RANKING MUSIC DATA BY RELEVANCE AND IMPORTANCE

*Maria M. Ruxanda** *Alexandros Nanopoulos*** *Christian S. Jensen** *Yannis Manolopoulos***

(*) Aalborg University
Department of Computer Science
Denmark

(**) Aristotle University of Thessaloniki
Department of Informatics
Greece

ABSTRACT

Due to the rapidly increasing availability of audio files on the Web, it is relevant to augment search engines with advanced audio search functionality. In this context, the ranking of the retrieved music is an important issue. This paper proposes a music ranking method capable of flexibly fusing the music based on its relevance and importance. The fusion is controlled by a single parameter, which can be intuitively tuned by the user. The notion of authoritative music among relevant music is introduced, and social media mined from the Web is used in an innovative manner to determine both the relevance and importance of music. The proposed method may support users with diverse needs when searching for music.

1. INTRODUCTION

With the proliferation of digital music on the Web, vast musical resources are becoming available to users. Today's search engines play a major role in helping users find information on the Web. Non-textual documents such as music audio, present search engines with new challenges—how to incorporate search by musical content.

Google Music Search and Yahoo!Music represent early attempts at supporting audio search; but these services remain text-based and are limited. To achieve better audio-based search, several issues must be addressed that relate to the concept of an audio object, which can be a musical piece or any music-related entity that aggregates or relates to audio content, such as artists and albums. Key issues in audio-based search include these: how to index the content of audio objects, how to present the user with intuitive methods of querying audio objects, and which audio objects to present to the user and in which order. This paper addresses the latter issue and proposes a novel approach for music ranking, when searching for audio objects relevant to a query audio object.

2. RELATED WORK

Two notable methods, Google's PageRank and HITS [1], exist for ranking textual documents on the Web. These rank web pages according to their *importance*, which is determined by analyzing the graph structure of the links between web pages. Various extensions to multimedia ranking were proposed. A Multimedia PageRank, which uses embedded links among multimedia objects on the web, was introduced in [2]. Similarly, [3] suggested to use the surrounding textual information that appears in the web pages that contain the multimedia objects. These works propose solutions that are generic to any type of multimedia, and they do not consider the specifics of audio. However, these generic techniques may be used complementary to methods that focus on audio search.

Numerous commercial services (e.g, last.fm, pandora.com) and research prototypes (e.g., [4, 5, 6]) support music search and recommendation. These systems rank the results of music search based on their *relevance*, quantified by a music similarity measure. In recent work [4, 5], hybrid similarity measures were proposed, which combine audio extracted features with social media (playlists, tags) mined from the Web. Along the same lines, [6] proposed an audio crawler that mines mp3 blogs to find initial audio files relevant to a query, from which new music is then discovered by means of audio similarity.

In web search, there have been proposed methods that rank the search results by both their *importance and relevance* to the query. For example, a modified HITS algorithm has been proposed in [7], that weights the score of each node in a links graph with its relevance to the query.

3. MOTIVATION AND CONTRIBUTION

In analogy to web search, it is useful to rank the results of music search by both importance, which reflects popularity, and by relevance, which reflects similarity. The reason is that, importance enables the retrieval of mainstream music (authoritative artists/songs), whereas similarity enables the discovery of new music (serendipity effect). Therefore, novel methods are required to attain a fused ranking of music-search results.

This paper's contribution is two-fold. First, it introduces the notion of importance in the ranking of music search results. The use of importance enables the detection of authoritative music. In our approach, music relevance is determined by employing the established method of hybrid simi-

This research was supported in part by the Danish Research Council for Technology and Production Sciences project no. 26-04-0092 Intelligent Sound (www.intelligentsound.org).

larity [4, 5, 6]. Music importance is determined by an innovative use of social media and by using the HITS algorithm [1]. Second, we propose to apply a fusion scheme to the results of music search. Unlike the fusion scheme in [7] which is fixed, our scheme is based on kernels [8] and produces a ranking that is capable of *flexibly* ranging between *importance* and *relevance*. The paper offers experimental results that support the above ideas.

The rest of the paper is organized as follows. Section 4 describes our proposed approach and the employed techniques. Section 5 presents the experimental results. Section 6 concludes the paper and outlines future research directions.

4. PROPOSED METHOD

4.1. Outline

Ranking the results of music search by fusing their importance and relevance, requires addressing three issues: (1) how to quantify the importance of the query results in music search, (2) what similarity measure to employ to decide the relevant results, and (3) what fusion scheme to apply to combine both.

By using an analogy, music search vs. web search, we can decide the importance of the search results by analyzing the *preference* between the searched objects. In web search, the preference between pages is expressed through links. Analogously, in music search, the preference between audio objects can be expressed through social media (playlists, blogs, tags). We use social media in the form of playlists because they target specific topics, are freely available in large quantities on the Web, and can be retrieved by a crawler. By using playlists, we shall work with audio objects that are artists and songs.

Recent research work has shown that hybrid similarity measures achieve the best performance in music search and recommendation [4, 5, 6]. Therefore, we shall employ a hybrid similarity approach that combines social similarity (co-occurrence in playlists) with audio similarity.

To flexibly fuse the music-search results by both their importance and relevance to the query, we use the Neumann kernels [9], which are controlled by one single parameter, that can easily be tuned by a user according to the user's needs.

4.2. Quantifying Importance

The HITS algorithm [1] assigns so-called authority and hub scores to each web document. An underlying assumption behind HITS is that mutual reinforcement relation exists between authorities and hubs: authoritative documents are cited by many hub documents, and hub documents are those that cite many authoritative documents.

By analogy, in music search, authoritative artists or songs are those that belong in many hub playlists, and hub playlists are those that contain many authoritative artists/songs.

Let A be an inclusion matrix that represent the inclusion relation between playlists and artists/songs. An element $A(i, j)$ is 1 if playlist i contains artist/song j, and 0 otherwise.

Computing HITS using this A matrix instead of the classical adjacency matrix of a citation graph, we can obtain the authority vector, which gives the authority scores of artists/songs, as the dominant eigenvector of matrix $M = A^T A$. Thus, the top-k important audio objects are those with the highest scores in the dominant eigenvector of M.

4.3. Music Similarity Measure

Social Similarity. Co-citation coupling is a classical means of defining the relatedness between web documents. By analogy, in music search, we propose to use the co-occurrence of artists/songs in playlists. Moreover, as shown in [10], playlists are capable of offering a good measure of artists/songs relatedness. Thus, the social similarity of an (artist, song)-pair is defined as the number of playlists containing the pair.

If A is the inclusion matrix of artists/songs in playlists as defined previously, then the matrix $M = A^T A$ is exactly the co-occurrence matrix.

Audio Similarity. We compute this quantity using a state-of-art technique: Mel Frequency Cepstral Coefficients (MFCCs) modeled by Gaussian Mixture Models (GMMs). The MFCCs are extracted from each song in the data set, and their global distribution is modeled as GMMs for each artist/song. Then the similarity distance between two artists/songs A and B is computed using the log-likelihood of one model given points sampled from the other model. The distance is made symmetrical and is normalized according to the formula:

$$d(A, B) = \frac{1}{2}[logP_A(A) + logP_B(B) - logP_B(A) - logP_A(B)]$$

where $logP_B(A)$ denotes the log-likelihood of A given B.

To obtain values normalized in $[0, 1]$, where 0 means no similarity and 1 means identical artists/songs, the above distances are further normalized and subtracted from 1.

Hybrid Similarity. Audio similarity may be used for (1) reducing the sparsity of the available social data (artists/songs may lack co-occurrence information) and (2) partially regulating the inherent bias of the social data towards popularity. To achieve these, we apply the following scheme:

1. Each element $M(i, j) \neq 0$ is weighted with the audio similarity of the i^{th} and j^{th} artists/songs.

2. Each element $M(i, j) = 0$ is replaced with one of the following non-zero values: **i.** The mean of the non-zero values in row i of M, weighted with the audio similarity of the i^{th} and j^{th} artists/songs. **ii.** The audio similarity of the i^{th} and j^{th} artists/songs, if the above mean is zero.

The top-k similar audio objects with respect to the i^{th} audio object as query, are those with the highest values in row i of matrix M denoting the hybrid similarity.

4.4. Fusion Scheme

The Neumann kernel [9] was proposed for computing the semantic similarity between documents comprising of terms. It defines document similarity and term similarity by using their complementary relation. This bears reminiscence to the complementary relation of authorities and hubs in HITS.

In original form, the Neumann kernel is defined based on the term-by-document matrix X with $X(i,j)$ being the frequency of term i occurring in document j. If $K = X^T X$ is the document correlation matrix, the kernel matrix representing the semantic similarity of documents is defined as:

$$K_\lambda = K(I - \lambda K)^{-1}, \quad 0 \leq \lambda < ||K||^{-1}$$

and $||K||^{-1}$ is the reciprocal of the spectral radius of K.

The interpretation of Neumann kernels and link analysis was discussed by Shimbo and Ito [8]: if A is the adjacency matrix of a citation graph then $K = A^T A$ coincides with the co-citation matrix. When $\lambda = 0$, it is easy to see that the kernel matrix K_λ is actually the co-citation matrix. Shimbo and Ito [8] show that when λ is approaching the value $||(A^T A)||^{-1}$, the ranking induced by the Neumann kernel is identical to the authority score of HITS.

Once again using the analogy web search vs. music search, if the inclusion matrix A defined in Section 4.2 is used instead of the adjacency matrix of a citation graph, then matrix $M = A^T A$ can be used instead of $K = X^T X$ in the original definition. With no alteration to what was discussed in [8], we can use the matrix M denoting the hybrid similarity (instead of the co-occurrence similarity only). Thus, the kernel matrix M_λ defining artists/songs ranking can be computed as:

$$M_\lambda = M(I - \lambda M)^{-1}, \quad 0 \leq \lambda < ||M||^{-1}$$

The top-k audio objects kernel-ranked with respect to the i^{th} audio object as query, are those with the highest values in row i of matrix M_λ. If we denote by $p = \frac{\lambda}{||M||^{-1}}$ then the kernel ranking can be controlled by the user by tuning the parameter p in the range $[0, 1]$: the kernel ranking yields more relevant music when p approaches 0, and it yields more important music when p approaches 1.

5. EXPERIMENTAL RESULTS

For ease of subsequent comparison, the experiments reuse an existing data set[1] composed of 8764 tracks from 400 US-popular artists [10]. This data is accompanied with social media mined from two different web sources: **a.** 3,245 user collections mined from OpenNap (a popular peer-to-peer service) containing 176,113 unique collection-to-artist relations; **b.** 23,111 playlists mined from the Art of the Mix web site containing 101,157 unique playlist-to-artist relations.

The proposed method can be used for the ranking of any type of audio object. Songs are perhaps the most obvious such objects, but artists are also audio objects, as they aggregate audio information (all the songs performed by an artist). While we consider artist ranking, songs may also be ranked. From each of the social media sources, an artist co-occurrence matrix was built. We denote these matrices by M_1 (for the OpenNap collections) and M_2 (for the Art of the Mix playlists), and denote the tuning parameter of the kernel ranking by $p = \frac{\lambda}{||M_i||^{-1}}$, $i = 1, 2$.

While M_1 is dense with 95.6% non-zero values, M_2 is sparse and contains more than half zero values. To eliminate the zero values and to partially regulate the bias of social media towards popularity, M_1 and M_2 are further processed to include the audio similarity as explained in Section 4.3. Alternatively, if the sparsity of the social media is not an issue, as in the case of matrix M_1, purely social similarity can be used. In this case, where audio similarity is not used, the bias of the co-occurrence data towards popularity can be eliminated by applying an adaptation of an existing formula [10] to M_1:

$$M_{1\_new}(i,j) = \frac{M_1(i,j)}{[M_1(i,i)\dot{M}_1(j,j)]^{1-p}}$$

Experiments show that such an unbiased co-occurrence matrix is promising: when p is varied from 0 to 1, the kernel ranking varies from unbiased co-occurrence to HITS. Due to space limitations, we do not include details.

To sanity check our data, we first applied HITS on both M_1 and M_2, and we plotted the distribution of the authority scores of artists (see Figure 1.a). These results, with few clear authorities and many non-authoritative artists, conform to those of previous studies on the frequency of artists and songs in playlists [11].

For all consequent experiments, we queried with each of the 400 artists, and we computed the average results.

Figures 1.b and 1.c show clearly how the kernel ranking increases gradually from the similarity ranking (which overlaps with the x-axis) to the HITS ranking, when p is increased from 0 to 1. The curves plot the average *Kmin* distance [12] for varying numbers of top artists ($k = 5, \ldots, 30$). The *Kmin* distance quantifies the difference between two top-k rankings and when normalized, it returns 0 for identical rankings and 1 for totally different rankings. For the plotted rankings, the *Kmin* distance is relative to the (hybrid) similarity ranking. For control, we also plotted the ranking induced by the weighed HITS [7]. As mentioned in section 3, it has a fix position somewhere in the middle of the two extremes.

Figures 1.d and 1.e offer strong evidence of the fact that there is a big difference in terms of similarity and authority between the top artists rankings of the two extremes (matrix M_2 reports similar results, omitted due to space limitations). This noticeable difference explains the need for the proposed ranking approach, namely a fused ranking that makes it possible to balance flexibly between the two extremes—relevance (similarity) and importance (authority).

Figure 1.f presents the complete picture of the simultaneous effects of kernel ranking towards both ranking extremes. For the top-30 artists, the average *Kmin* distance is plotted relative to the similarity ranking and relative to the HITS ranking. As p varies from 0 to 1, the kernel ranking gradually

[1]**Acknowledgments:** We thank Dan Ellis from Columbia University, USA, for providing us with MFCCs features for the data set used.

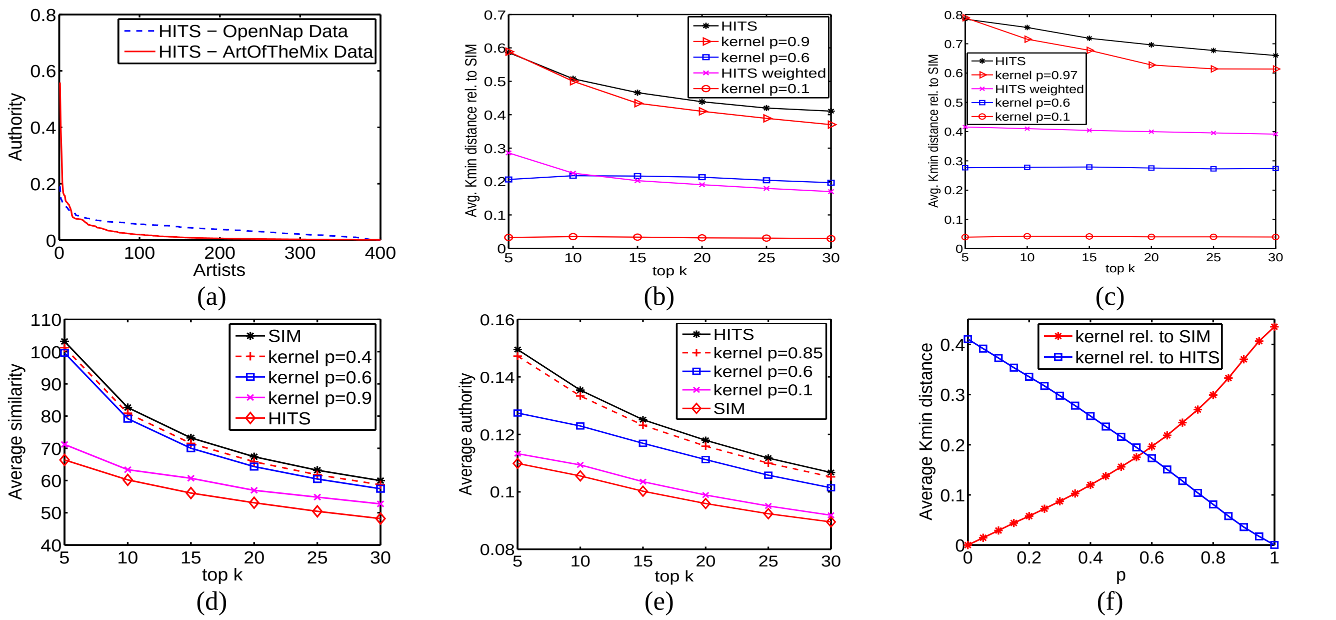

Fig. 1. Experimental results: **a.** Authority distribution of artists. **b.** and **c.** Rankings relative to similarity (SIM) ranking, for M_1 and M_2, respectively. **d.** M_1: Average similarity of top ranked artists. **e.** M_1: Average authority of top ranked artists. **f.** M_1: Kernel ranking when varying parameter p.

moves from similarity towards HITS, with an equilibrium point at around the median value of p.

6. CONCLUSION AND FUTURE WORK

We proposed a ranking-music method that is capable of flexibly fusing the relevance and the importance of music. The method's fusion capabilities address the diverse needs among users and can be tuned easily and intuitively. The ranking capabilities utilize a novel notion of authoritative music together with social media mined from the web. The proposed method can be useful when incorporated into music search engines.

In future work, other fusion schemes such as Laplacian or diffusion kernels will be investigated. Such kernels are controlled by more parameters than the Neumann kernels, but can better address the bias of social media towards popularity. As well, user studies will be conducted, to evaluate user's satisfaction with respect to the proposed music ranking method.

7. REFERENCES

[1] J.M. Kleinberg, "Authoritative sources in a hyperlinked environment," in *Proc. ACM-SIAM*, 1998, pp. 668–677.

[2] C.C. Yang and K.Y. Chan, "Retrieving multimedia web objects based on pagerank algorithm," in *Proc. WWW*, 2005, pp. 906–907.

[3] M. Sahami, V. Mittal, S. Baluja, and H. Rowley, "The happy searcher: Challenges in web information retrieval," in *Proc. PRICAI*, 2004, pp. 3–12.

[4] P. Knees, T. Pohle, M. Schedl, and G. Widmer, "A music search engine built upon audio-based and web-based similarity measures," in *Proc. SIGIR*, 2007, pp. 447–454.

[5] R. Stenzel and T. Kamps, "Improving content-based similarity measures by training a collaborative model," in *Proc. ISMIR*, 2005, pp. 264–271.

[6] O. Celma and P. Herrera, "Search sounds: An audio crawler focused on weblogs," in *Proc. ISMIR*, 2006, pp. 104–111.

[7] K. Bharat and M.R. Henzinger, "Improved algorithms for topic distillation in a hyperlinked environment," in *Proc. SIGIR*, 1998, pp. 104–111.

[8] M. Shimbo and T. Ito, *chapter: Kernels as Link Analysis Measures, in Mining Graph Data*, pp. 283–310, Wiley-Interscience, 2007.

[9] J. Kandola, J. Shawe-Taylor, and N. Cristianini, "Learning semantic similarity," in *Proc. NIPS*, 2002, pp. 657–664.

[10] A. Berenzweig, B. Logan, D. Ellis, and B. Whitman, "A large-scale evaluation of acoustic and subjective music similarity measures," *Computer Music Journal*, vol. 28:2, pp. 63–67, 2004.

[11] Y. Dou, E. Itoh, S. Hirokawa, and D. Ikeda, "An approach to analyzing correlation between songs/artists using itms playlists," in *Proc. CIMCA*, 2005, pp. 951–956.

[12] R. Fagin, R. Kumar, and D. Sivakumar, "Comparing top k lists," in *Proc. SODA*, 2003, pp. 28–36.