ABSTRACT
In the era of Web2.0 where users become active contributors instead of merely consumers, Music Information Retrieval (MIR) has turned into a similar new model: MIR2.0. In this new model, users' input is exploited for discovering and resolving a variety of research problems in MIR. This paper reviews representative MIR studies during the last decade that leverage users' input of various types ranging from playlists, social tags to microblogs and points out that the ever-growing user generated data available online can help researchers in both understanding users and exploring better ways to serve users’ music information needs.

1. INTRODUCTION
With the advent of Web 2.0, there is a large and growing amount of user generated data available online. Such data provide first-hand resources for studying and understanding users in daily life settings (a.k.a. the masses). In Music Information Retrieval (MIR), user generated data of various types have been exploited for a number of tasks such as music classification, recommendation, artist network identification etc. In general, the research problems to which user generated data have been applied can be categorized into two groups: 1) to understand users and thus give implications on MIR system and feature design; and 2) to improve MIR task performances. In particular, the second group includes building experimental datasets using users’ input as ground truth labels. This paper reviews seminal work in MIR settings (a.k.a. the masses).

2. USERS’ INPUT IN MIR
In recent years, online music repository and services have strived to elicit users’ input of various types. For example, last.fm is famous for its large amount of users’ social tags on tracks, albums and artists, boasting the largest music tagging site for western music [25]. Epinions.com encourages users to rate and write reviews on music CDs. These user generated data are valuable resources for MIR research. This section reviews and categorizes existing studies according to the types of the users’ input being exploited.

2.1 Playlists
User authored playlists have been used as a source of user perceived music similarity. The assumption is that songs/artists co-occurring in the same playlist would, in general, be more similar than songs/artists selected in random. Berenzweig et al. [4] calculated similarity among artists using a set of 29,000 playlists constructed by real users from the Art of the Mix database (www.artofthemix.org), a repository and community center for playlist hobbyists. Cano and Koppenberger [6] employed the same playlist dataset to construct artist networks.

2.2 Social Tags
In recent years, the increasing number of musical social tags on the Web has stimulated great interest in analyzing and exploiting social tags. A group of studies using musical social tags is on deriving semantic space for music. These studies take advantage of the fact that social tags are from real users and aim to identify semantic spaces in the music domain that can reflect the reality of music listening. For example, both [26] and [18] used social tags on last.fm to create a space of music mood (but using very different methods) and compared the results to classical psychological models respectively. Levy and Sandler [29] analyzed social tags published on last.fm and mystrands.com and concluded that social tags were effective in capturing music similarity.

Another set of research using social tags attempts to link social tags to audio features extracted from music content. Eck and colleagues [14] used last.fm social tags associated with nearly 100,000 artists to evaluate a method of predicting tags from audio input. Since then, the problem of automatically predicting social tags has attracted so much attention that the Music Information Retrieval Evaluation eXchange (MIREX) [12] has held the task of Audio Tag Classification since 2008. In this task, various systems are compared with regard to their abilities of associating 10-second audio clips of music with tags collected from the MajorMiner (http://majorminer.com/) game [31].

Social tags are also used as features in music classification. Bischoff et al. [5] tried to use social tags to predict mood and theme labels of popular songs. Symeonidis et al. [38] strived to make better recommendations based on the latent factors hidden in user-tag-item ternary relations. In [8], last.fm user tags have been used together with content-based features for automatic genre classification.
Other studies applied social tags in building ground truth datasets for experiments on various MIR tasks. Geleijnse et al. [15] investigated last.fm tags of 224 artists and their famous tracks and created a ground truth set of artist similarity based on these tags. Using tags of artist, album and tracks provided by last.fm, Hu et al. [19] derived a set of music mood categories as well as a ground truth track set corresponding to these categories. Building upon previous work, Hu [17] summarized the merits and problems of social tags in building ground truth dataset and constructed a ground truth set on music mood classification from social tags using linguistic resources and human expertise.

2.3 Metadata Labels

Metadata labels can be assigned by experts (e.g., genre categories on AMG (AllMusicGuide.com)) or by everyday users (e.g., usage labels on epinions.com). Both are good resources for investigating newly emerging music metadata (e.g., [21][22]). The former is especially of high quality and thus has been widely used as ground truth for evaluation purposes. For example, Hu and colleagues worked on the mood labels on AMG for deriving a set of 5 mood categories [21]. Berenzweig et al. [4] and Cano and Koppenberger [6] used “similar artists” labels on AMG as ground truth of an artist network.

A few studies utilized metadata labels as features in similarity calculation. For instance, Logan [30] improved the quality of automatically generated playlists by incorporating genre labels collected from AMG. Aucouturier et al. [1] combined acoustic features and Boolean metadata labels to predict high level music properties (e.g., “mood dreamy”, “situation party”) and achieved 5% improvement on precision over systems using audio content alone.

2.4 Web Search Results

There is no doubt that the Web contains millions of pages with music-related information, and under the auspice of Web 2.0, much of the content on the Web are contributed by users. With the popularity of search engines, obtaining music-related Web pages for MIR research has never been easier. In 2002, Whitman and Lawrence [42] proposed an approach to calculating artist similarity using “community metadata” (also called “cultural metadata”) which were words and phrases extracted from Web pages returned by search engines for queries made of artist names. Such “community metadata” took advantage of the Web and soon became popular for various MIR tasks including artist classification [24], music style detection [44], artist recommendation [3], etc. A series of work of Whitman et al. [43][40] then went one step further trying to find the link between audio content and text descriptions extracted from Web search results.

In addition to the textual content of Web search results, co-occurrence page counts returned by search engines also provide useful evidences for solving the puzzles of MIR tasks. For instance, Schedl et al. used co-occurrences of artist names, as returned by search engine’s page counts, to measured artist similarity [35] and categorize artists [36]. Besides, co-occurrence page counts have also been used as a representation of cultural source and combined with audio and symbolic features for the task of genre classification [32].

2.5 Q/A Sites and Forums

Music-related requests posted by users on online forums and question answering (Q/A) websites (e.g., Google Answers, Yahoo! Answers) are essentially music information queries expressed in the form of natural language. These queries are raised by real-world users for real tasks in real-life settings and thus provide a natural and authentic resource for investigating users’ information needs. Besides, posts on Q/A sites or forums usually include answers and/or comments from other users as well, which provides rich contextual details and/or background information for the information seeking process.

Because of these advantages, researchers have been applying users’ input on Q/A sites and online forums to identifying users’ music information needs. The earliest work in this direction can be attributed to [13] where the authors analyzed 161 music information requests posted on rec.music.country.old-time Usenet newsgroup and identified 1) features used to describe music information needs; 2) characteristics of information desired to satisfy the information needs; and 3) intended uses for requested music information. Building upon this work, Bainbridge et al. [2] analyzed 502 postings in the music category of Google Answers and identified the categories of information the posters used to describe their music information needs. Lee et al. [28] then compared music queries posted on a Korean Q/A site, Naver to those on Google Answers and explored the research questions related to cross culture, cross language music information seeking. Later, Cunningham and colleagues [10] analyzed forum postings on The Art of the Mix website (www.artofthemix.org) to investigate what information users needed for constructing playlists and music mix. Most recently, Lee continued her work and conducted a systematic study on Google Answers music queries [27]. The results of the study included a formal taxonomy of user needs, information features used in the queries, as well as recommendations for improving the evaluation of MIR systems.

2.6 Music Reviews

Online customer reviews on music represent a rich resource for examining the ways users describe their music preferences and possible impacts of those preferences. Online reviews can be very detailed, covering not only the reviewers’ personal opinions but also important background and contextual information about the music and artists under discussion. In addition, there is a large amount of review data online as most major online music stores (e.g., amazon.com) provide customer reviews. There are also non-retail websites devoted to customer reviews (e.g., epinions.com). These sources of user-generated information provide researchers with a rich resource for leveraging the collective knowledge of the music-listening public.

Whitman and Ellis [41] were among the first to make use of online music reviews for MIR. They used adjectives and noun phrases extracted from reviews on AMG and Pitchfork Media (pitchfork.com) to learn a model of the audio-to-term relations. Turnbull et al. [39] extracted music related words from editorial music reviews on AMG and trained a multi-class Naive Bayes model to annotate songs with these words as well as to retrieve songs given text queries consisting of these words. Hu et al. studied music reviews for various tasks including music genre classification, rating prediction, semantic polarity classification, authority classification and feature analysis [20].
2.7 Music Blogs

Music blogs are a special type of blogs. They usually contain links to music audio files as well as textual descriptions on the linked music written by the blog authors. This unique resource provides both music audio and natural language text related to the audio, and thus could be very useful for multi-modal MIR approaches that are increasingly adopted in recent years. However, very few MIR studies have used music blogs so far. The study by Celma et al. [7] is one of them. The authors designed an MIR system exploiting music blogs to support retrieving audio files with textual queries. This system can also recommend similar music based on audio content similarity.

Music blogs have been used more often in the field of business regarding to music album sales. For example, Dewan and Ramprasad [34] examined the interrelationship between music blog buzz and album sales. Their findings included that blog buzz had a positive and significant relationship with album sales and that this relationship was stronger for independently released music.

2.8 Microblogs

Microblogging services have surged into popularity within the last three years, with Twitter being the most notable service. Just as its name indicates, microblogs are like traditional blogs but in much smaller size. Music-related microblogs may demonstrate different characteristics from other social media such as social tags. However, to date very little research has done on music-related microblogs. Schedl’s study [34] is among the first works, if not the very first one, on leveraging microblogging posts for MIR tasks. Specifically, the author downloaded microblog posts on Twitter that included any artist name in a set of 224 artists and used the posts to calculate similarity among the artists. The results showed that the microblog data achieved comparable results to other textual data such as Web pages in the task of artist similarity estimation. It also compared the top N terms related to an artist other textual data such as Web pages in the task of artist similarity estimation. It also compared the top N terms related to an artist

2.9 Transaction Logs

Transaction logs are among the first user generated data being exploited in MIR studies. Strictly speaking, transaction logs are not part of Web 2.0 as they are usage data instead of content data and they are generated implicitly without users’ intention of contribution. Nevertheless, they are generated by users and can reveal (part) of users’ behaviors in interacting with music information systems.

Transaction logs are in large quantity and in great details – virtually all users’ interactions with the system can be recorded in the logs, and thus they provide a quantitative means to study users’ behaviors. McPherson and Bainbridge [33] analyzed server logs of the MELDEX digital music library of one year period of time and discovered usage patterns of its users. They found that, although alternative query methods were available, nearly half of the user-issued queries were textual ones. Another study [23] aimed to identify access points of music scores in online environment. The author surveyed 21,177 online catalog search logs in an academic music library. The findings included that more than half of the searches combined multiple access points. Such findings provide very useful implications and suggestions for designing MIR systems.

Besides user studies, transaction logs or server logs in general have been also used in collaborative filtering, a technique that makes recommendations to a user according to the tastes of other similar users. Cohen and Fan [9] did an early work on applying server logs to the problem of music recommendation. Using server logs associated with a large (2,800 album) repository of digital music, they calculated similarities among users according to the information of which songs were downloaded by which IP addresses. A variant of this study is [16] which used logs of requests to an Internet radio station to calculate user similarity. In fact, the recommendation service on last.fm is using the same idea – tracking which songs each user has been listening to.

The shortcoming of transaction logs is that they cannot tell the intention or satisfaction levels of users, nor can they provide contextual or background information as provided by users’ natural language queries or descriptions.

2.10 P2P File Sharing Services

Peer-to-peer (P2P) file sharing services allow users to share their resources with other users without a central server. P2P music-sharing services such as OpenNap and Gnutella have been very much welcomed among music listeners. The records of file sharing behaviors provide another resource for MIR researchers to investigate how users share music with one another. Shavitt and Weinsberg [37] collected the metadata of 530,000 songs from more than 1.2 million shared music files on Gnutella and discovered patterns music sharing behaviors including that most users in the P2P network shared similar music files.

Since users with similar musical taste often share files with one another, data about P2P music sharing can be applied to music recommendation using collaborative filtering. On the other hand, similar songs are often shared together by many different users and thus music sharing data can also be employed in calculating song similarity. Shavitt and Weinsberg [37] also used the Gnutella data for artist recommendation and song clustering, giving special emphasis to adjusting for the popularity bias.

Besides, data on P2P file sharing service can also provide information on individual user’s music collections, which can be used to calculate artist/song similarity. The assumption is: artists/songs co-occurring in someone’s collection have a better-than-average chance of being similar. Berenzweig et al. [4] retrieved about 3,200 user collection data from OpenNap and calculated artist similarity from about 176,000 user-to-artist relations contained in this dataset.

Despite of its usefulness for MIR research, P2P music sharing services are challenged by copyright issues. Many of OpenNap’s biggest networks have been denied by the Recording Industry Association of America (RIAA), and can only be brought back by complying with the RIAA legal terms.

3. CONCLUSIONS

With the advent and popularity of Web2.0, MIR also entered an era of MIR2.0 where user generated data are important resources for MIR research on both investigating users’ behaviors and
harnessing users’ input to improve information services. By reviewing MIR studies leveraging user’s input of 10 different types, this paper aims to provide insights and encouragement for further work on MIR2.0 exploiting user generated data.

4. REFERENCES


[34] Schedl, M. 2010. On the use of microblogging posts for similarity estimation and artist labeling. Accepted by of the 11th ISMIR.


