ARE TAGS BETTER THAN AUDIO FEATURES? THE EFFECT OF JOINT USE OF TAGS AND AUDIO CONTENT FEATURES FOR ARTISTIC STYLE CLUSTERING

Dingding Wang
School of Computer Science
Florida International University
Miami, FL USA
dwang003@cs.fiu.edu

Tao Li
School of Computer Science
Florida International University
Miami, FL USA
taoli@cs.fiu.edu

Mitsunori Ogihara
Department of Computer Science
University of Miami
Coral Gables, FL USA
ogihara@cs.miami.edu

ABSTRACT

Social tags are receiving growing interests in information retrieval. In music information retrieval previous research has demonstrated that tags can assist in music classification and clustering. This paper studies the problem of combining tags and audio contents for artistic style clustering. After studying the effectiveness of using tags and audio contents separately for clustering, this paper proposes a novel language model that makes use of both data sources. Experiments with various methods for combining feature sets demonstrate that tag features are more useful than audio content features for style clustering and that the proposed model can marginally improve clustering performance by combing tags and audio contents.

1. INTRODUCTION

The rapid growth of music the Internet both in quantity and in diversity has raised the importance of music style analysis (e.g., music style classification and clustering) in music information retrieval research [10]. Since a music style is generally included in a music genre (e.g., the style Progressive Rock within the genre of Rock) a style provides finer categorization of music than its enclosing genre. Also, for much the same reason that all music in a single genre has some commonality, all music in a single style has some commonality belonging to a same style, and the degree of commonality is stronger within a style than within its enclosing genre. These properties suggest that by way of appropriate music analysis, it is possible to computationally organize music sources into not only musicologically meaningful groups but also into hierarchical clusters that reflect style and genre similarities. Such organizations are likely to enable efficient browsing and navigation of music items.

Much of the past work on music style analysis methods is based solely on audio contents and various feature extraction methods have been tested. For example, [32] presents a study on music classification using short-time analysis along with data mining techniques to distinguish among five music styles. Pampalk et al. [17] combine different similarity sources based on fluctuation patterns and use a nearest neighbor classifier to categorize music items. More recently Chen and Chen [3] use long-term and short-term features that represent the time-varying behavior of music and apply support vector machines (SVM) to classify music into genres. Although these audio-content-based classification methods are successful, music style classification and clustering are difficult problems to tackle, in part because music style classes are more numerous than music genres and thus computation quickly reaches a limit in terms of the number of styles to classify music into. One then naturally asks whether adding non-audio features push style classification/clustering beyond the limit of audio-feature-based analysis.

Fortunately, the rapid development of web technologies has made available a large quantity of non-acoustic information about music, including lyrics and social tags, latter of which can be collected by a variety of approaches [24]. There has already been some work toward social tag based music information retrieval [1,11,13,16,23]. For example, Levy and Sandler [16] demonstrate that the co-occurrence patterns of words in social tags are highly effective in capturing music similarity, Bischoff et al. [1] discuss the potential of different kinds of tags for improving music search, and Symeonidis et al. [23] propose a music recommendation system by performing latent semantic analysis and dimensionality reduction using the higher order SVD technique on a user-tag-item tensor.

In this paper we consider social tags as the source of non-audio information. We naturally ask whether we can effectively combine the non-audio and audio information sources to improve performance of music retrieval. Some prior work has demonstrated that using both text and audio features can improve the ranking quality in music search systems. For example, Turnbull et al. [25] successfully combine audio-content features (MFCC and Chroma) with social tags via machine learning methods for music searching and ranking. Also, Knees et al. [12] incorporate audio contents into a text-based similarity ranking process.

However, few efforts have been made to examine the effect of combining tags and audio-contents for music style analysis. We thus the question of, given tags and representative pieces for each artist of concern, whether the tags and the audio-contents of the representative pieces complement each other with respect to artist style clustering, and if so, how efficiently those pieces of information can be combined.

In this paper, we study the above questions by treating the artist style clustering problem as an unsupervised clustering problem. We first apply various clustering algorithms using tags and audio features separately, and examine the usefulness of the two data sources for style clustering. Then we propose a new tag+content (TC) model for integrating tags and audio contents. A set of experiments is conducted on a small data set to compare our model with other methods, and then we explore whether combining the two information sources can improve the clustering performance or not.

The rest of this paper is organized as follows. In Section 2 we briefly discuss the related work. In Section 3 we introduce our proposed TC model for combining tags and contents for artist style clustering. We conduct comprehensive experiments on a real world dataset and the experimental results are presented in Section 4. Section 5 concludes.

2. RELATED WORK

Audio content based automatic music analysis (clustering, classification, and similarity search in particular) is one of the most important topics in music information retrieval. The most widely used audio features are timbral texture features (see, e.g., [26]), which usually consist of Short Term Fourier Transform (STFT) and Mel-Frequency Cepstral Coefficients (MFCC) [20]. Researchers have applied various data mining and statistically methods on these features for classifying or clustering artists, albums, and songs (see, e.g., [3, 5, 18, 19, 26]).

Music social tags have recently emerged as a popular information source for curating music collections on the web and for enabling visitors of such collections to express their feelings about particular artists, albums, and pieces. Social tags are free-text descriptions of any length (though in practice there sometimes is a limit in terms of number of characters) with no restriction on the words that are used. Social tags thus can be as simple as a single word and as complicated as a long, full sentence. Popular short tags include *heavy rock, black metal, and indie pop* and long tags can be like "I love you baby, can I have some more?"

As can be easily seen social tags are not as formal as descriptions that experts such as musicologists provide. However, by collecting a large number of tags for one single piece of music or for one single artist, it seems possible to gain understanding of how the song or the artist is received by the general listeners. As Lamere and Pampalk point out [13] social tags are widely used to enhance simple search, similarity analysis, and clustering of music items [13]. Lehwark, Risi, and Ultsi [15] use Emergent-Self-Organizing-Maps (ESOM) and U-Map techniques on tagged music data to conduct clustering and visualization in music collections. Levy and Sandler [16] apply latent semantic dimension reduction methods to discover new semantics from social tags for music. Karydisi et al. [11] propose a tensor-based algorithm to cluster music items using 3-way relational data involving song, users, and tags.

In the information retrieval community a few attempts have been made to complement document clustering using user-generated tags as an additional information source (see, e.g., [21]). In such work the role that social tags play is only supplementary because the texts appearing in the original data are, naturally, highly more informative than tags.

The situation in the MIR community seems different from this and the use of tags seems to show much stronger promise. This is because audio contents, which are the standard source of information, have to go through feature extraction for syntactic or semantic understanding and thus the distance between the original data source and the tag in terms of informativeness appears to be much smaller in MIR than in IR.

There has been some work exploring the effectiveness of joint use of the two types of information sources for retrieval, including including the work in [25] and [12] where audio contents and tags are combined for searching and ranking and the work in [30] that attempts to integrate audio contents and tags for multi-label classification of music styles. These prior efforts are concerned with supervised learning (i.e., classification) while the present paper is concerned with unsupervised learning (i.e., clustering).

3. TAG+CONTENT MODEL (TC)

Here we present our novel language model for integrating tags and audio contents and how to use the model for artistic style clustering.

3.1 The Model

Let $\mathcal{A}$ be the set of artists of interest, $\mathcal{S}$ the set of styles of interest, and $\mathcal{T}$ the set of tags of interest. We assume that for each artist, for each style, and for each artist-style pair, its tag set (as a multiset in which same elements may be repeated more than once) is generated by mutually independent selections. That is, for each artist $a \in \mathcal{A}$ and for each nonempty set of tags $t = (t_1, \ldots, t_n)$, $t_1, \ldots, t_n \in \mathcal{T}$, we define the language model, $p(t \mid a)$, by

$$p(t \mid a) = \prod_{i=1}^{n} p(t_i \mid a)$$

Similarly, for each style $s \in \mathcal{S}$, we define its language model $p(t \mid s)$, by

$$p(t \mid s) = \prod_{i=1}^{n} p(t_i \mid s)$$

Although we might want to consider the artist-style joint language model $p(t \mid a, s)$, we assume that the model is

dictated only by the style and that it is independent of the artist. Thus, we assume

$$p(t \mid a, s) = p(t \mid s)$$

for all tags $t \in \mathcal{T}$. Then the artist language model can be decomposed into several common style language models:

$$p(t \mid a) = \sum_{s \in \mathcal{S}} p(t \mid s) p(s \mid a).$$

Instead of directly choosing one style for artist a, we assume that the style language models are mixtures of some models for the artists linked to a, i.e.,

$$p(s \mid a) = \sum_{b \in \mathcal{A}} p(s \mid b) p(b \mid a),$$

where b is an artist linked to artist a. Combining these yields the following model:

$$p(\vec{t} \mid a) = \prod_{i=1}^{n} \sum_{s \in \mathcal{S}} \sum_{b \in \mathcal{A}} p(t_i \mid s) p(s \mid b) p(b \mid a).$$

We use the empirical distribution of the observed artists similarity graph for $p(b \mid a)$ and let $\mathbf{B}_{b,a} = \tilde{p}(b \mid a)$. The model parameters are $(\mathbf{U}, \mathbf{V})$, where

$$\mathbf{U}_{t,s} = p(t \mid s), \quad \mathbf{V}_{b,s} = p(b \mid s).$$

Thus, $p(t_i \mid a) = [\mathbf{U}\mathbf{V}^\top \mathbf{B}]_{t,a}$.

The artist similarity graph can be obtained using methods described in Section 3.2. Now we take the Dirichlet distribution, the conjugate prior of multinomial distribution, as the prior distribution of $\mathbf{U}$ and $\mathbf{V}$. The parameter estimation is maximum a posteriori (MAP) estimation. The task is

$$\mathbf{U}, \mathbf{V} = \arg\min_{\mathbf{U},\mathbf{V}} \ell(\mathbf{U}, \mathbf{V}), \tag{1}$$

where $\ell(\mathbf{U}, \mathbf{V}) = \mathrm{KL}\left(\mathbf{A} \| \mathbf{U}\mathbf{V}^\top \mathbf{B}\right) - \ln \Pr(\mathbf{U}, \mathbf{V})$.

Using an algorithm similar to the nonnegative matrix factorization (NMF) algorithm in [14], we obtain the following updating rules:

$$\begin{aligned} \mathbf{U}_{ts} &\leftarrow \mathbf{U}_{ts} \left[\mathbf{C}\mathbf{B}^\top \mathbf{V}\right]_{ts} \\ \mathbf{V}_{bs} &\leftarrow \mathbf{V}_{bs} \left[\mathbf{B}\mathbf{C}^\top \mathbf{U}\right]_{bs} \end{aligned}$$

where $\mathbf{C}_{ij} = \mathbf{A}_{ij} / [\mathbf{U}\mathbf{V}^\top \mathbf{B}]_{ij}$. The computational algorithm is given in Section 3.3.

3.2 Artist Similarity Graph Construction

Based on the audio content features, we can construct the artist similarity graph using one of the following popular methods, which is due to Zhu [33].

ϵ NN graphs A strategy for artist graph construction is the ϵ-nearest neighbor algorithm based on the distance between the feature values of two artists. For a pair of artists i and j, if the distance $d(i, j)$ is at most ϵ, draw an edge between them. The parameter ϵ controls the neighborhood radius. For the distance measure d, the Euclidean distance is used throughout the experiments.

exp-weighted graphs This is a continuous weighting scheme where $W_{ij} = \exp(-d(i, j)^2 / \alpha^2)$. The parameter α controls the decay rate and is set to 0.05 empirically.

3.3 The Algorithm

Algorithm 1 is our method for estimating the model parameters.

Algorithm 1 Parameter Estimation

Input: $\mathbf{A}$: tag-artist matrix.
$\mathbf{B}$: artist-artist relation matrix;
Output: $\mathbf{U}$: tag-style matrix;
$\mathbf{V}$: artist-style matrix.

begin
1. **Initialization:**
Initialize $\mathbf{U}$ and $\mathbf{V}$ randomly,
2. **Iteration:**
repeat
2.1 Compute $\mathbf{C}_{ij} = \mathbf{A}_{ij} / [\mathbf{U}\mathbf{V}^\top \mathbf{B}]_{ij}$;
2.2 Assign $\mathbf{U}_{ts} \leftarrow \mathbf{U}_{ts} \left[\mathbf{C}\mathbf{B}^\top \mathbf{V}\right]_{ts}$,
2.3 Compute $\mathbf{C}_{ij} = \mathbf{A}_{ij} / [\mathbf{B}\mathbf{U}\mathbf{V}^\top]_{ij}$;
2.4 Assign $\mathbf{V}_{bs} \leftarrow \mathbf{V}_{bs} \left[\mathbf{B}\mathbf{C}^\top \mathbf{U}\right]_{bs}$,
until convergence
3. **Return** $\mathbf{V}$
end

3.4 Relations with Other Models

The TC model uses mixtures of some existing base language models as topic language models. The model is different with some well-known topic models such as Probabilistic Latent Semantic Indexing (PLSI) [8] or Latent Dirichlet Allocation (LDA) [2] since they assume the topic distribution of each object is independent of those of others. However, this assumption does not always hold in practice since in music style analysis, artists (as well as songs) are usually related to each other in certain ways. Our TC model incorporates an external information source to model such relationships among artists. Also, when the base matrix $\mathbf{B}$ is an identity matrix, this model is identical to PLSI (or LDA), and the algorithm is the same as the NMF algorithm with Kullback-Leibler (KL) divergence loss [6, 29].

4. EXPERIMENTS

4.1 Data Set

For experimental purpose, we use the data set in [30]. The data set consists of 403 artists and one representative song per artist. The style and tag descriptions are obtained respectively from All Music Guide and Last.fm, as described below.

4.1.1 Music Tag Information

Tags were collected from Last.fm (http://www.last.fm). A total of 8,529 tags were collected. The number of tags for an artist ranged from 3 to 100. On average an artist had 89.5 tags. Note that, the tag set is a multiset in that the same tag may be assigned to the same artist more than once. For example, Michael Jackson was assigned "80s" for 453 times.

4.1.2 Audio Content Features

For each song we extracted 30 seconds of audio after the first 60 seconds. Then from each of the 30-second audio clips, we extracted 12 timbral features using short-term Fourier transform following the method described in [27]. The twelve features are based on Spectral Centroid, Spectral Rolloff, and Spectral Flux. For each of these three spectral dynamics, we calculate the mean and the standard deviation over a sliding window of 40 frames. Then from these means and variances we compute the mean and the standard deviation across the entire 30 seconds, which results in $2 \times 2 \times 3 = 12$ features. We mention here that we actually began our exploration with a much larger feature set of size 80, which included STFT, MFCC, and DWCH, but in an attempt to improve results all the features but STFT were consolidated which was consistent with the observations in [9].

4.1.3 Style Information

Style information was collected from All Music Guide (http://www.allmusic.com). All Music Guide's data are all created by musicologists. Style terms are nouns like Rock & Roll, Greek Folk, and Chinese Pop as well as adjectives like Joyous, Energetic, and New Romantic. Styles for each artist/track are different from the music tags described in the above, since each style name appears only once for each artist. We group the styles into five clusters, and assign each artist to one style cluster. In the experiments, the five groups of styles are: (1) Dance-Pop, Pop/Rock, Club/Dance, etc., consisting of 100 artists including *Michael Jackson*; (2) Urban, Motown, New Jack Swing, etc., consisting of 72 artists including *Bell Biv DeVoe*; (3) Free Jazz, Avant-Garden, Modern Creative, etc., consisting of 51 artists including *Air Band*; (4) Hip-Hop, Electronica, and etc., consisting 70 artists including *Afrika Bambaataa*; (5) Heavy Metal, Hard Rock, etc., consisting of 110 artists including *Aerosmith*.

4.2 Baselines

We compare our proposed method with several state-of-the-art clustering methods including K-means, spectral clustering (Ncuts) [31], and NMF [14]. For each clustering method, we perform it on two data matrices, i.e., the tag-artist matrix and the content-artist matrix, respectively. We also perform them on an artist similarity graph which is the linear combination of two similarity graphs generated based on tags and contents respectively using the graph construction method described in Section 3.2. NMF is not suitable for symmetric similarity matrices, there exists its symmetric matrix version, SNMF [28]. We use SNMF to deal with the artist similarity matrix. We also use PHITS-PLSI, a probabilistic model [4] which is a weighted sum of PLSI and PHITS, to integrate tag and audio content information for artist clustering. The summary of the baseline methods is listed in Table 4.2.

clustering methods	tags only	content only	both
K-means	√	√	√
Ncuts	√	√	√
NMF	√	√	
SNMF			√
PHITS-PLSA			√

Table 1. The implemented baseline methods.

	K-means	Ncuts	NMF
Accuracy	0.2953	0.4119	0.4020
NMI	0.0570	0.1166	0.1298

Table 2. Clustering results using tag information only.

4.3 Evaluation Methods

To measure the clustering quality, we use accuracy and normalized mutual information (NMI) as performance measures.

- Accuracy measures the relationship between each cluster and the ground truth class assignments. It is the total matching degree between all pairs of clusters and classes. The greater accuracy, the better clustering performance.
- NMI [22] measures the amount of statistical information shared by two random variables representing cluster assignment and underlying class label.

4.4 Experimental Results

4.4.1 Tags-only or Content-only

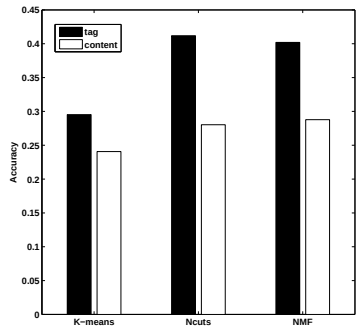
Tables 2 and 3 respectively show the clustering performance using tag information only and the performance using content features only. We observe that the tags are more effective than the audio content features for artist style clustering. Figure 1 better illustrates this observation.

4.4.2 Combining Tags and Content

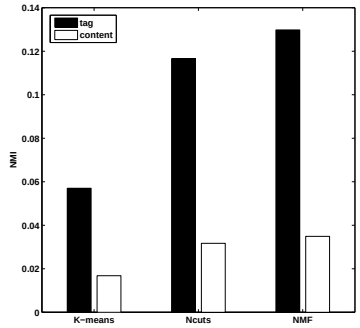
Table 4.4.2 show the performance of different clustering methods using both tag and content information. Since the

	K-means	Ncuts	NMF
Accuracy	0.2407	0.2803	0.2878
NMI	0.0168	0.0317	0.0349

Table 3. Clustering results using content features only.



(a) Accuracy



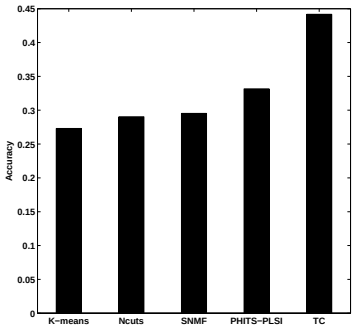
(b) NMI

Figure 1. Clustering performance using tag or content information.

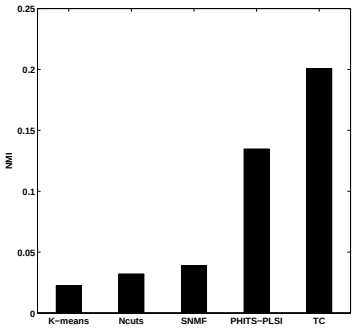
first three clustering algorithms are originally designed for clustering one data matrix, we first construct an artist similarity graph as follows. (1) We compute the pairwise Euclidean distances of artists using the tag-artist matrix (normalized by tags (rows)) to obtain a symmetric distance matrix d_t, and another distance matrix d_c can be calculated in the similar way using the content- artist matrix. (2) Since d_t and d_c are in the same scale, we can simply combine them linearly to obtain the pairwise artist distance. (3) The corresponding artist similarity graph can be constructed using the strategies introduced in Section 3.2. Once the artist similarity graph is generated, the clustering can be conducted using any clustering method. Since both PHITS-PLSI and our proposed method are designed to combine two types of information, we can directly use the tag-artist matrix as the original data matrix, and the similarity graph is constructed based on content features. Figure 2 illustrates the results visually.

From the results, we observe the following:

- The artist clustering performance is not necessarily improved by incorporating content features. This means that the tags are more informative than contents for clustering artist styles.
- Advanced methods, e.g. PHITS-PLSI and our proposed method, can naturally integrate different types of information and they outperform other traditional clustering methods. In addition, our proposed method outperforms PHITS-PLSI because PHITS-PLSI is more suitable for incorporating explicit link information while our method is more suitable for handling implicit links (graph).
- Continuous similarity graph construction such as exp-weighted method performs better than discrete methods, e.g. ϵ NN.
- Our proposed method with combined tags and contents using ϵ NN graph construction outperforms all the methods using only tag information. This demonstrates our model is effective for combining different sources of information, although the content features do not contribute much.



(a) Accuracy



(b) NMI

Figure 2. Clustering performance combining tags and contents.

5. CONCLUSION

In this paper, we study artistic style clustering based on two types of data sources, i.e., user-generated tags and audio content features. A novel language model is also proposed to make use of both types of information. Experimental results on a real world data set demonstrate that tag information is more effective than music content information for artistic style clustering, and our model-based method can marginally improve the clustering performance by combining tags and contents. However, other simple combination methods fail to enhance the clustering results by incorporating content features into tag-based analysis.

		K-means	Ncuts	SNMF	PHITS-PLSI	TC
ϵ NN graph	Acc	0.2680	0.2804	0.2630	0.3152	0.3648
	NMI	0.0193	0.0312	0.0261	0.0709	0.1587
exp-weighted graph	Acc	0.2730	0.2903	0.2953	0.3316	0.4417
	NMI	0.0226	0.0321	0.0389	0.1347	0.2008

Table 4. Clustering results combining tags and content.

6. ACKNOWLEDGMENT

The work is partially supported by the FIU Dissertation Year Fellowship, NSF grants IIS-0546280, CCF-0939179, and CCF-0958490, and an NIH Grant 1-RC2-HG005668-01.

7. REFERENCES

[1] K. Bischoff, C. Firan, W. Nejdl, and R. Paiu: "Can all tags be used for search?," *Proceedings of CIKM*, 2008.

[2] D. Blei, A. Ng, and M. Jordan: "Latent Dirichlet allocation," *NIPS*, 2002.

[3] S. Chen and S. Chen: "Content-based music genre classification Usinu timbral feature vectors and support vector machine," *Proceedings of ICIS*, 2009.

[4] D. Cohn and T. Hofmann: "The missing link - a probabilistic model of document content and hypertext connectivity," *NIPS*, 2000.

[5] H. Deshpande, R. Singh, and U. Nam: "Classification of music signals in the visual domain," *Proceedings of the the COST-G6 Conference on Digital Audio Effects*, 2001.

[6] C. Ding, T. Li, and W. Peng: "On the equivalence between Non-negative Matrix Factorization and Probabilistic Latent Semantic Indexing,". *Comput. Stat. Data Anal.*, 52(8):3913-3927.

[7] C. Ding, T. Li, W. Peng, and H. Park. "Orthogonal nonnegative matrix tri-factorizations for clustering," *SIGKDD*, 2006.

[8] T. Hofmann: "Probabilistic latent semantic indexing," *SIGIR*, 1999.

[9] T. Li, M. Ogihara, and Q. Li: "A comparative study on content-based music genre classification," *SIGIR*, 2003.

[10] T. Li and M. Ogihara: "Towards intelligent music information retrieval," *IEEE Transactions on Multimedia*, 8(3):564-575, 2006.

[11] I. Karydis, A. Nanopoulos, H. Gabriel, and M. Spiliopoulou: "Tag-aware spectral clustering of music items," *ISMIR*, pp. 159–164, 2009.

[12] P. Knees, T. Pohle, M. Schedl, D. Schnitzer, K. Seyerlehner, and G. Widmer: "Augmenting text-based music retrieval with audio similarity," *ISMIR*, 2009.

[13] P. Lamere and E. Pampalk: "Social tags and music information Retrieval," *ISMIR*, 2008.

[14] D. Lee and H. Seung: "Algorithms for non-negative matrix factorization," *NIPS*, 2001.

[15] P. Lehwark, S. Risi, and A. Ultsch: "Data analysis, machine learning and applications," in *Visualization and Clustering of Tagged Music Data*, pp. 673–680. Springer Berlin Heidelberg, 2008.

[16] M. Levy and M. Sandler: "Learning latent semantic models for music from social tags" *Journal of New Music Research*, 37:137–150, 2008.

[17] E. Pampalk, A. Flexer, and G. Widmer: "Improvements of audio-based music similarity and genre classificaton," *ISMIR*, 2005.

[18] W. Peng, T. Li, and M. Ogihara: "Music clustering with constraints," *ISMIR*, 2007.

[19] D. Pye: "Content-based methods for managing electronic music," *ISCASSP*, 2000.

[20] L. Rabiner and B. Juang: *Fundamentals of Speech Recognition*, Prentice-Hall, NJ, 1993.

[21] D. Ramage, P. Heymann, C. Manning, and H. Garcia: "Clustering the tagged web," *ACM International Conference on Web Search and Data Mining*, 2009.

[22] A. Strehl and J. Ghosh: "Clustering ensembles - a knowledge reuse framework for combining multiple partitions," *Journal of Machine Learning Research*, 3:583-617, 2003.

[23] P. Symeonidis, M. Ruxanda, A. Nanopoulos, and Y. Manolopoulos: "Ternary semantic analysis of social tags for personalized music Recommendation," *ISMIR*, 2008.

[24] D. Turnbull, L. Barrington, and G. Lanckriet: "Five approaches to collecting tags for music," *ISMIR*, 2008.

[25] D. Turnbull, L. Barrington, M. Yazdani, and G. Lanckriet: "Combining audio content and social context for semantic music discovery," *SIGIR*, 2009.

[26] G. Tzanetakis and P. Cook: "Musical Genre Classification of Audio Signals," *IEEE Transactions on Speech and Audio Processing*, 10:5, 2002.

[27] G. Tzanetakis: "Marsyas submissions to MIREX 2007," *MIREX 2007*.

[28] D. Wang, S. Zhu, T. Li, and C. Ding: "Multi-document summarization via sentence-level semantic analysis and symmetric matrix factorization," *SIGIR*, 2008.

[29] D. Wang, S. Zhu, T. Li, Y. Chi, and Y. Gong: "Integrating clustering and multi-document summarization to improve document understanding," in *CIKM*. pp. 1435-1436, 2008.

[30] F. Wang, X. Wang, B. Shao, T. Li, and M. Ogihara: "Tag integrated multi-label music style classification with hypergraph," in *ISMIR*, pp. 363–368, 2008.

[31] J. Shi and J. Malik: "Normalized Cuts and Image Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):888–905, 2002.

[32] Y. Zhang and J. Zhou: "A study on content-based music Classification," *IEEE Signal Processing and Its Applications*, 2003.

[33] X. Zhu: "Semi-supervised learning with graphs," *Doctoral Thesis*, Carnegie Mellon University, 2005.