ENRICHING MUSIC MOOD ANNOTATION BY SEMANTIC ASSOCIATION REASONING

Jun Wang, Xavier Anguera, Xiaoou Chen, Deshun Yang
Institute of Computer Science and Technology, Peking University, Beijing, P.R. China 100871
Telefonica Research, Via Augusta 177, Barcelona, Spain

ABSTRACT

Mood annotation of music is challenging as it concerns not only audio content but also extra-musical information. It is a representative research topic about how to traverse the well-known semantic gap. In this paper, we propose a new music-mood-specific ontology. Novel ontology-based semantic reasoning methods are applied to effectively bridge content-based information with web-based resources. Also, the system can automatically discover closely relevant semantics for music mood and thus a novel weighting method is proposed for mood propagation. Experiments show that the proposed method outperforms purely content-based methods and significantly enhances the mood prediction accuracy. Furthermore, evaluations show the system's accuracy could be promisingly increased with the enrichment of metadata.

Keywords— Social music, Mood, Semantic reasoning, Ontology, Annotation

1. INTRODUCTION

Semantic information retrieval of music is becoming more and more important based on the ever-growing predominance and amount of available digital music. Recent music information retrieval (MIR) research pays increasing attention to reveal the aspects of music composition and performance which can influence listeners' emotional responses. A comprehensive investigation in [1] gives high importance for affective/emotive descriptors, and indicates that people enjoy discovering new music by entering mood-based queries. Automatic music mood tagging and annotation can poten[...] of the core technologies for a wide [...] Currently content-b[...]inant player[...] in the field of auto[...]. With M[...] researchers' growing in[...]sifying music by moods, significant progress is being made in the content-based area. One of the most representative works is the Audio Mood Classification (AMC) evaluation task held by the Music Information Retrieval Evaluation eXchange (MIREX) [2]. In the AMC evaluation, the top resulting accuracies for 5-cluster mood classification progressed from 61.5% in 2007, to 63.7% in 2008, and to 65.67% in 2009.

Nevertheless, music mood classification based only on audio-based methods is quite an arduous challenge because of its subjectivity and the influence of social and cultural factors. Generally, the acoustic aspect is only one side of music. Another important aspect to mood classification is how the acoustical events are combined, which constitutes the high level semantic part of music. Besides, socio-cultural aspects are also very important, since certain other music perceptual aspects can not be captured well enough by content-only based algorithms, as the categorization into pieces such as neutral, autumnal and passionate are too abstract and too subtle to be recognized by audio features only. Therefore, alternative sources of information, possibly extra-musical, need to be incorporated in order to further improve the results of music mood classification.

Nowadays the Web has become a primary host of a sizeable amount of text-based and semantic information. Technologies such as Web 2.0-- e.g., Last.fm [1], MusicBrainz[2]-- have drastically augmented social media with rich context, such as user-provided tags, comments, editorial metadata, etc. This allows researchers to bring up novel web-based methods for MIR tasks. Some works [3] [4], etc. have recently proposed working on the automatic extraction of cultural metadata about music from the web. In conclusion, combining information from sources like web-based, text and other sorts of multimodal information with content-based features in an efficient way could be one of the solutions to break the bottleneck of pure content-based method.

The extension of semantic annotation to the field of knowledge representation for[...] not limited to the music domain. In the m[...] community, great [...] been given [...]ects of [...]ntologies. Ther[...]n the [...]sed for manag[...] [...]ding image [...]ation [5, 6, [...] the domain c[...] and framework[...] [...] high level semantics. For exa[...] et al. [5] give a concrete example considering th[...]tion of images of a sports event, and show how retrieval and interpretation of image data can be obtained by abductive

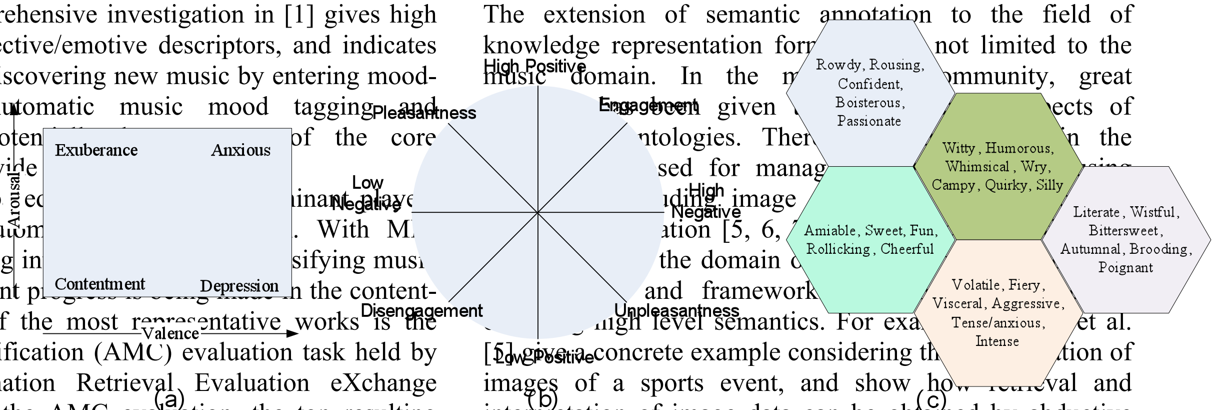

Fig. 1. Different mood models: (a) TWC model; (b) Thayer's model; (c) MIREX Adjective-cluster model.

reasoning. All these works are using the benefits of ontology systems, which have scalability and extendibility capabilities to achieve effective image retrieval. However, to the best of our knowledge, music annotation enrichment via combining high level semantic aspects like social metadata with low level features has been comparatively rarely studied in an ontology-based system. In the paper we present an ontology-based system and the novel points of our work can be summarized as follows:

1) Combining the low-level with the social metadata-based information for mood annotation via an ontology-based system.
2) Automatically discovering closely relevant semantics. In our case, detecting high-level concepts that are closely related to mood.
3) Discovering new knowledge. It is proven to predict mood with promising accuracy, even if the background knowledge is not necessarily directly related to mood.

2. RELATED WORK

Existing mood annotation research is mainly about how to automatically classify music into mood taxonomy, based on the audio signal itself. It generally includes the following parts: mood taxonomy modeling, acoustic feature extraction, features subset selection (FSS) and classification methods.

Existing mood models in the MIR field could be categorized into two classes: dimensional models and adjective-cluster models. There are several widely used dimensional models, including Tellegen-Watson-Clark (TWC) model [9] and Thayer's model [10] [11], as shown in Fig. 1 (a) and (b). On the other hand, ISMIR MIREX group [2] has proposed a five-cluster model for automatic music mood classification, as shown in Fig. 1 (c).

Previous works reveal that certain sophisticated features have significant relevance with music mood perception. The most commonly mentioned features can be grouped into intensity features, timbre features and rhythm features. Intensity features are correlated to "energy", "amplitude" of the signal. Timbre features are spectral shape features. Rhythm features are analysis of rhythm regularity, rhythm strength, beat, tempo, etc.

Yang et al. [10] use all features above plus psychoacoustic features to construct a 114 dimension feature space. The psychoacoustic features represent parameters based on some psychoacoustic models. They use FSS algorithm to decrease the large feature space, and greatly improves accuracy. K-nearest neighbors (k-NN) classifier is adopted in their work. Instead of using FSS algorithms, Lie Lu et al. [11] divide their feature subset according to physical attributes. Gaussian Mixture Models (GMM) is adopted as classifier. They report a classification accuracy as high as 86.3% by combining the timbre and rhythm features for a task limited to a classical music album, which has much less ambiguity in mood. Moreover, SVM has been found in many cases [2] [12] superior to the other classifiers.

[1] http://www.last.fm/
[2] http://musicbrainz.org/

3. PROPOSED SYSTEM

3.1. System Overview

As shown in Fig. 2, the proposed system's framework evolves from low semantic concept level (audio signal) to high semantic concept level (mood). In the web-based extraction module, we extract information from the web, including socio-cultural tags, editorial data, etc. In the audio feature extraction part, we apply a state-of-the-art music mood machine learning method based on SVM. Both the information from SVM and web are stored in an extended Music Ontology. At last, we predict mood annotation with semantic association via semantic reasoning.

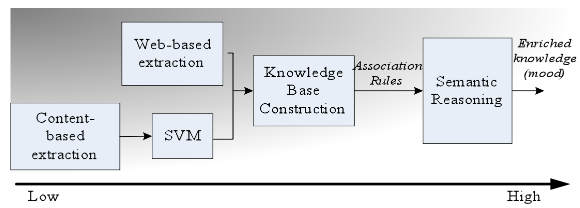

Fig. 2. Framework of the proposed system.

3.2. Knowledge Base Construction

TBox is an ontology's terminology, and ABox is the ontology's assertional axioms. Music Ontology (MO) [13] is a generic ontology in music domain well-designed for the extendibility, especially in terms of its extendibility of content-based information and Linked Data services. To follow the coherence principle of ontology development, the mood-oriented TBox in our work is constructed on the base of MO terms and has music-mood-specific refinement. There are two main parts of TBox refinement: the "Web-based part" refines high-level social metadata information; the "Audio-based part" refines audio-based information. Besides, the ABox underlines the combination of the content-based and high-level social metadata information.

3.2.1. TBox Construction

Professional databases, web services and ontologies are resources created by professional data entry staff, editors, and writers. There are rich editorial metadata such as names, titles, product numbers, biographies, nationalities, reviews etc., relational content such as similar artists and albums, influences, etc. There are standard taxonomies force objects into predefined categories and the information are normally very precise, trustful and useful. AllMusic[3] is a large-scale music database that provides professional reviews and metadata for albums, tracks and artists. Its datasets are frequently used for MIR research purposes [2],

[3] http://allmusic.com

[11]. We construct TBox by refining MO with the taxonomies of Allmusic and explicitly represent socio-cultural association information and semantic links (SLs) between terms of metadata concepts. Due to space limitation, the detailed SLs and class hierarchies are omitted here. Some typical SLs among different concepts are demonstrated in a triple <Domain, Property, Range> format as follows:

<MusicArtist/ Track/ Style/ SocialTag/..., express, MusicMood>

<Genre, genre, MusicArtist/ Track/ ...>

<Instrument, instrument, MusicArtist/ Track/ ...>

<Style, style, MusicArtist/ Track/ ...>

<Style, style, MusicArtist/ Track/ ...>

<SocialTag, tag, MusicArtist/ Track/ ...>

<Genre/ Instrument/ MusicArtist/ Style/ Track, similarTo, Genre/ Instrument/ MusicArtist/ Style/ Track>

<MusicArtist, made, Track>

<MusicArtist, follows/ followedBy/ influences/ influencedBy/ inSameBand/ coworkWith, MusicArtist>

<Track, publicationOf, Signal>

For comparison with state-of-the-art approaches, we adopt the MIREX Adjective-Cluster Model [2] mood classification taxonomy model and LibSVM [15] as automatic classification tool. Probabilistic representation is important as music mood information is intrinsically both complex and uncertain. In this work the mechanism for representing probabilities is by defining data-type properties for domains with probabilistic memberships. We defined a data-type property "estimateMood", which describes the probability of a certain mood that an object has. Given the MIREX Adjective-Cluster Model, we further refined it into five sub-properties: "moodCluster1", "moodCluster2",… "moodCluster5".

The ontology-based system has a principal advantage that it is applicable to facilitate semantic reasoning among different concepts. The SLs and properties can later be exploited in section 4 to conduct semantic reasoning for mood annotation.

3.2.2. ABox Construction

According to TBox, ABox is constructed with information extracted from two sources: raw audio and web information. We extract web-based information from metadata-rich websites such as Last.fm, AllMusic, etc., which support abundant ID3 metadata, tags, annotations, editorial information, comments, etc. Moreover, emerging resources such as Linking Open Data (LOD)[4] could be used to extend the Web by publishing various open music data sets as RDF on the Web and by setting RDF links between data items from different data sources. Typical LOD resources in the music domain include MusicBrainz, AudioScrobbler, Jamendo, Magnatune, etc. Since our system is based on ontology and OWL, it could be more convenient to extract data from LOD. Yet in practice, due to state of the art of current available LOD resources, the available RDF data-- in particular of the music-mood-related domain-- is still rather sparse. As a result, we currently extract data from metadata-rich websites in order to get enough information to conveniently demonstrate our method. For this work, we crawled Last.fm tags and identify their classes to WordNet concept hierarchies. In the collection of over 60,000 tags, 47% are identified into genre subclasses, 13% into mood subclasses, 3% into usage subclasses and 2% into instrument subclasses. We crawled web pages from AllMusic, and parsed the HTML content and tags into MMR nodes and links, and map the structured information into the KB.

To predict mood classes via raw audio, we need to estimate the probability that the music mood belongs to any of N classes, i.e., $mpv_i = P(y = i \mid \chi),\ \ i = 1,2,...,N$, given a set of observed feature vectors χ. The feature space χ was constructed by selecting features which have been successfully applied to mood/emotion classification in existing state of the art: On the one hand, we extract a 138-

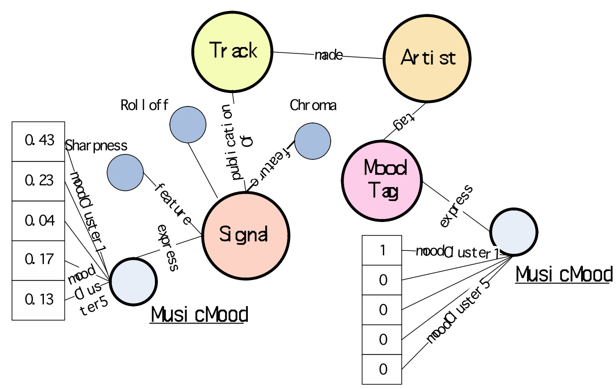

Fig. 3 Illustration of initiation of mood annotation in knowledge base

dimension feature set using the Marsyas and PsySound softwares [10], then apply FSS algorithm (implemented as a toolkit in Weka [16]) to reduce the feature dimension; On the other hand, We model Mel-frequency cepstral coefficients (MFCCs) with GMM, and then apply Probability Product Kernel (PPK) as the SVM kernel. Radial Basis Function (RBF) kernel is applied for the other features. To combine PPK and RBF kernel in SVM we adopt a similar method as proposed in [3], where they find SVMs trained using the combined kernel is superior to SVMs trained using individual kernel for music retrieval. Then we use LibSVM as the classifier tool, and applied it to output the mood probablities $\text{MPV} = \{mpv_i\},\ \ i = 1,2,...,N$ (N equals 5 in our case).

As shown in Fig.3, a group of nodes in the KB can act as "seed nodes": for MoodTag nodes which contain the same adjectives in a certain mood cluster i, we initiate their

4 http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

linked MusicMood nodes with data property moodCluster i =1; Secondly, for Signal nodes we initiate their linked MusicMood nodes with the data property moodCluster i using values of $MPV = \{mpv_i\}$ obtained via the content-based extraction method as described above.

4. SEMANTIC REASONING

Until now, we have constructed the KB for the ontology-based system, then semantic reasoning can be conducted on the KB to discover new knowledge. Firstly, a global propagation step is processed to mine the association weighting of different concept nodes in KB. Then the weighted propagation is performed to decide the mood cluster probabilities of target nodes.

The processing of global Propagation is illustrated in Fig.4 in the left. Seed nodes are firstly considered as activated nodes S and propagate their data properties into their semantic linking target nodes, as denoted by the curves with "1"; then in the following iteration, the activated target nodes keep being propagated to their semantic linking nodes, as denoted by the curves with "2". The iteration continues

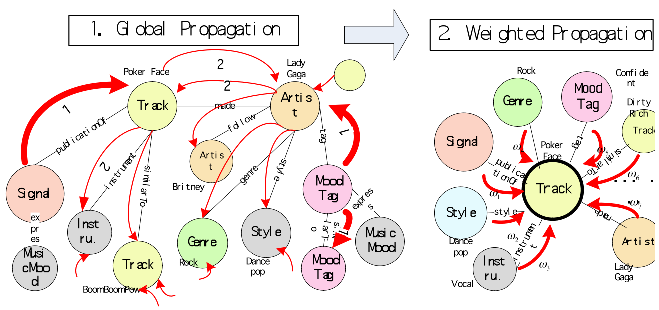

Fig. 4 Illustration for global propagation and weighted propagation

until reaching certain limits, e.g. the percentage of activated nodes. For brevity, in Fig.4 we only illustrates two iterations.

Global propagation can be efficiently implemented via New Racerpro Query Language (NRQL) in Racerpro [14] on the whole KB. Some DL-safe non-recursive rules in our work are illustrated below:

$$\begin{aligned} \textit{Rule I}: \ musicCluster_i(l,p) \leftarrow\ & x \in Track \wedge y \in Signal \wedge y \in S \\ & \wedge z, l \in MusicMood \wedge express(x,l) \\ & \wedge express(y,z) \wedge musicCluster_i(z,p) \end{aligned}$$

RuleI means that a Track x is the publication of a Signal y in the activated set S, and y has been attached class probability value p to its MusicMood data property $musicCluster_i$, then the Track x has p as the value of its i th sub-property of Mood, i.e. $musicCluster_i$.

$$\begin{aligned} \textit{Rule II}: \ musicCluster_i(l,p') \leftarrow\ & x, y \in MoodTag \wedge z, l \in MusicMood \\ & \wedge express(x,l) \wedge similarTo(x,y) \wedge y \in S \\ & \wedge express(y,z) \wedge musicCluster_i(z,p) \end{aligned}$$

Rule II means that a MoodTag x is similar to a MoodTag y in the activated set S, and y has been attached class probability value p to its MusicMood data property $musicCluster_i$, then the MoodTag x has p' as the value of its sub-property of mood $musicCluster_i$, where p' is the weighted mean of the original data property value and p.

Similarly, there are reasoning rules among other different semantic concepts. The following *Rule III* and *Rule IV* are two illustrative examples:

$$\begin{aligned} \textit{Rule III}: \ musicCluster_i(l,p') \leftarrow\ & x \in MusicArtist \wedge y \in Style \wedge y \in S \\ & \wedge z, l \in MusicMood \wedge express(x,l) \\ & \wedge style(x,y) \wedge express(y,z) \wedge musicCluster_i(z,p) \end{aligned}$$

$$\begin{aligned} \textit{Rule IV}: \ musicCluster_i(l,p') \leftarrow\ & x, y \in MusicArtist \wedge y \in S \\ & \wedge z, l \in MusicMood \wedge express(x,l) \wedge followedBy(x,y) \\ & \wedge express(y,z) \wedge musicCluster_i(z,p) \end{aligned}$$

Intuitively not all semantic nodes linking to a track necessarily have a close semantic association in terms of mood. This motivates the weighted propagation step to finally predict the music mood for a track, as shown in the right part of Fig.4.

Global propagation step takes a cumulative processing over the whole KB and it is able to reveal interesting implicit knowledge that is "unknown". For example, in our system after global propagation, the genre node "Children's" is attached with MPV = {0.0463, 0.5328, 0.022, 0.3987, 0}, which indicates that "Children's" has very dominant music mood in cluster 2 – "amiable, sweet, fun, etc.". Meanwhile, it also reveals some semantic concepts which do NOT have a close semantic association with mood. For example, over the 25 Instrument nodes evaluated in our system, most of them are attached with MPV that has very vague classification among different mood clusters.

To scale the extent to which a semantic concept is related to mood, we propose to examine ω as defined in Eq.1:

$$\omega = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(mpv_i - \overline{mpv})^2} \qquad (1)$$

where mpv_i is the i th mpv value attached to the node, $i = 1,2,...N$, and N is the number of mood clusters. ω indicates how biased a node is to be classified into a particular mood cluster. For instance, if a node has MPV = {0.9, 0.0, 0.1, 0.0, 0.0}, which can be clearly classified into Cluster1, its ω equals 0.3937 and is relatively high; in contrast, if a node has MPV = {0.2, 0.2, 0.2, 0.2, 0.2}, which means it is not biased to any mood cluster, its ω equals 0. Fig.5 gives the histogram of ω for several different semantic concept classes examined in our system: from the histogram, we can observe that most of the MoodTag nodes have a very high ω. It indicates that MoodTag class has a very close semantic association with mood. In contrast, most of the Instrument nodes have a very low ω. It indicates that this class has alienated or very noisy associations with mood. Moreover, many of Genre and MusicArtist nodes have comparatively high ω . This indicates some remarkable associations between some genre and mood, and between some artist and mood. Consequently, we consider it reasonable to set weighting factor (WF) being function or proportional to ω. In our current work, we simply attached each class with a WF, as given in Eq.2, instead of attaching each node with different WF, however, it would be interesting to look into setting WF for each node in future work.

$$WF_{ClassK} = \alpha * \sum_{all\ n_j \in ClassK} \omega_j * \frac{1}{\sum_{all\ n_j \in ClassK} 1} \quad (2)$$

where α is a constant factor, K is the indicator of different classes, ω_j is related to node n_j belonging to class K.

At last, the mood probability of a track can be predicted by weighted propagation: firstly the track is initiated with $RuleI$, then a set of DL-safe rules are applied to implement the weighted propagation. An illustrative rule is listed as below:

$$Rule\ V: \ musicCluster_i(l, p') \leftarrow x \in Track \wedge y \in Genre \wedge y \in S \wedge z, l \in MusicMood \wedge express(x,l) \wedge genre(y,x) \wedge express(y,z) \wedge musicCluster_i(z,p)$$

$Rule\ V$ means that a Track x has Genre y, and y has been attached class probability value p to its $musicCluster_i$, then the Track x has p' as the value of its MusicMood $musicCluster_i$, where p' is the weighting mean of the original value with weighted p (with weighting factor WF_{Genre}).

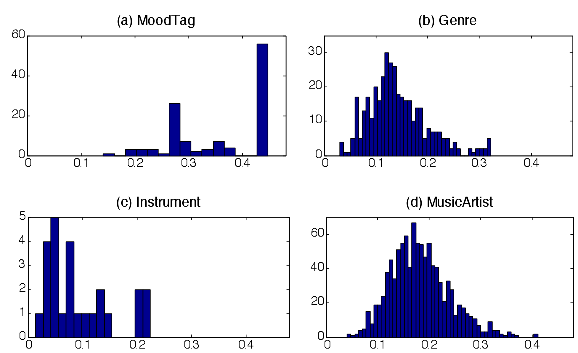

Fig.5. Histograms of ω for classes (a) MoodTag, (b) Genre, (c) Instrument and (d) MusicArtist.

For each track node, similar non-recursive rules like $Rule\ V$ are applied until all its linked nodes are traversed. Finally normalization is implemented to ensure the sum of all cluster probabilities equals to 1.

5. EXPERIMENTS

5.1. Dataset

We collected an audio album, consisting of 1804 tracks, covering about 21 major genres and 56 sub genres, and including 1022 different artists. These tracks are pre-labeled with one of 5 mood clusters according to their metadata provided by AMG professional editors, and distributed almost evenly for each cluster. Each track is processed into 30s, mono, 22.05 kHz, .wav audio clips. The album is divided into three independent sub-datasets TRAIN_FSS, TRAIN_SVM and TEST. TRAIN_FSS (145 tracks) is the dataset used for training the FSS; TRAIN_SVM (400 tracks) is the dataset used for training SVM. TEST (1259 tracks) is the dataset applied to test the FSS and SVM. The tracks in each sub-datasets are evenly distributed among clusters.

We crawled the web-based information from AllMusic and Last.fm as described in section 3.2.2. Here are some statistics of the web-based information: 101 “Genre”, 387 “style”, 179 “MoodTag” and 25 “Instrument”; 1027 “KnownArtist” -- artists that are well labeled with metadata like genre, style, instrument, etc.; 11496 “UnknownArtist” -- artists without any metadata except some editorial information of their social association with “KnownArtist”, such as “similar to”, “follows”, “influences”, etc. On the other hand, we implemented our content-based system as described in section 3.2.2, namely PPK-RBF, and output its content-based classification result $MPV = \{mpv_i\}$ to initiate the data property moodCluster i of related ABox nodes. To better compare with state-of-the-art approaches in the future, we’ve published the dataset and KB[5].

5.2. Evaluation of Content-based and Semantic Reasoning Method

MIREX datasets (MIREX_TEST) are NOT attainable as they are secretive for the contest. To try our best to look into state-of-the-art systems, we took one of the systems in MIREX 2009 Mood Classification contest, i.e. FCY1 (skipping the citation due to anonymous review).. The result of FCY1 tested on MIREX_TEST, TEST, and PPK-RBF tested on TEST are given in the upper part and the lower left part of Tab.1.

The average accuracy of PPK-RBF on TEST is 39.64%. The average accuracy of FCY1 degrades from 60.33% on MIREX_TEST to 39.32% on TEST. This mainly due to the following reasons: 1) Album diversity and scale: 600 tracks of 27 genres in MIREX_TEST and 1259 tracks of 101 genres in TEST; 2) MIREX_TEST greatly eliminates the album’s ambiguity among different clusters. It is claimed by them to have an important influence on system performances. For realistic systems, however, it is not practical to have a dataset pre-filtered by human assessors, especially for large-scale social media dataset.

In the above tested systems, Cluster1 and Cluster2 are more or less badly confused with Cluster4, which might reveals certain limitation of content-based systems. We consider that, for the moods in such clusters, it is hard to capture the distinguishing information purely via audio features.

[5] Skipped here due to the anonymous submission

In the following, we examine the Mood Semantic Reasoning (MoodSR) system which uses the semantic reasoning method to bridge content-based annotation with web-based information. MoodSR goes through the steps as described in section 4. After the "weighted propagation" step, the track nodes take their data properties MPV as their probabilistic mood annotations, and are classified into the mood cluster with the highest probability.

The tests are run on the same dataset, i.e. TEST. The lower right part of Tab.1 shows the confusion matrix for MoodSR system. Comparing to the average accuracy of 39.32% for FCY1 system and 39.64% for PPK-RBF system, MoodSR system achieves an average accuracy of 60.6%. It shows that MoodSR performs remarkably better prediction accuracies than FCY1 and PPK-RBF. The result proves that the semantic reasoning method is an effective way to combine web-based information with audio-based features. Especially, it improves significantly for Cluster1 and Cluster2. This demonstrates that the extra-musical information gathered from the web is able to capture some complicated semantic part of music mood.

Table 1. Confusion matrix of FCY1, PPK-RBF and MoodSR on MIREX_TEST and TEST dataset

True	FCY1 (MIREX_TEST)					FCY1 (TEST)				
	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
C1	**0.53**	0.10	0.02	0.15	0.14	0.03	0.03	0.01	0.03	0.02
C2	0.10	**0.43**	0.11	0.21	0.01	0.32	**0.36**	0.07	0.18	0.14
C3	0.08	0.22	**0.82**	0.12	0.02	0.12	0.25	**0.64**	0.18	0.12
C4	0.14	0.25	0.06	**0.52**	0.11	**0.34**	0.32	0.25	**0.50**	0.34
C5	0.15	0.01	0	0.01	**0.73**	0.19	0.04	0.03	0.11	**0.38**
	PPK-RBF (TEST)					MoodSR (TEST)				
	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
C1	0.24	0.18	0.04	0.13	0.14	**0.54**	0.22	0.03	0.09	0.14
C2	0.08	0.06	0.01	0.04	0.01	0.09	**0.44**	0.05	0.06	0.00
C3	0.11	0.24	**0.64**	0.17	0.07	0.11	0.18	**0.77**	0.18	0.09
C4	**0.36**	**0.46**	0.25	**0.51**	0.32	0.11	0.15	0.10	**0.59**	0.09
C5	0.21	0.06	0.05	0.15	**0.46**	0.15	0.01	0.05	0.07	**0.68**

5.3. KB Enrichment Simulation

A community-contributed social media system could get saturated with more and more data. In traditional systems, however, the metadata enrichment doesn't necessarily benefit improving mood annotation as they are considered as disjunctive attributes. For example, in a social music system, more artists are annotated with metadata such as genres, similar artists, styles, etc., yet this enrichment can hardly affect the mood annotation.

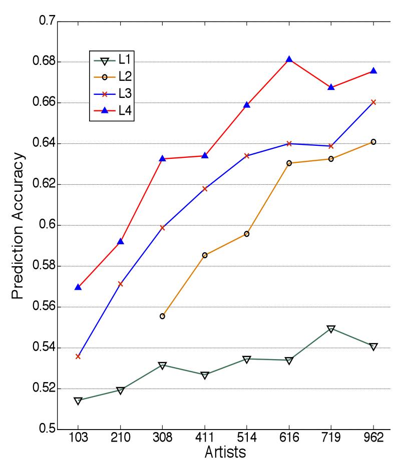

Fig.6. Prediction accuracy varying

To evaluate how our system's mood prediction performance varies while it gains more metadata information-- in particular-- which are not related to mood, we simulate an enrichment progress by increasing the amount of artists annotated with metadata (genres, styles, instruments, followers, similar artists, followed artists, etc.). As shown in Fig.6, the x-axis indicates the number of such artists. They are randomly selected from the original 1027 KnownArtist nodes.

Meanwhile, we evaluate how the performance varies while applying different reasoning rules and propagation iteration times, as shown in Fig.6:

L1: applying rules with two iterations of propagation between MusicArtist and other metadata classes including Genre, Style, Track, Instrument, etc.;

L2: applying rules with two iterations of propagation inside MusicArtist and Track classes exclusively;

L3: applying rules with three iterations of propagation inside MusicArtist and Track classes exclusively;

L4: applying rules with three iterations of propagation inside MusicArtist and Track classes as well as propagation between MusicArtist and other metadata classes.

We conduct the above evaluation on the whole KB. As shown in Fig.6, in general the mood prediction accuracies rise almost consistently when increasing the amount of artists annotated with metadata. Among different lines, L1 gives the lowest accuracy. This may be due to that L1 only considers semantic links between artist and other metadata nodes, and these nodes such as "Genre" are of low Granularity and can only obtain rough prediction via global propagation. L2 considers semantic links among artists and has richer contributing nodes than L1. One more iteration is applied for L3 than L2, which helps obtaining better prediction accuracy. Finally, L4 gives the best prediction accuracy, since it considers comprehensive semantic links among different classes for reasoning mood. To sum up, the evaluation proves that the proposed ontology-based system turns disjunctive knowledge into useful, efficient, special-purpose expertise, and the semantic reasoning method can naturally generate new assertions, which are not inferable in attribute-based systems, with promising accuracies. The accuracies keep improving while the system gains more knowledge which hardly ever aids mood annotation in traditional systems.

6. CONCLUSION

In this paper, we look at music mood annotation problem and suggest an ontology-based method to efficiently bridge audio content with web-based information. A new propagation algorithm using semantic reasoning method is proposed for the first time in the domain of music mood annotation. To improve the propagation efficiency, a weighted propagation method is proposed in our work, seen as an adequate way to reveal the association between mood and high-level metadata. The evaluation result on our test dataset shows that our system outperforms pure content-based method and improves the mood prediction accuracy. Moreover, the annotation accuracy can benefit from the enrichment of other metadata. This feature is particularly promising for a large-scale social media environment, where the social media are augmented with ever growing rich context and social metadata information.

More future work is stimulated, such as mapping tags and folksonomies to domain taxonomies, learning rules from training knowledge base, etc. One of our ongoing work is about a semi-automatical approach of association rule mining.

7. REFERENCES

[1] M. Lesaffre, M. Leman, J.P. Martens, "A user-oriented approach to music information retrieval", in Content-Based Retrieval. Wadern Germany: Dagstuhl Seminar Proceedings, 2006.

[2] X. Hu, J. S. Downie, C. Laurier, M. Bay, A. F. Ehmann, "The 2007 MIREX Audio Mood Classification Task: Lessons Learned", ISMIR'08.

[3] L. Barrington, D. Turnbull, M. Yazdani, G. Lanckriet, "Combining Audio Content and Social Context for Semantic Music Discovery", ACM SIGIR'09, 2009.

[4] M. Levy, M. Sandler, "Music Information Retrieval Using Social Tags and Audio", IEEE Trans. on Multimedia, vol.11, No. 3. April 2009.

[5] S. E. Peraldi, A. Kaya, S. Melzer, R. Moller, M. Wessel, "Multimedia Interpretation as Abduction", 2007.

[6] A. Penta, A. Picariello, L. Tanca, "Multimedia Knowledge Management Using Ontologies", IEEE Intelligent Systems, 2003.

[7] L. Ballan, M. Bertini, A. D. Bimbo, and G. Serra, "Video Annotation and Retrieval Using Ontologies and Rule Learning", IEEE Multimedia, 2010.

[8] A. Ferrara, L. Ludovico, S. Montanelli, S. Castano, and G. Haus, "A Semantic Web Ontology for Context-based Classification and Retrieval of Music Resources", Transactions on Multimedia Computing, Communications and Applications, Vol. V, April, 2006.

[9] D. Yang, W. Lee, "Disambiguating music emotion using software agents", In Proc. Int. Conf. Music Inf. Retrieval, 2004.

[10] Y.H. Yang, Y.C. Lin, Y.F. Su, and H. H. Chen, "A Regression Approach to Music Emotion Recognition", IEEE Trans. on Audio, Speech, and Language Processing, Vol. 16, No. 2, 2008.

[11] L. Lu, D. Liu, and H. J. Zhang. Automatic Mood Detection and Tracking of Music Audio Signals. IEEE TRANS on Audio, Speech, and Language Processing, Vol. 14, No. 1, Jan. 2006.

[12] S. R. Ness, A. Theocharis, G. Tzanetakis, L. G. Martins, "Improving Automatic Music Tag Annotation Using Stacked Generalization Of Probabilistic SVM Outputs",ACM Multimedia'09, Oct., 2009.

[13] F. Giasson and Y. Raimond, "Music ontology specification. Online ontology", 2008.

[14] V. Haarslev, R. Moller, M. Wessel, "RacerPro User's Guide and Reference Manual Version 1.9.2", 2007.

[15] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines", 2001 [Online]. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

[16] P.; Witten I.H. 2009. "The Weka Data Mining Software: An Update", SIGKDD Explorations, Vol. 11, Issue1.