A Self-Organizing Map Based Knowledge Discovery for Music Recommendation Systems

Shankar Vembu[1], Stephan Baumann[2]

[1] Technical University of Hamburg, Harburg
D-21071, Hamburg, Germany
shankar.vembu@tuhh.de
[2] German Research Center for Artificial Intelligence
Erwin Schrödinger Str., 67663 Kaiserslautern, Germany
stephan.baumann@dfki.de

Abstract. In this paper, we present an approach for musical artist recommendation based on Self-Organizing Maps (SOMs) of artist reviews from Amazon web site. The Amazon reviews for the artists are obtained using the Amazon web service interface and stored in the form of textual documents that form the basis for the formation of the SOMs. The idea is to spatially organize these textual documents wherein similar documents are located nearby. We make an attempt to exploit the similarities between different artist reviews to provide insights into similar artists that can be used in a recommendation service. We introduce the concept of a modified weighting scheme for text mining in the musical domain and demonstrate its role in improving the quality of the recommendations. Finally, we present results for a list of around 400 musical artists and validate them using recommendations from a popular recommendation service.

1 Introduction

Music similarity perception plays an important role in recommendation services for musical information. Given the huge amount of musical data available online, there is an increasing demand to provide quality-enriched services to allow access and browse musical content over the web. Various services are being offered by online music recommendation systems that use different approaches to compute music similarities [1], [2]. We present here in this paper an approach for musical artist recommendations using the album reviews for the artists available from the Amazon web site [3]. The basic idea of our approach is to spatially organize these reviews that are in the form of textual documents using an unsupervised learning algorithm called Self-Organizing Maps [4] and thus be able to give recommendations for similar artists by making use of the model built by the algorithm.

An approach to musical artist recommendations based on cultural metadata has been presented in [5], [6]. A comparison of audio-based music similarity measures in a peer-to-peer setting has been made in [7]. The use of SOMs in the field of text mining has been reported in [8], [9] and the WEBSOM project [10]. SOMs have also been used in the field of Music Information Retrieval for automatic analysis and organization of musical archives based on the audio content [11], [12] and in the SOM-enhanced JukeBox Music Digital Library project [13]. In this paper we investigate the use of SOMs for artist recommendations not based on the content-based audio analysis but by relying on textual information available in the web.

1.1 Amazon Web Service

The online shopping mall Amazon has made its products available through an interface [14] based on web service standards [15]. Using this interface, one is able to retrieve product information like artists/musicians, authors, ASIN and ISBN directly from the Amazon servers.

1.2 Self-Organizing Maps

The Self-Organizing Map (SOM) is an unsupervised learning algorithm used to visualize and interpret large high-dimensional data sets. The map consists of a regular grid of processing units called "neurons". Each unit is associated with a model of some high dimensional observation represented by a feature vector. The map attempts to represent all the available observations with optimal accuracy using a restricted set of models. Map units that lie nearby on the grid are called *neighbors*. After the formation of a map for a particular data set, the model vectors are arranged in such a manner that nearby map units represents similar kind of data and distant map units represent different kinds of data. The reader is referred to [4] for a detailed description on this subject.

2 Self-Organizing Maps of Amazon Reviews

2.1 Textual Data Mining

The album reviews for artists can be accessed from the Amazon site using the Amazon Web Service interface [14] that is available as a standard development kit. Every artist can then be represented as a collection of his/her album reviews. The interface provides facilities to query a broad range of products from the Amazon site. The queries can be submitted either as a web service SOAP [15] message or using XML [16] over HTTP [16]. The latter uses URIs (Uniform Resource Indicators) with specific name/value pairs to invoke methods and processes within Amazon's Web Services framework. The returned response, which is a SOAP message in the former case and a well-formed XML document in the latter case, contains the complete product information.

The literature on Information Retrieval [17] provides techniques to preprocess and represent textual documents for mining operations. Preprocessing techniques include stripping unwanted characters/markup (e.g. HTML tags, punctuation, numbers, etc.) and removing common stop words (e.g. a, the, of, etc.). The documents are represented in the form of a bag-of-words where each document is considered as a point (or vector) in an n-dimensional Euclidean space where each dimension corresponds to a word (term) of the vocabulary. The i^{th} component d_i of the document vector expresses the number of times the word with index i occurs in the document, or a function of it. Furthermore, each word can be assigned a weight signifying its importance. Commonly used weighting strategy is the tf * idf (term frequency – inverted document frequency) scheme [18], where tf stands for term frequency within the document, and idf stands for the inverse of the number of documents in which the term appears. The scheme is based on the notion that a word that occurs frequently in the document but rarely in the rest of the collection is given more importance. A variant of the general tf * idf scheme that assigns a weight to every word is given by

$$w_{ij} = tf_{ij} * idf_i = tf_{ij} * \log_2 (N/ df_i) \qquad \textbf{(1)}$$

where,

w_{ij} is the tf-idf weight for the i^{th} word in j^{th} document in a collection of N documents,

tf_{ij} is the term frequency of the i^{th} word in the j^{th} document and

$idf_i = \log_2 (N/df_i)$ is the inverse document frequency of the i^{th} word over the entire collection.

2.2 Modified Weighting Scheme for Text Mining in the Musical Domain

The tf * idf weighting scheme described above does not take into consideration any domain knowledge to determine the importance of a word. The words are given importance only based on the frequencies of their occurrence in the document as well as the collection. But when trying to find similarities between two documents in a musical context, it is desirable to exploit any domain knowledge that is inherently present in the documents. We propose one such mechanism to accomplish this by introducing the concept of a modified weighting scheme in the musical domain or context.

The album reviews for artists that are in plain English contain a few words that are more relevant and ought to be given more importance in the musical context. For example, consider an album review for a popular *rap* artist "Eminem" from the Amazon web site that says,

"…This is Eminem before he became the hottest rapper on Earth, when he was a lot more funny and creative. There are only like 3 songs that I don't like. The beats are just ok, but he makes up for it with his witty lyrics and tight flow. If you're into rap, you'll like it…"

Words like "rap" in the above example definitely say a lot about the artist being a rap artist and thus deserve more importance when the above document is represented. Eventually, one would like to find similarities between documents with words like the one described above being given more importance in comparison to other words occurring in the English language. Therefore, in addition to the weighting importance given to a word by the tf *idf scheme, it would be worthwhile to increase the weight of a word by a certain factor if it is pertaining to the musical domain. A few sample words that deserve more attention in a musical context are *samba, rap, hip-hop,* *metal, instrumental*. This would allow one to find similarities between documents not only by the occurrence frequencies of the words in them but also based on the musical importance of the words. Coming up with such a list of musical words, also plays an important role in the quality of the recommendations. We came up with such a word list of 324 from the genre taxonomies in [2]. The quality of document representation and thereby the results of the recommendations would definitely be increased if this modified importance weighting were incorporated into the system. The modified weighting scheme gives rise to a new weight for musical words that is given by

$$w^m_{ij} = tf^m_{ij} * idf^m_i = tf^m_{ij} * \log_2 (N/df^m_i) * \alpha \quad \textbf{(2)}$$

where the superscript m indicates words belonging to the musical context and α is the weighting increment.

2.3 Map Formation

The preprocessed textual documents represented in the form of n-dimensional vectors can be used to train a Self-Organizing Map in an unsupervised way. The learning starts with a set of reference vectors also called the model vectors that are the actual map units of the network. As the learning proceeds, the model vectors gradually change or arrange themselves so as to approximate the input data space. The final arrangement is such that the model vectors that are nearby are similar to each other. The model vectors are usually constrained to a two-dimensional regular grid, and by virtue of the learning algorithm, follow the distribution of the data in a nonlinear fashion. The model vectors are fitted using a sequential regression process. Given a sample vector x(t) at iteration step t the model vector $m_i(t)$ with index *i* is adapted as follows:

$$m_i(t + 1) = m_i(t) + h_{c(x),i}(t)[x(t) - m_i(t)], \quad \textbf{(3)}$$

where the index of the "winner" model, c for the current sample is identified by the condition,

$$\forall i,\ \|x(t) - m_c(t)\| \leq \|x(t) - m_i(t)\|. \quad \textbf{(4)}$$

$h_{c(x),i}(t)$ is called the *neighborhood function*, which acts as a smoothing kernel over the grid, centered at the "winner" model $m_c(t)$ of the current data sample. The neighborhood function is a decreasing function of the distance between the *i*th and *c*th nodes on the map grid. The regression is usually reiterated over all the available samples. Thus, with this unsupervised learning algorithm we can spatially arrange all the documents i.e. the album reviews of all the artists, resulting in a topological ordering of the artists. In addition to this, the SOM algorithm also obtains a clustering of the data onto the model vectors wherein the artists present in a particular cluster are similar to each other. In Fig. 1, if we consider map unit A, the artists in itself and its neighboring units show similarities.

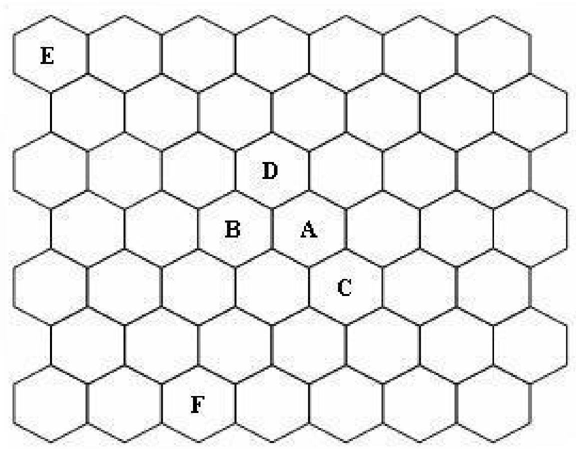

Fig. 1. A 7 * 7 SOM for the artist reviews with a hexagonal grid. Artists in map unit *A* are more similar to artists in nearby units like *B*, *C* or *D* than in far away units like *E* or *F*, which is by virtue of the self-organization of the map due to the learning algorithm

3 Experiments and Results

3.1 SOMs of the Whitman Artist List

We implemented the SOM based approach for artist recommendations described in this paper for the Whitman musical artist list [6] to build an SOM model for 398 popular artists. The most important album reviews for each artist of this list were obtained using the Amazon Web Service interface and stored as textual documents one for each artist. Each document was preprocessed and represented using standard techniques from the field of Information Retrieval [17]. Preprocessing techniques include stripping unwanted characters/markup (e.g. HTML tags, punctuation, numbers, etc.) and removing common stop words (e.g. a, the, of, etc.). The documents were then represented as a bag-of-words with each word being assigned a weight according to the tf-idf schema. We came up with a list of 324 musical words, to incorporate the concept of the modified weighting scheme described above in the musical context. If a document had words present in this list, its weight was increased by a factor of $\alpha = 4$. The weights were then normalized so that the size of the documents would not have any effect in their representation. A full-term indexing of the documents thus yielded a feature vector of 36,708 words for the data set of size 398. We then removed the words that were present in less than 5% and more than 90% of the collection as these words do not play an important role in classification [13]. This dramatically reduced the size of the feature vector to 3313. We thus had all the documents represented as a 398 * 3313 matrix.

For the map formation, we used the SOM toolbox [19] available for MATLAB.

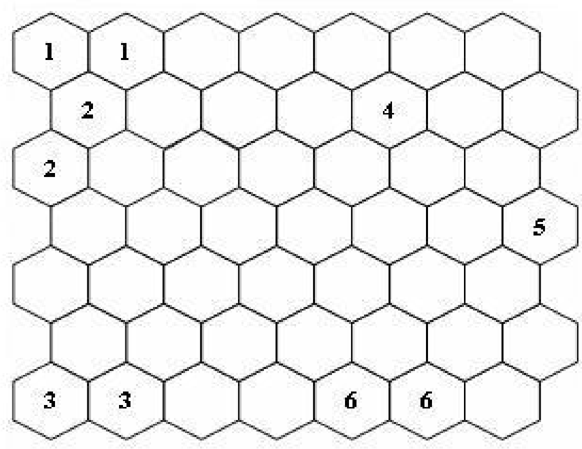

Fig. 2. Results for the Whitman artist list. The labels of the numbered map units and some of the artists present in them are given in Table 1

Table 1. Labels and artists for the numbered map units shown in Fig. 2

Map unit	Labels	Artists
1	Rap, Funk	Limp Bizkit, Korn, Linkin Park
2	Reggae, Rap	Bob Marley, Shaggy, Police, 311
3	Metal, Heavy	Metallica, Nirvana, Smashing Pump- kins
4	Techno, Remix	Alice Deejay, Prodigy, Moby, Chemical Broth- ers
5	Spanish, Latin	Ricky Martin, Enrique Iglesias
6	Rock, Roll, Blues	Dire Straits, Queen, Eric Clapton, Bob Dylan

We trained a Self-Organizing Map consisting of 49 map units in a hexagonal grid of size 7*7 for all the 398 artists using the 398 * 3313 matrix mentioned above. After the training phases, each map unit represented a model obtained from the collection of reviews in the high dimensional space. The entire data collection of 398 artists were divided according to the obtained map models on the two-dimensional grid and spatially organized based on their similarities.

3.2 SOM Labeling and Visualization

Labeling plays an important role in the visualization of the self-organizing map [20, 21]. We employed a simple labeling technique where a map unit is represented by a label or a keyword that has a higher weight, as calculated by the tf * idf weighting scheme, when compared to other words that appear in the map unit. The modified weighting scheme described in the previous sections also aided in labeling the map units. Since we increase the weight of a word that pertains to the musical context, many, if not all, of the labels that we obtained were from the musical list of words. This is indeed desirable when we are labeling an SOM of artists as we would like to see labels that are musical words like *rap, rock, metal, blues* and not plain English words. We present here the results of the SOM model described in the previous section with labeling. Fig. 2 shows a few distinct sections of the map with their respective labels and artists in Table 1. As can been seen from the results, we were able to obtain a clear categorization of artists based on different musical genres (rap, metal, rock, blues, techno etc.)

3.3 Validations

The results of our experiments were validated using Echocloud [1], a web-based music recommendation engine. Echocloud works by crawling peer-to-peer networks to capture users' file lists in order to discover correlations between musical artists. We would like to mention some limiting factors that made the validations difficult.

1. The Echocloud approach to finding similar artists differs widely from ours. Our approach entirely depends on the artist reviews from the Amazon site, whereas Echocloud is based on the notion that if two artists are found together in users' file lists, they are similar in some sense as having one tends to imply having the other [1].
2. The Echocloud builds the similarity model using a database of around 120k artists [1] in the so-called *open world* environment of users' music collection, whereas we performed experiments on the much-restricted Whitman artist list of around 400 artists.
3. Music similarity perception is *subjective* in nature. Even though we do not present a *subjective* validation in this paper, we believe that such a validation is more efficient for experiments related to music similarity perception. We presented such an approach in an ecological environment in [22].

We compare the Top 10 recommendations from Echocloud with our Top 10 recommendations for all the artists. We also compare Echocloud recommendations with

Table 2. Validations of our results for 398 artists with Echocloud's Top 10 recommendations without modified weighting scheme

Comparisons	Total number of recommendation matches	Average recommendation match in percentage
Top 10	482 / 3980	12.1 %
3 BMUs	685 / 3980	17.2 %
5 BMUs	982 / 3980	24.7 %

Table 3. Validations of our results for 398 artists with Echoclould's Top 10 recommendations with modified weighting scheme

Comparisons	Total number of recommendation matches	Average recommendation match in percentage
Top 10	493 / 3980	12.4 %
3 BMUs	785 / 3980	19.7 %
5 BMUs	1038 / 3980	26.1 %

the artists that are present in the 3 and 5 Best Matching Units (BMUs) of the artist in question. A Best Matching Unit for an artist i.e. textual document for the review is the SOM map unit that best models it. The Euclidean distance measure is used in finding the BMUs for an artist.

As can be seen from the results in Tables 2 and 3, the quality of the recommendations increases as we move up from the Top 10 comparisons to 5 BMUs. This leads to an interesting observation that even though all the Echocloud recommendations do not match with our Top 10 recommendations, some of them do certainly lie in the nearby regions namely the 3 or 5 BMUs, which is by virtue of the map's self organization. Most importantly, we were able to see an increase in the quality of recommendations when the modified weighting scheme that exploits the musical domain knowledge was incorporated into the system. We believe that the list of 324 words pertaining to the musical context that was used for this can play an important role in a recommendation service to yield better results, even though there is human intervention in coming up with such a list.

4 MYMO: A Prototypical Application for Mobile Music Information Retrieval

4.1 Existing Framework

We came up with a mobile music recommendation engine called MYMO that uses internal as well as external services to provide recommendations for similar songs and artists. The song similarities are computed internally using a trimodal global similarity measure to provide the user a set of similar songs given an anchor song. This measure is realized as a weighted linear combination of three different local similarity metrics, namely sounds-alike similarity, similarity of lyrics and cultural or stylistic similarity.

$$S = wso * Sso + wly * Sly + wst * Sst \quad (\mathbf{5})$$

where
Sso: sounds-alike similarity
Sly: similarity of lyrics
Sst: similarity by style/cultural aspects and
wso, wly, wst are the respective weights.

For a detailed description of the local similarity metrics, the reader is referred to [22]. For the artist similarities, we interfaced with an external recommendation service, Echocloud [1]. The engine was realized as a web based solution (Fig. 3). The web site can be displayed on a small screen typical of PDA (personal digital assistant) or PIM (personal information management) devices. To allow the user subjective and interactive feedback, a virtual joystick was included that can be easily accessed using the pen of the PDA. The recommendation engine uses the song similarity measure described above. The position of the joystick has a direct influence on the individual weights in the linear combination. In this way the individual users can select different settings and find their favorite combination.

4.2 Integrating SOM Based Artist Recommendation Approach into MYMO

The SOM based approach for musical artist recommendations described in the previous sections can be integrated into our existing framework of mobile music recommendation service to provide recommendations for similar artists. This can be used as a replacement for the external Echocloud service that is currently being used to provide artist recommendations. The SOMs provide a spatial representation of the high dimensional artist reviews. A model can thus be formed for a set of artists where similar artists are located nearby spatially and dissimilar artists lie farther away. Furthermore, it is not mandatory that the queried artist be present in the model formed by the SOMs. The review for such a query can be obtained online from the Amazon site and the best-matching unit (BMU) for this review can be found from the model. The artists present in this BMU are therefore similar to the queried artist.

5 Conclusions and Discussions

We have presented in this paper a novel approach to musical artist recommendations using Amazon reviews and Self-Organizing Maps. We introduced the concept of a modified weighting scheme for text mining in the musical domain and demonstrated its role in improving the quality of recommendations. We presented results for the Whitman artist list of around 400 musical artists and validated them with the recommendations from Echocloud, a web based recommendation service for musical artists.

Fig. 3. MYMO: Mobile Music Recommendation Engine

We also demonstrated how this approach can be used in a music recommendation service. Exploiting the semantics of the language behind the artist reviews by using Natural Language Processing techniques could be a possible future work. The goal is to extract as much relevant information as possible from the artist reviews by using the state-of-the-art technologies from the fields of Information Retrieval and Natural Language Processing. We believe that the approach presented in this paper would be well suited for use in a recommendation service to provide quality services to all the music lovers of the world.

6 Acknowledgements

We wish to thank the *Stiftung Rheinland-Pfalz für Innovation* for supporting this work.

References

1. http://www.echocloud.net
2. http://www.allmusic.com

3. http://www.amazon.com
4. Kohonen, T.: *Self-Organizing Maps*, Springer, Berlin, Heidelberg, 1995
5. Baumann, S., Hummel, O.: "Using Cultural Metadata for Artist Recommendation", Proc. Of the Third Conference on Web Delivering of Music (WEDELMUSIC 2003), Leeds, UK, September, 2003
6. Whitman, B., Lawrence, S.: "Inferring Descriptions and Similarity for Music from Community Metadata", Proceedings of the 2002 ICMC, Göteborg, Sweden, 16-21 Sep.2002, pp 591-598
7. Baumann, S., Pohle, T.: "A Comparison of Music Similarity Measures for a P2P Application", Proc. of the 6th Int. Conference on Digital Audio Effects (DAFX-03), London, UK, September 8-11, 2003
8. Lagus, K.: "Text retrieval using self-organized document maps", Technical Report A61, Helsinki University of Technology, Laboratory of Computer and Information Science, ISBN 951-22-5145-0, 2000
9. Merkl, D. and Rauber, A.: "Document Classification with Unsupervised Neural Networks", in Fabio Crestani, Gabriella Pasi (Eds.), Soft Computing in Information Retrieval, Physica Verlag & Co, ISBN:3790812994, Germany, 2000, pp. 102 - 121
10. Lagus, K.,: "Text Mining with the WEBSOM" Acta Polytechnica Scandinavica, Mathematics and Computing Series no. 110, Espoo 2000, 54 pp. D.Sc.(Tech) Thesis, Helsinki University of Technology, Finland, 2000
11. Rauber, A. and Frühwirth, M.: "Automatically Analyzing and Organizing Music Archives", In Proceedings of the 5th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2001), Darmstadt, Germany, Sept. 4-8 2001, Springer Lecture Notes in Computer Science, Springer, 2001
12. Pampalk, E., Dixon, S. and Widmer, G.: "Exploring Music Collections by Browsing Different Views", Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR'03), Baltimore, MD, October 26-30, 2003, pp 201-208
13. Rauber, A., Pampalk, E. and Merkl, D.: "The SOM-enhanced JukeBox: Organization and Visualization of Music Collections based on Perceptual Models", In: Journal of New Music Research (JNMR), 32(2):193-210, Swets and Zeitlinger, June 2003
14. Amazon Web Service, http://www.amazon.com/gp/aws/landing.html
15. Web Services Activity, http://www.w3.org/2002/ws/
16. World Wide Web Consortium, http://www.w3.org/
17. Ricardo Baeza-Yates and Berthier Ribeiro-Neto: *Modern Information Retrieval*, Addison-Wesley, 1999
18. Salton, G. and Buckley, C.: "Term weighting approaches in automatic text retrieval", Tech. Rep. 87-881, Cornell University, Department of Computer Science, Ithaca, NY, 1987
19. SOM Toolbox for MATLAB, Helsinki University of Technology, Laboratory of Computer and Information Science, Neural Networks Research Centre
20. Rauber, A.: "LabelSOM: On the Labeling of Self-Organizing Maps", Proceedings of the International Joint Conference on Neural Networks (IJCNN'99), Washington, DC, July 10 - 16, 1999
21. Lagus, K. and Kaski, S.: "Keyword selection method for characterizing text document maps", Proceedings of ICANN99, Ninth International Conference on Artificial Neural Networks volume 1, pages 371-376, IEE, London, 1999
22. Baumann, S. and John Halloran: "An ecological approach to multimodal subjective music similarity perception", to appear in Proc. Of the First International Conf. On Interdisciplinary Musicology, Graz 15.-18.April, 2004