MOOD-EX-MACHINA: TOWARDS AUTOMATION OF MOODY TUNES

Sten Govaerts, Nik Corthaut, Erik Duval
Katholieke Universiteit Leuven
Department of Computer Science
Celestijnenlaan 200A
B-3001 Heverlee, Belgium
{Nik.Corthaut, Sten.Govaerts, Erik.Duval}@cs.kuleuven.be

ABSTRACT

In 2006, the rockanango system was developed for music annotation by music experts. The system allows these experts to create new musical parameters within a flat data structure [1]. Rockanango is deployed in a commercial environment of hotels, restaurants and cafés. One of the main concerns is the time it takes to manually annotate the music and to introduce new parameters. In this paper, we investigate the possibilities to assist the experts by means of automatic metadata generation. Two case studies are described. One focuses on the use of association rules, in combination with lower level metadata like mode and key. The other case study concerns the generation of a topic or subject marker for songs through harvested lyrics and a keyword generator. From our evaluation, we conclude that the generated keywords are relevant and that the music experts value them higher then laymen. Data mining techniques provide means for monitoring the metadata in terms of interparametric relationships that can be used to generate metadata.

1. INTRODUCTION

Last year, we proposed a platform for music annotation, the rockanango project [1]. This project is a co-operation between the Katholieke Universiteit Leuven and the company Aristo Music (http://www.aristomusic.com). In this project, we developed a context-based music player for the horeca (hotels, restaurants and cafés). The main idea behind this music player is a musical context. This is a description of a situation, based on atmospheres and musical parameters. The bartender or restaurant owner can describe a desired musical context by selecting different atmospheres and musical parameters, like genre, instrument, dance style, rhythm... With this description, a matching playlist is generated. This system is commercialized and currently running at around 400 places in the Benelux and France. The music collection consists of around 30,000 songs.

The system relies heavily on good quality metadata. Each song is manually annotated for each of the 22 musical parameters by a team of music experts (dance teachers, music teachers, DJ's...) with a tool [1] that we created for this purpose. The two main problems are the quality and cost of annotating.

The cost of classifying wrongly is considerable: imagine playing a mood-killing song on a Saturday night party and the effect on the dance floor... To counter this, the metadata is annotated in multiple iterations. Furthermore, the indexers look at each other's work in order to refine the metadata. Regarding quality as fitness for purpose, our purpose is making the acquired experience of music experts in creating moods while rendering music available to pub owners. For that reason we cannot discard the music experts.

When a new song enters the system, it has only trivial metadata. Therefore, the playlist generator will not select it. It can take days if not weeks to actually reach a full annotation. Not only is this process slow, it is also expensive, as it involved manual labour by the indexers.

To assist in this task, different complementary approaches are possible. The metadata can be computationally generated through various artificial intelligence techniques utilising our annotated database as ground truth. The vast knowledge pool of the Internet can be mined and automated classifiers used on the mined results [2].

Studies [14] show that many have an interest in music. Their knowledge can be used to assist our music experts. Another approach is collaborative filtering, as used by for instance Last.fm (http://www.last.fm). Using human algorithms [3] could be another possibility. This is ideal for metadata that is typically hard to compute.

We want to bundle the power of computational techniques, the knowledge of the masses and the experience of our music experts. The data from the computational and collaborative approach can be combined and used to suggest values for atmospheres and musical parameters during manual annotation to the experts, while they are annotating music. The music experts can then decide based on their knowledge and context-awareness.

2. TOOLS OF THE TRADE

We will look at two different paths we have taken: first we will look at an approach to assist with the annotation of the topic of a song, then we will look at the extraction of key, mode and chords, and last we derive association rules out of the annotated music.

2.1. With lyrics towards topic markers

The musical experts planned the introduction of a new data field in the metadata scheme, a field to mark the subject of a song. This should allow people to select songs based on their spoken content, for example "love songs" or "songs about a dog". To introduce this new data field manually for 30,000 songs would take much time and thus be very costly.

Plenty of tag-based systems are available, like Last.fm, AllMusic (http://www.allmusic.com), Plurn (http://www.plurn.com). For example, here is a sample of the tags of "Paint it Black" of the Rolling Stones on Last.fm: "60s", "favourites", "guitar", "psychedelic rock", "rock", "seen live". The problem is that tags mostly indicate genres, sometimes moods and favourites. Even if there would be subjects in the tag cloud of a song, it would be hard to select the valid subject candidates from genres, favourites and moods.

It is clear that having the lyrics can assist with this task. This transforms the retrieval of topic from an acoustic problem to a text-retrieval problem. Adding lyrics to our database isn't new and can assist in solving another problem. When people search for a song, they often know parts of recurring words in the song and not the title. It is not uncommon that the title doesn't occur in the lyrics at all. If this is the case the user needs to know the title song to find it, e.g. Queen's 'Bohemian Rhapsody'. Enriching our search engine with lyrics would solve this. Whether to search on the complete text or on the most important parts of the lyrics, for example the intro and chorus, to limit the search results, should be evaluated. But we think that there definitely is an advantage to be able to search on 'Galileo' or 'Scaramouche' and find 'Bohemian Rhapsody'.

The Internet contains many sites dedicated to song texts, mostly entered by fans. In order to harvest the lyrics, there are in general two ways: through an API or by web crawling and screen scraping. We decided to use a source that supports the first approach.

LyricWiki (http://www.lyricwiki.org) is a free source to search for or add lyrics and it has a web services API available.

2.1.1. Lyrics Keyword Extraction

The importance of a keyword in a given lyric makes a reasonable metric for the quality of a generated topic. It is very hard to extract information about the context the song was written for. Our approach to suggesting a topic is using keyword extraction on the lyrics.

Simple keyword extraction is based on term frequency while complex approaches rely on statistical techniques [7] or natural language processing [8], sometimes supported by domain specific ontologies [9]. There are plenty of keyword extraction techniques in IR literature, most are either proprietary or experimental; hence there are not much freely available products that can be used. The limited options are search engine optimisation (SEO) keyword analyser tools, Kea [10], or the Yahoo API term extractor [11].

Kea requires extensive training in a specific domain to generate reasonable results. SEO tools mostly look for popular search terms in a webpage while extracting keywords and the techniques used are very basic, e.g. word frequency/count. Yahoo uses a context-based technique to extract keywords [12]; this means that they can generate results based on the context of a document, so no training is needed. Yahoo can generate multi-word keywords by recognizing, with NLP, parts of sentences. The Yahoo term extractor seems most suited for lyrics.

2.1.2. Evaluation

We evaluated whether the generated keywords are good topic markers for a song. This is done by measuring the difference in perception between the music experts, who will be assisted by the keywords during their work, and ordinary people. As a significant sample set, 5 out of 7 music experts and 10 laymen performed a survey on 10 songs out of a pool of 40. The laymen are people working on non-music related research at the department of Computer Science at the

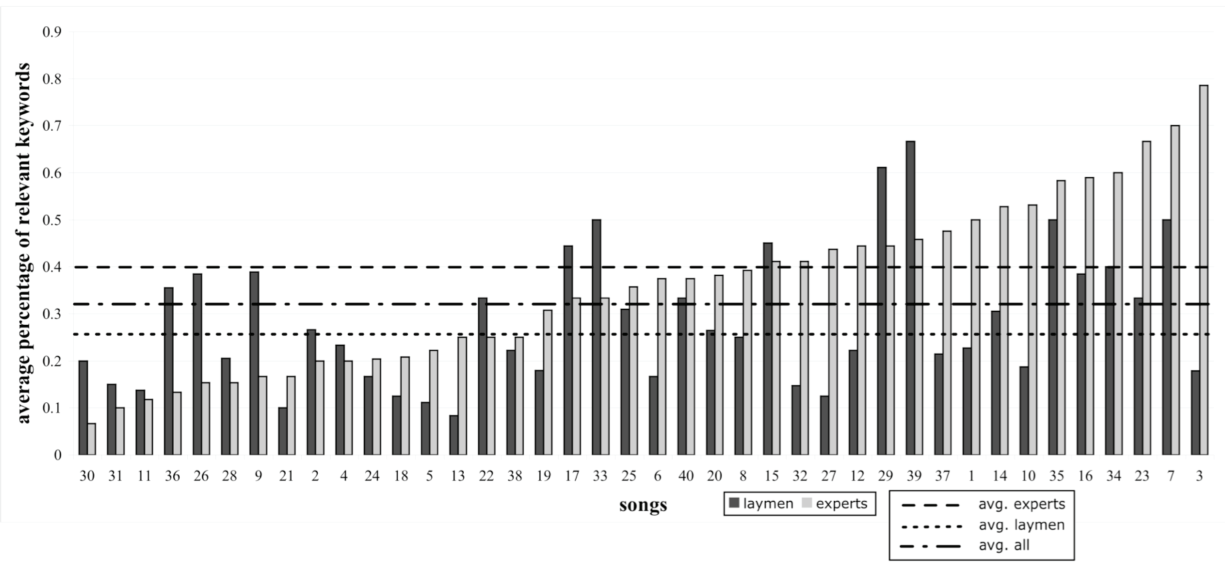

Figure 1. Bar chart with average percentage of relevant keywords of laymen and experts

K.U. Leuven. We made two musical contexts, one with very popular songs and one with unpopular songs, harvested the lyrics and generated 474 keywords for the 40 songs, selected evenly amongst the two contexts. For each song we made a survey, containing the artist and title of the song, the lyrics, the keywords and some questions. We asked whether the evaluator knew the song, whether a given keyword was relevant for the song and their name. We distributed the surveys randomly over laymen, assuring that every song was at least evaluated twice and for the experts at least once. This gives 150 samples of keyword validation data, this is sufficient for statistical relevance.

Figure 1 shows a chart with the (ordered) average percentage of relevant keywords for each song for the experts and the laymen. We observe that there were relevant topic markers for all of the songs. Further more, experts value the keywords higher – on average.

This can be formalised by studying the mean values of the experts (40.3%) and the laymen (26.9%). We performed a two-sample t-test with equal variances not assumed (see Table 1), since each song was considered multiple times. This gives us a t-value of 3.8 with significance 1.14e-4. The large difference of 13.4% in mean between the laymen and the experts is supported by a 95% significance interval of [0.06,0.21]. Lyrics have a special textual structure, separated in verses and chorus. Choruses are likely to contain good pointers to the topic of the song. It is possible that the retrieved lyrics contain a chorus that gets repeated, or that marginally changes over time. Due to the high similarity of these words, they could be considered as stop words by the algorithm. There has been some research on finding complete lyrics [13], but a lyrics aware keyword extractor would be needed too.

Musical exclamations, e.g. 'Yeah, baby', 'Oh, no, no, no', 'Na, na, na, na', also showed up between the keywords. For this reason it is not surprisingly that about 2/3 of the keywords are irrelevant. Some participants asked us why we did not treat them as stop-words. We are considering doing so in our future work.

average # keywords	11.85 keywords		
		laymen	all
average # relevant keywords	4.44	3.03	3.61
std. dev. avg. # relevant keyw.	3.29	2.01	2.69
average % relevant keywords	40.3%	26.9%	32.3%
std. dev. avg. % relevant keyw.	25.9%	16.1%	21.6%
tests			
t-value	3.82	significance	1.14E-4
95%- confidence interval between expert-laymen means	[0.06, 0.21]		

Table 1: statistical results of keyword evaluation.

2.2. Data mining and association rules

The systematic approach of the experts posed the question whether association rules could be retrieved from the existing metadata. To test this hypothesis, we used the a priori association rule algorithm from the WEKA [5] toolkit on the data. We made a selection of subjective parameters: genre, subgenre, mood, dancing style and party factor. These parameters are the most time consuming for the music annotators and we know from experience that these parameters are key in generating musical contexts.

2.2.1. Observations

Not much to our surprise, we could identify the genre given the main subgenre with remarkable confidence level: 100%. More interesting however is the lack of association rules containing mood, dancing style or rhythm style. Table 2 shows the derived rules with confidence level above 75%. We used the a priori algorithm for its memory efficiency due to the large dataset (30000 songs). Remarkable, but not surprising is the association of a low danceability with a low party factor (86%).

If we limit to pop songs according to the first generation metadata scheme (3.2.2), other parameters appear on the right hand side of the rules. Subgenre, dancing style and rhythm style with respective values Pure Pop, Pop Rhythm 4/4 and Slow interact with one another and with danceability and party factor. These rules make actually sense: Music that is not danceable will most likely not drive the dance floor wild and hence will not make a good party song.

Subgenre = Ambiance Café	1437	→	Genre = Schlager	1437	100%
Subgenre = Schlager Café	1359	→	Genre = Schlager	1359	100%
Subgenre = Pure Pop	1226	→	Genre = Pop	1220	100%
Dance Style = Slow	1808	→	Party Factor = 1	1731	96%
Genre = Rock and Danceability = 1	2006	→	Party Factor = 1	1762	88%
Danceability = 1	3463	→	Party Factor = 1	2990	86%

Table 2. Rules with a priori, confidence level > 75%. The grey boxes show the number of songs, the black ones the confidence level of the rules.

Clearly, subjective parameters alone do not have sufficient predicting power for parameter automation. In an attempt to counter this problem, we introduced 2 additional parameters: the key of the song and what we call the mode of a song. These parameters are generated from the chords generated by the ChordsExtractor [15] embedded in the CLAM Annotator framework (http://clam.iua.upf.edu/). Histograms were used to determine the main key. The mode of a song is determined by counting the occurrences of minor and major chords. The value for the mode is either minor or major. It was expected that those parameters are relevant for the mood related parameters on the idea of chord quality [6]. Unfortunately, they are by far not expressive enough to predict any subjective parameter by means of a classifier for instance. To accomplish this more predictive parameters are needed.

2.2.2. *Metadata scheme changes*

The rockanango metadata annotation methodology [1] allows for changes in the metadata scheme. The musical experts claim to have found new insights in the classification of mood and other musical parameters like rhythm style. Additionally, they suggest new parameters that can be used for association rules.

Instead of using flat labels for music annotation, the experts have developed a tendency towards categories with a limited set of related values. This allows for a more granular approach. The values for the mood parameter are now grouped in four categories. The values in category are ordered, allowing for a more convenient comparison.

The reason behind the granularity is the gap in musical knowledge between experts and (most) end users. The typical bartender cannot distinguish a cumbia from a rumba; he just wants 'this Latin flavour'. A Latin dance teacher, however, needs the difference. Additionally, it is expected that classifiers will perform better with bigger grained metadata because of reduced value spaces for algorithms.

3. CONCLUSION AND FUTURE WORK

In conclusion of the two presented case studies we can state that metadata generation by means of MIR techniques or collaborative filtering is a valid option in the process of contextualising music. Furthermore, data mining techniques provide means for monitoring the metadata in terms of interparametric relationships that can be used to generate metadata. The experience of the music experts is the binding factor that ensures the quality of the metadata. Their context- and purpose-awareness should be reflected in the metadata and only they can validate this.

Current work includes a conceptual design stage of human algorithms for mood classification in the form of a collaborative, competitive game [4]. As a starting point, we are thinking of experiments to map a musical mood to the mood of something else, e.g. a picture. We are also looking for closer integration of metadata generation, data mining techniques and music experts.

4. ACKNOWLEDGEMENT

We gratefully acknowledge the financial support of the IWT Vlaanderen (http:///www.iwt.be), through the project: Music Metadata Generation IWT-060237.

5. REFERENCES

[1] S. Govaerts, N. Corthaut, E. Duval, "Moody Tunes: The rockanango project.", *Proc. of the 7th Int. Conf. on Music Information Retrieval*, 2006.

[2] Schedl M., Pohle T., Knees P., Widmer G., "Assigning and visualising Music Genres by Web-based Co-Occurrence Analysis", *Proc. of the 7th Int. Conf. on Music Information Retrieval*, 2006,.

[3] L. von Ahn, L. Dabbish, "Labelling images with a Computer Game", *CHI 2004*, 24-29 April 2004.

[4] Raph Koster, *A Theory of Fun for Game Design*, Paraglyph, 1st. edition, Nov. 2004.

[5] Ian H. Witten and Eibe Frank, *Data Mining: Practical machine learning tools and techniques*, 2nd Edition, 2005.

[6] Mark Levine, *The Jazz Piano Book*, Sher Music, June 1, 2005

[7] Matsuo, Y. and M. Ishizuka, "Keyword Extraction from a Single Document using Word Co-occurrence Statistical Information.", *Int. Journ. on Artificial Intelligence Tools*, 13(1): 157-169, 2004.

[8] Sado, W. N., et al., "A Linguistic and Statistical Approach for Extracting Knowledge from Documents", *Proc. of the 15th Int. Workshop on Database and Expert Systems Applications (DEXA'04)*, 2004.

[9] Hulth, A., et al., "Automatic Keyword Extraction Using Domain Knowledge", *Proc. of the 2nd Int. Conf. on Computational Linguistics and Intelligent Text Processing*, 2001.

[10] Witten I.H., Paynter G.W., Frank E., Gutwin C. and Nevill-Manning C.G., "KEA: Practical automatic keyphrase extraction." *Proc. DL '99*, pp. 254-256.

[11] Yahoo Web Services API Term Extraction, http://developer.yahoo.com/search/content/V1/termExtraction.html

[12] Kraft, R. et al., "Y!Q: Contextual Search at the Point of Inspiration", *ACM Conf. on Information and Knowledge Management (CIKM'05)*, 2005.

[13] G. Geleijnse, J. Korst, "Efficient Lyrics Extraction from the Web", *Proc. of the 7th Int. Conf. on Music Information Retrieval*, 2006.

[14] Stevens, F., Van Den Broek, A., & Vandeweyer, J., "Time use of adolescents in Belgium and the Netherlands", *25th IATUR Conf. on Time Use Research*

[15] Amatriain, X., de Boer, M., Robledo, E. & Garcia, D., "CLAM: an OO framework for developing audio and music applications", *Companion of the 17th annual ACM SIGPLAN Conf. on Object-oriented Programming Systems, Languages, and Applications*, 2002.