Categorizing Music Mood in Social Context

Xiao Hu
Graduate School of Library and Information Science,
University of Illinois at Urbana-Champaign
501 E. Daniel St.; Champaign, IL 61820
xiaohu@illinois.edu

Music mood is an emerging metadata type of music, but there are no well-accepted mood categories. This research proposes a new method to categorize music mood in the social context of music listening. This method combines the strength of social tags, linguistic resources and human experts. Preliminary results show that the proposed method is promising in identifying mood categories that better reflect users’ actual music information behaviors.

The Challenge
Research on music information behavior has identified music mood or the affective aspects of music as an important criterion in music information seeking and organization (e.g., Cunningham et al., 2004). However, most existing music repositories do not support access to music by mood. In fact, music mood, due to its subjectivity, has been far from well studied in information science. First and foremost, there are no standard music mood categories. Although music psychologists have proposed a number of music mood models over the years, these models were developed in pure laboratory settings and thus mood categories in these models generally lacked the social context of music listening (Juslin and Laukka, 2004). Therefore, it remains a challenge to identify mood categories that reflect the context of real-life music listening. Such categories would better serve users’ needs in organizing and accessing music by mood.

Can Social Tags Help?
With the birth of Web 2.0, the general public can now post text tags on music pieces and share these tags with others. The accumulated user tags can yield so called “collective wisdom” that can augment values of music itself and create the social context of music seeking and listening. Specifically, there are two major advantages of social tags. First, social tags are uncontrolled and thus contain much noise or junk tags. Second, many tags have ambiguous meanings. For example, “love” can be the theme of a song or a user’s attitude towards a song. Third, a majority of tags are tagged to only a few songs, and thus are not representative (so called “long-tail” problem 1). Fourth, some tags are essentially synonyms (e.g., “cheerful” and “joyful”), and thus do not represent separate and distinguishable categories. To address these problems, this paper builds on the initial research in this area by Hu et al. (2007) and proposes a new method for deriving more realistic mood categories from social tags.

Combining Social Tags, Linguistic Resources and Human Expertise
The proposed method combines the strength of social tags, linguistic resources and human expertise. Starting from a large set of social tags on music pieces, the method employs linguistic resources (e.g., domain-oriented lexicons) to filter out junk tags and tags with little or no affective meanings. To solve the synonym problem, the method follows the recommendations from a previous research on music mood metadata by Hu and Downie (2007). Specifically, synonyms are identified with linguistic resources (e.g. thesaurus) and are grouped into the same categories. Each new category is then defined collectively by all terms in it instead of picking one term as a category label. In the next step, the ambiguity problem is untangled by human experts in the music domain with the help of tag co-occurrences. Specifically, to disambiguate tag_a, tags co-occuring with tag_a are identified and presented to human experts. With their music knowledge and the context created by the co-occurring tags, human experts can find out whether tag_a takes an affective meaning in this context. Finally, the unrepresentative “long-tail” is chopped by removing tags that were assigned to few (e.g., less than 10) music pieces.

A dataset used in this preliminary study was about 6,000 songs accessible to the author. 4.5 million social tags on these songs were collected from last.fm, the most popular tagging site for Western music2. Among these tags only 19,627 were unique and were tagged to at least two songs.

The linguistic resource used to solve the uncontrolled vocabulary problem was WordNet-Affect, an affective extension of WordNet (Strapparava and Valitutti, 2004). In WordNet-Affect, affective labels are assigned to concepts (synsets) representing emotions, moods, situations eliciting emotions, or emotional responses. As a major resource in text sentiment analysis, it has a good coverage of mood related words. There were 1,113 mood-related words extracted from WordNet-Affect and 203 of them exactly matched the 19,627 unique tags collected from last.fm.

WordNet is a natural resource for the synonym problem, because it organizes words into synsets. Words in the same

---

1 “long-tail” means the tag distribution follows a power law: many tags are used by a few users while only a few tags are used by many users (Guy and Tonkin, 2006).

synset are synonyms in the linguistic point of view. Hence mood-related tags appearing in the same synset of WordNet-Affect are merged into one group. Moreover, tags with the same morphological root are further merged, by using a stemmer. This is because synsets in WordNet only contain synonyms with the same Part of Speech. For example, “sorrow” and “sorrowful” are in different synsets because one is a noun while the other is an adjective. However, they represent the same kind of mood and should be in the same category. At the end of this step, the 203 mood related tags are merged into 93 groups.

Two human experts were consulted in disambiguating the tags. Both are music information retrieval (MIR) researchers with music background and native English speakers. They firstly removed tags with special music meanings that did not involve an affective aspect, such as “trance” and “beat”. Secondly, since only descriptive terms could be used as category labels, judgmental tags such as “bad”, “good” and “great” were removed. Then, co-occurring tags were obtained from last.fm\textsuperscript{3} to provide more contexts for the human experts. For instance, the tag, “romantic”, mostly refers to a romance-related feeling in a general setting, but its similar tags on last.fm included “symphony”, “czech classical”, “chopin” and “romantic classical”. This indicated that “romantic” in last.fm frequently referred to a time period in the history of classical music, thus was not a mood category candidate. After this disambiguation step, 36 tag groups remained in the experimental dataset.

Several tag groups were further merged if they were deemed musically similar by the experts. For instance, the group of (“panic”, “terror”) was merged with (“scared”, “horror”). An advantage of this approach is to let patterns naturally emerge from the data, but still make them under control of human experts. Finally, after removing the groups of tags applied to less than 10 songs, 25 tag groups left. They represented 25 mood categories each of which was collectively defined by all the tags in the group. Table 1 shows the major categories with their total occurrences in the dataset as well as the number of songs tagged with at least 1 tag in the category.

### Discussion and Future Research

Mood categories have been a much debated topic in MIR. This paper proposes a method that combines the strength of linguistic resources and expert knowledge to identify mood categories from social tags on online music communities. The resultant categories can provide a realistic and user-centric guidance in organizing music and facilitating music access by mood. To further understand the value of the method, the resultant mood categories will be compared to emotion models proposed in music psychology, which will disclose, among other things, whether the categories derived from empirical music listening data reflect theoretical models developed in laboratory settings. Particularly the following questions will be addressed: (1) Is there any correspondence between the resultant categories and the categories in psychological models? (2) How do music pieces distribute across the mood categories? (3) If distances between identified mood categories can be calculated (e.g., by song co-occurrences), how do such distances differ from the relative distances of categories in the psychological models? In general, such comparison will open up new directions in text mining and humanities. By comparing results derived from empirical datasets against what have been proposed in theories of the domain, researchers can refine or adapt theories and models and make them better fit the reality of users’ information behaviors.

<table>
<thead>
<tr>
<th>Groups of mood tags</th>
<th>Occu.</th>
<th>#. songs</th>
</tr>
</thead>
<tbody>
<tr>
<td>sadness, sad, unhappy, melancholic, melancholy</td>
<td>15907</td>
<td>994</td>
</tr>
<tr>
<td>chill, die, fear, frightful, horrible, panic attack,</td>
<td>8959</td>
<td>840</td>
</tr>
<tr>
<td>happiness, content, euphoric, happy, euphoria</td>
<td>5005</td>
<td>528</td>
</tr>
<tr>
<td>blue, dark, depressed, depressing, depressive,</td>
<td>5112</td>
<td>521</td>
</tr>
<tr>
<td>amazing, astonishing, staggering, amazement</td>
<td>1915</td>
<td>370</td>
</tr>
<tr>
<td>angry, anger, choleric</td>
<td>1427</td>
<td>198</td>
</tr>
<tr>
<td>grief, heartache, heartbreak, heartbreaking</td>
<td>1294</td>
<td>123</td>
</tr>
<tr>
<td>aggression, aggressive</td>
<td>782</td>
<td>123</td>
</tr>
<tr>
<td>cheerful, cheery, festive, gala, jolly, joyful,…</td>
<td>1028</td>
<td>117</td>
</tr>
<tr>
<td>fury, madness, rage, mad, sore</td>
<td>506</td>
<td>52</td>
</tr>
<tr>
<td>tenderness, warmth, caring, tender, warm….</td>
<td>291</td>
<td>40</td>
</tr>
<tr>
<td>excitement, exhilaration, exciting, exuberant, …</td>
<td>151</td>
<td>35</td>
</tr>
<tr>
<td>sickness, wicked, sick</td>
<td>135</td>
<td>29</td>
</tr>
<tr>
<td>regret, sorrow, sorrowful, desolate, joyless …</td>
<td>289</td>
<td>29</td>
</tr>
</tbody>
</table>

Table 1: major mood categories derived by social tag analysis

### References


