A Method for Obtaining Semantic Facets of Music Tags

Mohamed Sordo
Universitat Pompeu Fabra
Barcelona, Spain
mohamed.sordo@upf.edu

Fabien Gouyon
INESC Porto
Porto, Portugal
fgouyon@inescporto.pt

Luís Sarmento
LIACC/FEUP, Univ. do Porto
Porto, Portugal
las@fe.up.pt

ABSTRACT

Music folksonomies have an inherent loose and open semantics, which hampers their use in structured browsing and recommendation. In this paper, we present a method for automatically obtaining a set of semantic facets underlying a folksonomy of music tags. The semantic facets are anchored upon the structure of the dynamic repository of universal knowledge Wikipedia. We illustrate the relevance of the obtained facets for the description of tags.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing—*Dictionaries, Linguistic processing*; H.5.5 [**Information Storage and Retrieval**]: Sound and Music Computing

General Terms

Algorithms, Experimentation, Languages

Keywords

Music tagging, Last.fm, Wikipedia, Social music

1. INTRODUCTION

Music is a complex phenomenon that can be described according to multiple *facets*. Descriptive facets of music are commonly defined by experts (e.g. stakeholders in the music industry) in professional taxonomies. Multifaceted descriptions are especially useful for music browsing and recommendation. For instance, recommendations of the Pandora Internet radio use around 400 music attributes grouped in 20 facets,[1] as for instance Roots (e.g. "Afro-Latin Roots"), Instrumentation (e.g. "Mixed Acoustic and Electric Instrumentation"), Recording techniques (e.g. "Vinyl Ambience"), or Influences (e.g. "Brazilian Influences").

[1] `http://en.wikipedia.org/wiki/List_of_Music_Genome_Project_attributes`

WOMRAD 2010 Workshop on Music Recommendation and Discovery, colocated with ACM RecSys 2010 (Barcelona, SPAIN)

However, there exists no consensual taxonomy for music. Previous research showed the music industry uses *inconsistent* taxonomies [6], even when restricting to a single and widespread facet such as the music genre. Also, expert-defined taxonomies (music-related or not) have two fundamental problems. First, they are very likely to be *incomplete*, since it is impossible for a small group of experts to incorporate in a single structure all the knowledge that is relevant to a specific domain. Second, since domains are constantly evolving taxonomies tend to become quickly *outdated* –in music, new genres and techniques are constantly emerging.

An alternative strategy for describing music consists in relying on the broadness of the web and making use of the "wisdom of the crowds". Many music websites allow users themselves to assign their own descriptive tags to music items (artists, albums, songs, playlists, etc.). For instance, users of the website Last.fm tagged the band Radiohead as "90s", "00s", "alternative", "post-punk", "britpop", "best band ever", among other things. The combination of annotations provided by thousands of music users leads to the emergence of a large body of domain-specific knowledge, usually called *folksonomy*. Due to its informal syntax (i.e. direct assignment of tags), the tagging process allows the collective creation of very rich tag descriptions of individual music items.

When compared to taxonomies defined by experts, music folksonomies have several advantages. First, completeness, they ideally encompass all possible "ways to talk about music", including both *lay* and *expert* points of view. Second, due to the continuous nature of the tagging process, folksonomies tend to be well updated. Third, they usually incorporate both *commonly accepted* and *generic* concepts, as well as *very specific* and *local* ones.

It seems reasonable to assume that folksonomies tend to encompass various groups of tags that should reflect the underlying semantic facets of the domain including not only traditional dimensions (e.g. instrumentation), but also more subjective ones (e.g. mood). However, the simplicity and user-friendliness of community-based tagging imposes a toll: there is usually no way to *explicitly* relate tags with the corresponding music facets. For instance, a user may assign a number of tags related with music genre without ever actually explicitly specifying that they are about "music genre". For providing a flexible browsing experience, this is a significant disadvantage of folksonomy-based classification in relation to classification based on taxonomies, where the information about which facets are being browsed can be made explicitly available to the user.

In this paper, we approach an essential research question that is relevant to bridging this gap: Is it possible to *automatically* infer the semantic facets inherent to a given music folksonomy? A related research question is whether it is then possible to classify elements of that music folksonomy with respect to the inferred semantic facets?

We propose an automatic method for (1) uncovering the set of semantic facets implicit to the tags of a given music folksonomy, and (2) classify tags with respect to these facets. We anchor semantic facets on metadata of the semi-structured repository of general knowledge Wikipedia. Our rationale is that as it is dynamically maintained by a large community, Wikipedia should contain *grounded* and *updated* information about relevant facets of music, in practice.

2. RELATED WORK

Music tags have recently been the object of increasing attention by the research community [3, 4]. A number of approaches have been proposed to associate tags to music items (e.g. a particular artist, or a music piece) based on an analysis of audio data [1, 9], on the knowledge about tag co-occurence [5], or on the extraction of tag information from community-edited resources [8]. However, in most cases, such approaches consider tags independently, i.e. not as elements in structured hierarchies of different music facets. When hierarchies of facets are considered, they are usually defined *a priori*, and greatly vary according to authors. For example, [4] groups tags in the following facets: genre, locale, mood, opinion, instrumentation, style, time period, recording label, organizational, and social signaling.

To our knowledge, however, few efforts have been dedicated to the particular task of *automatically* identifying the relevant facets of music tags. In their work on inferring models for genre and artist classification, Levy et al. apply dimensionality reduction techniques to a data set of tagged music tracks in order to obtain their corresponding compact representations in a low-dimensional space [5]. They base their approach on tag co-occurrence information. Some emerging dimensions can be associated to facets such as Era (e.g. the dimension [90s]). However, most of the dimensions thus inferred are, in fact, a combination of diverse music facets, such as for example the dimension [guitar; rock], which includes concepts of instrumentation and of genre.

Cano et al. use the WordNet ontology to automatically describe sound effects [2]. Albeit the very large amount of concepts in WordNet, they report that it accounts for relatively few concepts related to sound and music, and propose an extension specific to the domain of sound effects. On the one hand, they illustrate that browsing can indeed be greatly enhanced by providing multifaceted descriptions of items. On the other hand however, it is our belief that, because of their necessary stability, existing ontologies are not the most adapted tool to describe domains of knowledge with inherent open and dynamic semantics, such as music.

3. METHOD

Our method consists in using metadata from Wikipedia to infer the semantic facets of a given music folksonomy. This is performed in two steps. In the first step, we specialize the very large network of interlinked Wikipedia pages to the specific domain of the music folksonomy at hand. This is done by maximizing the overlap between Wikipedia pages and a list of frequent tags from the folksonomy. As the resulting network still represents a very large number of nodes, in a second step, we focus on the most relevant ones (node relevance being defined as an intrinsic property of the network). This step also includes additional refinements.

3.1 Obtaining a Music-Related Network

Wikipedia pages are usually interlinked, and we use the links between two particular types of pages (i.e. *articles* and *categories*) to construct a music-related network. Concretely, we use the DBpedia knowledge base (`http://dbpedia.org/`) that provides structured, machine-readable descriptions of the links between Wikipedia pages (DBpedia uses the SKOS vocabulary, in its 2005 version).[2] In particular, we make use of two properties that connect pages in DBpedia: (1) the property *subjectOf*, that connect articles to categories (e.g. the article "Samba" is a *subjectOf* of the category "Dance_music", and (2), the property *broaderOf*, that connect categories in a hierarchical manner (e.g. the category "Dance" is a *broaderOf* of the category "Dance_music", which is a *broaderOf* of the category "Ballroom_dance_music").

We start from the seed category "Music" and explore its neighbourhood from the top down, checking whether connected categories can be considered relevant to the music domain. A category is considered relevant if it satisfies any of the two following conditions:

- It is a tag from the folksonomy, such as for example "Rock and Roll". (This condition will be referred to as *isMusical*);
- At least one of its "descendants" is a tag from the folksonomy *and* the substring "music" is included in the title or the abstract of the corresponding Wikipedia article. (This condition is further referred to as *isTextMusical*.)

The "descendants" of a category are fetched from DBpedia using the two connecting properties previously described. These descendants can be either "successors" (i.e. all direct *subjectOf* and *broaderOf* of this category), or successors of successors, and so on. This iterative search is limited by a maximum depth, empirically fixed to a value of 4. Indeed, experiments with smaller values yielded a significant reduction of the tag coverage, while experiments with greater values did not increase significantly the coverage.

If any of the previous conditions is satisfied, the category, its successors and their edges are added to the network. Otherwise, the category and all incident edges are removed. The algorithm proceeds iteratively (following a Breadth-First search approach) until no more categories can be visited. A summarized version of the method for obtaining a music-related network is described in algorithm 1.

3.2 Finding Relevant Facets

Once the network of music-related categories is built, the next step is to find the nodes that are potentially more relevant to the network than others.

We invert the direction of the edges of the network in order to point back in the direction of the most generic category, i.e. "Music", and we compute the PageRank of the

[2] `http://www.w3.org/TR/2005/WD-swbp-skos-core-spec-20051102/`

Data: $C = \emptyset$, a list of categories (a queue, initially empty); $N = (V, E)$, a directed network with a set of nodes V and a set of edges E (initially empty);
Result: N, network with music nodes;
$C \leftarrow C \cup$ "$Music$";
while $C \neq \emptyset$ **do**
  $c \leftarrow first\ element\ of\ C$;
  $C \leftarrow C - c$;
  if *(c isMusical)* $\vee$ *((at least one descendant of c isMusical)* $\wedge$ *(c isTextMusical))* **then**
    $N \begin{cases} V \leftarrow V \cup c \cup successors(c) \\ E \leftarrow E \cup edges\ between\ c\ and\ successors(c) \end{cases}$
    $C \leftarrow C \cup successors(c)$
  else
    $N \begin{cases} V \leftarrow V - c \\ E \leftarrow E - all\ edges\ incident\ in\ c \end{cases}$
  end
end

Algorithm 1: Pseudo-code for the creation of a network of music-related categories from Wikipedia.

resulting network. PageRank [7] is a link analysis algorithm that measures the relative relevance of all nodes in a network. In PageRank, each node is able to issue a relevance vote on all nodes to which it points to (thus the need for reorienting the edges). The weight of the vote depends on the relevance of the voting node (i.e. relevant nodes issue more authoritative votes). The process runs iteratively, and (under certain conditions) converges to a stable relative ranking, where nodes to which more edges from other relevant nodes converge (directly or indirectly) are considered more relevant. For initializing the PageRank algorithm, we set the initial weight of each node to 0.

In order to capture general yet complementary facets of music, we aim at reducing semantic overlap as much as possible by applying the following filters:

Stub Filter: We remove all categories with substring "_by_" and "_from_". We noticed that many categories in Wikipedia are actually combinations of two more general categories, as for instance "Musicians_by_genres", which is halfways between "Musicians" and "Music_genres" (see also figure 1). Further, we also remove categories that include "_music(al)_groups" (e.g. "Musical_groups_from_California" that has hundreds of connected categories, hence a high PageRank). Most of these categories are used as *stubs*, even sometimes explicitly so we also excluded categories with the word "stub".

Over-Specialization Filter: We exclude all categories that include lexically a more relevant category. Many relevant categories are *specializations* of other more relevant ones, this occurs mostly with concepts related to anglophone music, which are described in great detail in Wikipedia (e.g. "American_Musicians" includes "Musicians" that has a higher PageRank).

Tag Filter: We remove all categories that are tags. Our objective is to uncover music facets that are implicit to the tags that make up a folksonomy. In general, tags are *elements* of such facets, not the facets themselves.

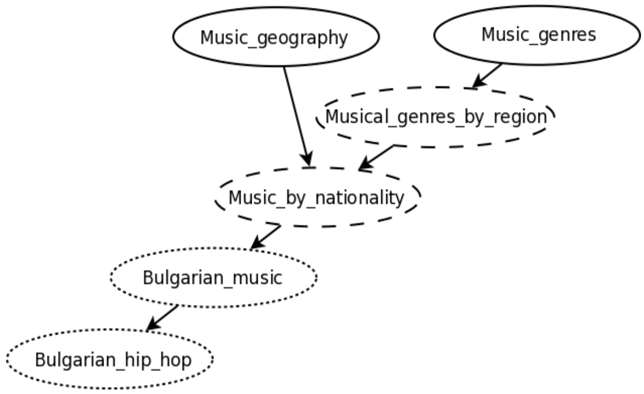

Figure 1: Example of subnetwork in our data. Dotted lines correspond to Wikipedia categories that are also Last.fm tags. Dashed lines correspond to categories not kept. Plain lines correspond to facets kept.

4. RESULTS

We experimented our method on a large dataset of artist tags, gathered from Last.fm during April 2010. The dataset consists of around 600,000 artists and 416,159 distinct tags. This dataset was cleaned in order to remove noisy/irrelevant data: (1) tags were edited in order to remove special characters such as spaces, etc.; (2) tags were filtered by weight[3], only tags with a weight ≥ 1 were kept; and (3) tags were filtered by popularity, keeping only tags with popularity ≥ 10, i.e. keeping only tags that were assigned to at least 10 artists. As a result, the final dataset consists of 582,502 artists, 39,953 distinct tags, and 9.03 tags per artist.

After running both stages of our method, we obtained a list of 333 candidate facets. Table 1 contains the top-50 facets, ordered by pagerank (top to bottom, left to right).

Table 1: Top-50 Wikipedia music facets

Music_genres	Aspects_of_music
Music_geography	Hip_hop_genres
Musical_groups	Music_of_California
Music_industry	Music_theory
Musicians	Rock_and_Roll_Hall_of_Fame_inductees
Musical_culture	Musical_subcultures
Occupations_in_music	Recorded_music
Music_people	Musical_quartets
Record_labels	Music_festivals
Music_technology	East_Asian_music
Sociological_genres_of_music	Centuries_in_music
Music_publishing_companies	Musical_composition
Musical_instruments	Musical_quintets
Anglophone_music	Southern_European_music
Music_of_United_States_subdivisions	Music_software
Western_European_music	Incidental_music
American_styles_of_music	Years_in_music
Radio_formats	Music_websites
Music_publishing	Guitars
Albums	Music_competitions
Musical_techniques	Musical_eras
Wiki_music	Music_and_video
Music_history	Musical_terminology
Music_performance	Music_halls_of_fame
Music_publishers_"people"	Dates_in_music

4.1 Assigning facets to tags

In order to assign a set of facets to a given Last.fm tag, we process the subnetwork of Wikipedia pages specialized to the Last.fm folksonomy (obtained in section 3.1), as described in algorithm 2 (Note that this process is restricted to tags that can be matched to one of the nodes in the network).

[3] i.e. Last.fm "relevance weight", which goes from 0 to 100

Table 2: Sample of the top tags for various music facets inferred

Music_genres	Occupations_in_music	Musical_instruments	Aspects_of_music
Sufi_music	Troubadour	Melodica	Rhythm
Dance_music	Bandleaders	Tambourine	Melody
Indietronica	Pianist	Drums	Harmony
Minimalism	Singer-songwriter	Synthesizers	Percussion
Singer-songwriter	Flautist	Piano	Chords
Music_software	**Music_websites**	**Music_competitions**	**Musical_eras**
Nanoloop	Mikseri.net	Nashville_Star	Baroque_music
Scorewriter	PureVolume	American_Idol	Ancient_music
MIDI	Allmusic	Melodifestivalen	Romantic_music
DrumCore	Jamendo	Star_Search	Medieval_music
Renoise	Netlabels	Eurovision_Song_Contest	Renaissance_music

Data: $C = \emptyset$, a list of categories (initially empty); F, a list of top-N music facets; t, a Last.fm tag;
Result: TF, list of facets applied to tag t;
$iter \leftarrow 1$;
$TF = \emptyset$;
while $(F \neq \emptyset) \vee (iter \leq maxIter)$ **do**
    $C \leftarrow C \cup predecessors(t)$;
    if $(\exists f \in (F \cap C))$ **then**
        $TF \leftarrow TF \cup f$
        $F \leftarrow F - f$
    end
    $iter \leftarrow iter + 1$
end

Algorithm 2: Pseudo-code for assigning Wikipedia facets to Last.fm tags

Given a Last.fm tag t, we look at its "predecessor" categories c, or more formally:

$$predecessors(t) = \{c | (t\ broaderOf(c)) \vee (t\ subjectOf(c))\}.$$

If any of these predecessors is a top-N facet, it is then assigned to t. The process continues iteratively until no more facets can be assigned to the tag, or a maximum number of iteration ($maxIter$) is exceeded. We empirically set this value to 8. This condition can be interpreted as the maximum distance in the network between a tag and a facet.

Table 2 presents a small subset of the obtained facets, followed by a subset of their corresponding list of top tags. Top tags are chosen based on the distance (in number of successive edges in the music network) to the given facet.

The relevance R_{tf} of a music facet f to a tag t is computed as the normalized inverse distance d_{tf} –in number of successive edges– between t and f:

$$R_{tf} = \frac{\frac{1}{d_{tf}}}{\sum_i \frac{1}{d_{ti}}}$$

For example, in figure 1, given the tag *bulgarian hip-hop*, our method starts navigating through the predecessors of this tag until finally reaching two music facets: *Music_genres* and *Music_geography*:

```
bulgarian hip-hop: {(Music_genres, 0.4),
                    (Music_geography, 0.6)}
```

5. SUMMARY AND FUTURE WORK

Although potentially more complete and up-to-date than taxonomies, music folksonomies lack structured categories, a particularly relevant aspect to browsing and recommendation. In this paper, we addressed the problem of uncovering the underlying semantic facets of the Last.fm folksonomy, using Wikipedia as backbone for semi-structured semantic categories.

There are many avenues for future work. First and foremost, we intend to evaluate the relevance of the obtained facets via systematic evaluations of tag classification. We will also study the distributions of music facets with respect to artist popularity. Further work should also relate to evaluating the usefulness of the obtained facets in a number of tasks, such as music recommendation, or tag expansion. We also intend to release the data (and code used to obtain it) in order to stimulate its use by fellow researchers.

6. ACKNOWLEDGMENTS

Thanks to Òscar Celma (BMAT), Eduarda Mendes Rodrigues (FEUP) and anonymous reviewers for useful comments. This work was partly supported by the *Ministerio de Educación* in Spain, and the *Fundação para a Ciência e a Tecnologia* (FCT) and QREN-AdI grant for the project Palco3.0/3121 in Portugal.

7. REFERENCES

[1] T. Bertin-Mahieux, D. Eck, F. Maillet, and P. Lamere. Autotagger: A model for predicting social tags from acoustic features on large music databases. *JNMR*, 37(2):115–135, 2008.

[2] P. Cano, M. Koppenberger, P. Herrera, O. Celma, and V. Tarasov. Sound effect taxonomy management in production environments. In *AES*, 2004.

[3] O. Celma. *Music Recommendation and Discovery - The Long Tail, Long Fail, and Long Play in the Digital Music Space*. Springer, 2010.

[4] P. Lamere. Social tagging and Music Information Retrieval. *JNMR*, 37(2):101–114, 2008.

[5] M. Levy and M. Sandler. Learning latent semantic models for music from social tags. *JNMR*, 37(2):137–150, 2008.

[6] F. Pachet and D. Cazaly. A taxonomy of musical genres. In *RIAO*, 2000.

[7] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.

[8] L. Sarmento, F. Gouyon, and E. Oliveira. Music artist tag propagation with wikipedia abstracts. In *ECIR-WIRSN*, 2009.

[9] D. Turnbull, L. Barrington, D. Torres, and G. Lanckriet. Semantic annotation and retrieval of music and sound effects. *IEEE TASLP*, 2(16):467–476, 2008.