach. However, to analyze the music content itself requires some computational power. Thus, with modern digital music libraries, which contain millions of music data, this approach is less efficient for commercial use than the Collaborative filtering approach.

A common problem in these two approaches is that they both neglect the user's contextual information when recommending music. Reynolds *et al*. introduced the necessity to use user's contextual information when recommending music [10]. Recently, there have been several attempts to utilize this information for music recommendation. Su *et al*. extracted user's data such as heartbeat, body temperature, air temperature, noise volume, humidity, light, motion, time, season, and location. They used these data to indirectly infer the user's contextual information and by combining with content analysis, they proposed a recommendation system which proved to provide more effective recommendation list [11]. Han *et al*. allowed the users to select a situation from a predefined situation list and this situation information along with other information such as genre preference, favorite artist, age, occupation, and hobby, were used to infer the user's overall contextual information to utilize in recommending music [12]. Hariri *et al*. suggested to use mined social tags when recommending music [13]. However, these approaches indirectly infer the user's context and thus could be prone to errors in predicting the actual context.

In this paper, we propose a music recommendation system that uses text documents written explicitly by users indicating their situation. These documents contain the background for requesting a song, and hence can be interpreted as valid context.

II. Proposed method

A. Korean Radio Episodes

In order to use the text documents written explicitly by users, we mined Korean radio episodes from the radio channel's internet bulletin board. Users use the internet bulletin

978-0-7695-5120-3/13 $26.00 © 2013 IEEE
DOI 10.1109/IIH-MSP.2013.123

board to post their personal stories about an interesting event that occurred to the writer. Along with the story, the writer requests a song to be played which we assume to have close relationship to the context of the story. The staff of the radio program manually selects some of these stories and the program host introduces the story prior to playing the requested song. We believe that the stories contain situational information, and thus these documents can provide significant link between music and its relevant contextual information. Therefore, we can compute document similarity to recommend music in similar context.

B. Bag-of-Words Model

We used bag-of-words (BoW) model to represent each document as a vector composed of the frequency of words used in the document. Prior to representing the document in a word frequency vector, we performed morpheme analysis tool to remove stop-words and to discover the stem word [14]. Stop-words, such as 'and', 'the', and 'at', are irrelevant information in text analysis and can result in degrading the performance of the text analysis algorithm. Also, it is important to find the stem of a word. Without this process, words in different tense would all be regarded as different words, and as a consequence would lower the performance of the text analysis algorithm as well.

C. Probabilistic Latent Semantic Analysis

Since the number of words used in a single document is significantly small compared to the complete word dictionary, the BoW representation of the document is extremely sparse. Such sparse BoW vectors do not provide accurate analysis result and is a common problem using the BoW model. In our previous work, we showed that Latent Semantic Analysis (LSA) is a feasible algorithm in finding similar documents represented as sparse BoW vectors and thus, it is possible to use text analysis in recommending music [15]. However, in the field of language processing, probabilistic Latent Semantic Analysis (pLSA) has been proved to outperform LSA [16]. Therefore, in this paper we adopt pLSA as the text analysis algorithm to find similar documents. PLSA discovers the latent meaning between documents and words to process the sparse word-document matrix. This processed word-document matrix containing the latent meaning can then be used to find similar documents by applying conventional distance metrics, such as cosine distance or euclidean distance. PLSA uses a statistical latent class model or aspect model to estimate the transformation matrix, which discovers the underlying latent meaning from the sparse word-document matrix [17]. This is achieved by assigning probability distributions over classes to words and documents. The joint probability of document d and word w based on the latent variable z is shown in 1.

$$P(d,w) = P(d)\sum_{z} P(w|z)P(z|d) \tag{1}$$

To fit the model to a document collection, the Expectation-maximization (EM) algorithm is used. The EM algorithm consists an E step where $P(z|d,w)$ is updated and a M-step where $P(w|z)$, $P(d|z)$, and $P(z)$ are updated. The equation for the E-step is shown in 2, and the equation for the M-step is shown in 3, 4, and 5.

$$P(z|d,w) = \frac{P(z)P(d|z)P(w|z)}{\sum_{z'} P(z')P(d|z')P(w|z')} \tag{2}$$

$$P(w|z) = \frac{\sum_{d} f(d,w)P(z|d,w)}{\sum_{d,w'} f(d,w')P(z|d,w')} \tag{3}$$

$$P(d|z) = \frac{\sum_{w} f(d,w)P(z|d,w)}{\sum_{d',w} f(d',w)P(z|d',w)} \tag{4}$$

$$P(z) = \frac{\sum_{d,w} f(d,w)P(z|d,w)}{\sum_{d,w} f(d,w)} \tag{5}$$

The parameters used in the EM-algorithm are randomly initialized. Following the folding-in process described by Hofmann, we perform folding-in process to insert the query document q into the model [16]. The folding in process keeps all the $P(w|z)$ constant and recalculates $P_f(z|q)$ of the M-step while performing the same E-step mentioned above. The representation of the newly projected query document is shown in 6.

$$P_f(w|q) = \sum_{z} P(w|z)P_f(z|q) \tag{6}$$

After insertion of the query document into the model, we use cosine distance to find similar documents. Since we assumed that people prefer similar music in similar situation, once we find similar documents via pLSA of an input query document, we can recommend the associated or requested song of the top n similar documents. A controllable parameter in pLSA is the number of latent variables. We used 128 latent variables for this work and we leave as future work to find the optimal number of latent variables.

III. Evaluation

A. Dataset

We collected 15,000 documents sent by the listeners of the radio show aired between 10:00 PM and 12:00 AM, from the radio show website. As a filtering stage, we removed documents with less than 10 words since it would be close to impossible to include contextual information along with a request song using too few words, and hence would be inappropriate for training and evaluating. After the filtering stage, we had 13,520 documents which we divided into training set containing 8,364 documents and evaluation set containing the remaining 5,156 documents. For the evaluation set, we extracted the requested songs and used songs that were requested by more than 20 documents. Documents that requested the same song will be denoted as relevant documents from here on. The evaluation set is shown in Tab. I. There are total 12 songs and 383 associated documents included in the evaluation set.

B. Precision and Recall

Precision and recall is a widely used metric in evaluating information retrieval systems. Given an evaluation song S associated with relevant items $D_1, D_2, ..., D_n$, when a relevant item is found, we calculated the precision and recall until all the relevant documents were retrieved. We then averaged them to compute the average precision. For each song within the evaluation set, we used every relevant document D_i where $1 \geq i \geq n$ as an input resulting in n average precisions. We then aggregated all these to compute the mean average precision (MAP). Along with MAP we computed the interpolated precision P_{inter} at each recall level r according to 7.

$$P_{inter}(r) = max_{r' \geq r} Precision(r') \qquad (7)$$

We then obtained the eleven-point interpolated average precision which is the precision point at recall levels of 0.0, 0.1, 0.2, ..., 1.0.

C. Mean Reciprocal Rank

Another conventional information retrieval system evaluation metric is the mean reciprocal rank (MRR). For a given input document, we first find the rank of the first discovered relevant document $R_{highest}$ and inverse it to find the reciprocal rank. Similar to MAP, since we use every relevant document D_i where $1 \geq i \geq n$ as an input, we end up with n reciprocal rank for each evaluation song. Therefore, we aggregate the reciprocal ranks to compute the mean reciprocal rank (MRR) for each evaluation song.

IV. Results and Discussion

A. Results

We compared the MAP and MRR of our system with the random based MAP and MRR. To obtain the random based MAP and MRR, we first sorted the evaluation set randomly and then computed the MAP and MRR using the same evaluation set explained in Section III-A. The MAP and MRR comparison between our system and that obtained randomly is shown in Fig. 1 and Fig. 2 respectively.

In most cases our system's result outperformed that obtained randomly. Especially, songs 1, 2, 6, and 12 showed the most significant difference. We manually read the documents associated to these songs and discovered that those songs were requested in specific context. Song 12 showed an extraordinary good result so we looked at the actual documents closely. The reason for such extraordinary result was that most of the relevant documents, requesting song 12, were identical. Some of the words were different but it was basically the same, which indicates that someone was

Table I
Evaluation set with song ID, song title, and the number of relevant documents associated to the song.

Song ID	Song Title	No. Relevant Documents
1	Thank you	22
2	Start	27
3	Smiling angel	29
4	The way to me	30
5	It should be good	22
6	Two people	65
7	On the road	22
8	Become a song	29
9	I like it	48
10	Solar system	22
11	Oh my goddess	46
12	Hoi Hoi	21

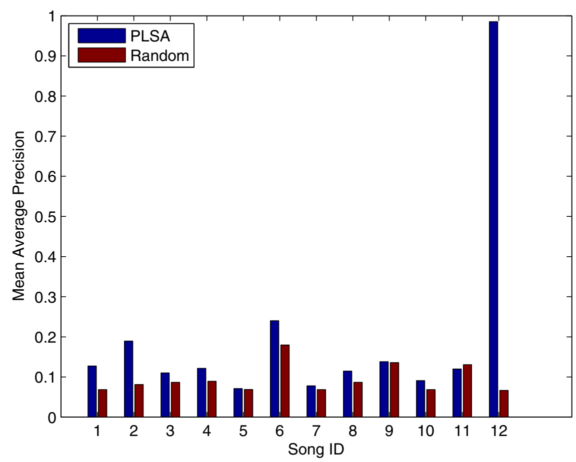

Figure 1. MAP comparison between our system and that obtained randomly.

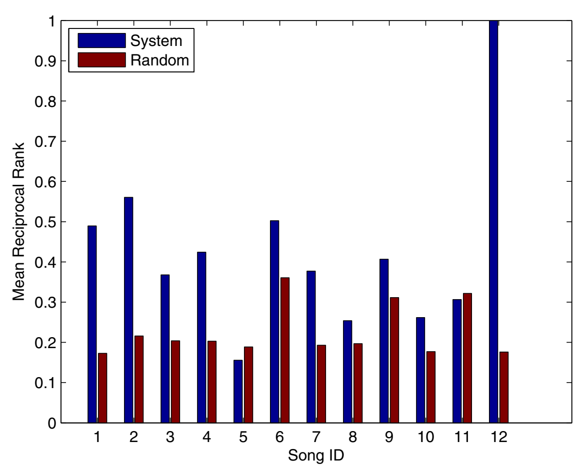

Figure 2. MRR comparison between our system and that obtained randomly.

spamming the bulletin board which we didn't filter out prior to creating the evaluation set. In the future, we plan to create a filter that could remove such spam. A possible explanation for other songs to show good result is that while the creators of the documents wrote using different words, they shared a similar situation for requesting the song. For example, the mainstream of the people requesting song 2 shared a situation where the writer wanted to start something new. This is a perfect match for the song as the lyric is about a new start and the tone of the music is also delightful. Regarding each relevant document of song 2 as an input, we computed the eleven-point interpolated average precision, which is explained in Section III-B. The result is shown in Fig. 3, which supports our assumption of people sharing similar context prefer similar music.

On the other hand, there were cases where our system didn't perform that well such as song 5, 8, and 11. A possible explanation is that the lyric of the song doesn't match the tone of the music. For example, the lyric of song 5 is about reminiscent of past love. However, the tone of the song is bright. The mismatch of lyric and the tone of the song resulted in people in different context to request this song. Some people with higher preference in lyric than the tone asked for this song reminiscing their unsuccessful love while some people with higher preference in tone than lyric requested this song in a cheerful situation. Some people requested this song because they liked the artist of the song. Due to such diversity and having no mainstream reason, our system failed to find the relevant documents.

Despite some limitations, the results for MAP and MRR were promising and thus, our assumption that people share similar music preferences in similar situations proved correct. Therefore, our approach to perform text analysis to gather contextual information when recommending music showed future potential.

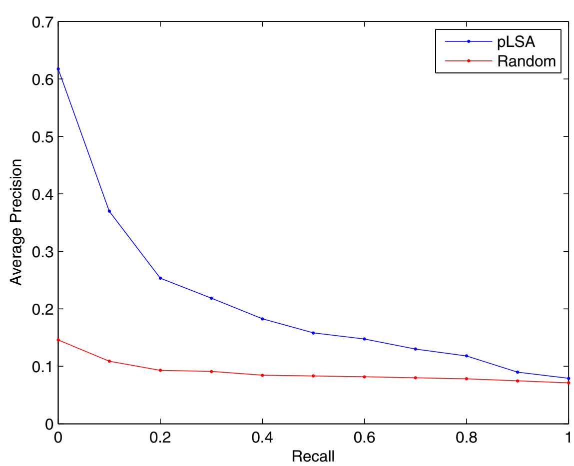

Figure 3. Eleven-point interpolated average precision for song 2.

V. Conclusion

In this paper, we proposed a context-based music recommendation system purely based on textual analysis. We showed that it is possible to use documents, written by the users, to extract the contextual information explicitly. To this end we gathered documents from the radio station's website and performed pLSA to identify the semantic meanings and find similar documents. Our assumption was that people prefer similar music in similar situation. Since each document is associated with a song request, we could simply recommend music associated to similar documents.

As mentioned in Section II-C, we intend to find the optimal number of latent variables in the future. Additionally, we intend to implement sparse coding algorithm for text analysis which is the state-of-art in machine learning. Also, we tend to expand our research by performing user evaluation to compute the actual quality of our system. Our main contribution is that we proposed a novel recommendation system purely based on text analysis. Furthermore, since our approach is purely based on text analysis, we believe that the system can be implemented to various existing social network services and remains as a future work.

VI. Acknowledgements

This work was supported under the framework of international cooperation program managed by the National Research Foundation of Korea (A307-K001).

References

[1] A. I. Schein, A. Popescul, L. H. Ungar, and D. M. Pennock, "Methods and metrics for cold-start recommendations," in *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2002, pp. 253–260.

[2] M. Zhang and N. Hurley, "Avoiding monotony: improving the diversity of recommendation lists," in *Proceedings of the 2008 ACM conference on Recommender systems*. ACM, 2008, pp. 123–130.

[3] C. Anderson, *Long tail, the, revised and updated edition: Why the future of business is selling less of more*. Hyperion, 2008.

[4] Ò. Celma and P. Cano, "From hits to niches?: or how popular artists can bias music recommendation and discovery," in *Proceedings of the 2nd KDD Workshop on Large-Scale Recommender Systems and the Netflix Prize Competition*. ACM, 2008, p. 5.

[5] Q. Li, B. M. Kim, D. H. Guan *et al.*, "A music recommender based on audio features," in *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2004, pp. 532–533.

[6] T. Li and M. Ogihara, "Content-based music similarity search and emotion detection," in *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, vol. 5. IEEE, 2004, pp. V–705.

[7] S. Pauws, W. Verhaegh, and M. Vossen, "Fast generation of optimal music playlists using local search," in *Proceedings of the 7th International Conference on Music Information Retrieval*. Citeseer, 2006, pp. 138–143.

[8] R. Ragno, C. J. Burges, and C. Herley, "Inferring similarity between music objects with application to playlist generation," in *Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval*. ACM, 2005, pp. 73–80.

[9] K. Yoshii, M. Goto, K. Komatani, T. Ogata, and H. G. Okuno, "Hybrid collaborative and content-based music recommendation using probabilistic model with latent user preferences," in *Proceedings of the International Conference on Music Information Retrieval*, 2006.

[10] G. Reynolds, D. Barry, T. Burke, and E. Coyle, "Interacting with large music collections: towards the use of environmental metadata," in *Multimedia and Expo, 2008 IEEE International Conference on*. IEEE, 2008, pp. 989–992.

[11] J.-H. Su, H.-H. Yeh, P. S. Yu, and V. S. Tseng, "Music recommendation using content and context information mining," *Intelligent Systems, IEEE*, vol. 25, no. 1, pp. 16–26, 2010.

[12] B. Han, S. Rho, S. Jun, and E. Hwang, "Music emotion classification and context-based music recommendation," *Multimedia Tools and Applications*, vol. 47, no. 3, pp. 433–460, 2010.

[13] N. Hariri, B. Mobasher, and R. Burke, "Context-aware music recommendation based on latent topic sequential patterns," in *Proceedings of the sixth ACM conference on Recommender systems*. ACM, 2012, pp. 131–138.

[14] S. Kang. (Last accessed on October 13, 2012) Korean language technology. [Online]. Available: http://nlp.kookmin.ac.kr/HAM/kor/index.html

[15] Z. Hyung, M. Lee, and K. Lee, "Music recommendation based on text mining," in *IMMM 2012, The Second International Conference on Advances in Information Mining and Management*, 2012, pp. 129–134.

[16] T. Hofmann, "Probabilistic latent semantic indexing," in *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 1999, pp. 50–57.

[17] ——, "Probabilistic latent semantic analysis," in *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 1999, pp. 289–296.