Visualization and Clustering of Tagged Music Data

Pascal Lehwark[1], Sebastian Risi[2], and Alfred Ultsch[3]

[1] Databionics Research Group, Philipps University Marburg pascal@indiji.com
[2] Databionics Research Group, Philipps University Marburg sebastian.risi@gmail.com
[3] Databionics Research Group, Philipps University Marburg ultsch@informatik.uni-marburg.de

Abstract. The process of assigning keywords to a special group of objects is often called tagging and becomes an important character of community based networks like Flickr, YouTube or Last.fm. This kind of user generated content can be used to define a similarity measure for those objects. The usage of Emergent-Self-Organizing-Maps (ESOM) and U-Map techniques to visualize and cluster this sort of tagged data to discover emergent structures in collections of music is reported. An item is described by the feature vector of the most frequently used tags. A meaningful similarity measure for the resulting vectors needs to be defined by removing redundancies and adjusting the variances. In this work we present the principles and first examples of the resulting U-Maps.

1 Introduction

The increased interest in folksonomies like Flickr, Last.fm, YouTube, del.icio.us and other community based networks shows, that tagging is already used by many users to discover new material and becomes a collaborative way of classifying items, being controlled by the creator and consumer of the content. One popular way to visualize tag relations is the use of *tag clouds*. They are used to visualize the most used tags on a website. More frequently used tags have a larger font and they are normally ordered alphabetically. For our study we chose to analyse the data provided by the music community *Last.fm*, an internet radio featuring a music recommendation system. The users can assign tags to artists and browse the content via tags allowing them to only listen to songs tagged in a certain way.

Tags make it possible to organize the media (artists and songs) in a semantic way and states a useful base for discovering new music. Because of the huge amount of artists and songs, an intuitive user interface is required to avoid losing the overview. We propose the *Emergent-Self-Organizing-Map (ESOM)* (Ultsch (2003)) to cluster tagged data because it has some advantages over

other clustering algorithms. It is topology preserving and combined with the *U-Map* it provides a visually appealing user interface and an intuitive way of exploring new content. The remainder of this paper is organized as follows. First some related work on tagged data, clustering music and documents with the ESOM is presented. Then we describe the main learning algorithm of the ESOM in section 3 together with the U-Map visualization. Next, the dataset is presented together with the used methods of data preparation. In section 5 we present our experimental results. We round off the paper by giving the conclusion in section 6 together with future research directions.

2 Related Work

There has been some work on enhancing the user interface based on tags and we will briefly mention some here. Flickr uses *Flickr clusters* which can provide related tags to a popular tag, grouped into clusters. Begelman (2006) uses clustering algorithms to find strongly related tags visualizing them as a graph. Hassan-Montero et al. (2006) propose a method for an improved tag cloud and a technique to display these tags with clustering based layout.

The ESOM has already been used successfully to visualize collections of music, photos and on clustering documents. Most of these works have in common that they cluster the data based on features extracted directly from the media. An example is MusicMiner (Mörchen (2005)) which uses the *timbre distance*, a measure based on frequency analysis of audio data. The *websom project* (Kaski (1998)) is an ESOM based approach in free text mining. Here each document is encoded as a histogram of word categories which are formed by the ESOM algorithm based on the similarities in the contexts of the words.

Although our approach is different because we are not using information that can be extracted from the objects' raw data itself but instead user generated content, the works mentioned previously show that the ESOM is a powerful tool in visualizing high dimensional data.

3 Emergent Self Organizing Maps

The ESOM is an artificial neural network that performs a mapping from a high dimensional data space R^n onto a two-dimensional grid of neurons. The unsupervised training process is partly motivated by how visual information is handled in the cerebral cortex of the mammalian brain and equals a regression of an ordered set of model vectors $m_i \in R^n$ into the space of observation vectors $x \in R^n$ by performing the following process:

$$m_i(t+1) = m_i(t) + h_{c(x),i}(x(t) - m_i(t))$$

where t is the sample index of the regression step, whereby the regression is performed recursively for each presentation of a sample of x. Index c, the *bestmatching unit* (BMU) or *winner*, is defined by the condition

$$||x(t) - m_c(t)|| \leq ||x(t) - m_i(t)|| \forall i$$

The so called *neighbourhood function* h is often taken to be the Gaussian

$$h_{c(x),i} = \alpha(t) exp(-\frac{||r_i - r_c||^2}{2\sigma^2(t)})$$

where $0 < \alpha(t) < 1$ is the learning-rate factor, which decreases monotonically with the regression steps, r_i and r_c are the vectorial locations in the display grid and $\sigma(t)$ corresponds to the width of the neighbourhood function, which is also decreasing monotonically with the regression steps. For a more detailed discussion of the SOM see Kaski (1997).

3.1 U-Map Visualization

The U-Map (Ultsch (2003)) is constructed on top of the map of ESOM. The *U-Height* for each neuron n_i equals the accumulated distances of n_i to its immediate neighbors $N(i)$. It is calculated as follows:

$$\text{U-Height}(n_i) = \sum_{j \in N(i)} d(m_i, m_j)$$

where $d(x, y)$ is the distance function used in the SOM algorithm to construct the map and $N(i)$ denotes the indices of the immediate neighbours of neuron i.

A single U-Height shows the local distance structure of the corresponding neuron. The overall structure of densities emerges, if a global view of a U-Map is regarded. A U-Map is usually displayed as a three dimensional landscape and has become a standard tool to display the distance structures of the ESOM. Therefore the U-Map delivers a 'landscape' of the distance relationships of the input data in the data space. It has the property that weight vectors of neurons with large U-Heights are very distant from other vectors in the data space and that weight vectors of neurons with small U-Heights are surrounded by other vectors in the data space. Outliers and other possible cluster structures can easily be recognized. U-Maps have been used in a number of applications to detect new and meaningful knowledge in data sets.

4 Data

We extracted 1200 artists from the Last.fm website together with the 250 most frequently used tags like *rock*, *pop*, *metal*, etc.

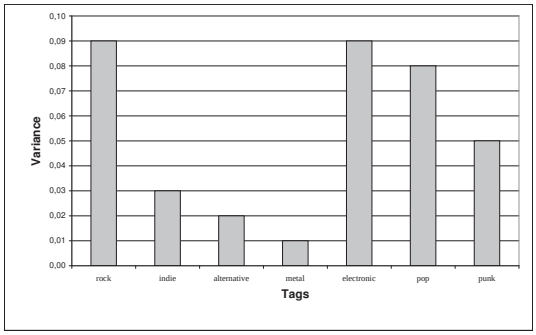


Fig. 1. The variances of the seven most popular tags

4.1 Peparation of the datasets

Before the ESOM can be trained, special demands have to be fulfilled. Tags from the Last.fm dataset which do not stand for a certain kind of music genre, like *seen-live*, *favourite albums*, etc. were excluded. Highly correlated tags were condensed to a single feature. For the preparation of the tagged data we used a modification of the *Inverse Document Frequency (IDF)*.

Last.fm provides the number of people ($t_{ij} = tagcount_{ij}$) that have used a specific tag i for an artist j. We scaled t_{ij} to the range of $[0, 1]$. Then we slightly modified the term frequency to be more appropriate for tagged data:

$$\mathrm{tf}_{ij} = \frac{t_{ij}}{\sum_k t_{kj}}$$

with the denominator being the accumulated frequencies over all tags used for artist j. The *IDF* of tag i is defined as

$$\mathrm{idf}_i = log \frac{|D|}{\sum_k t_{ik}}$$

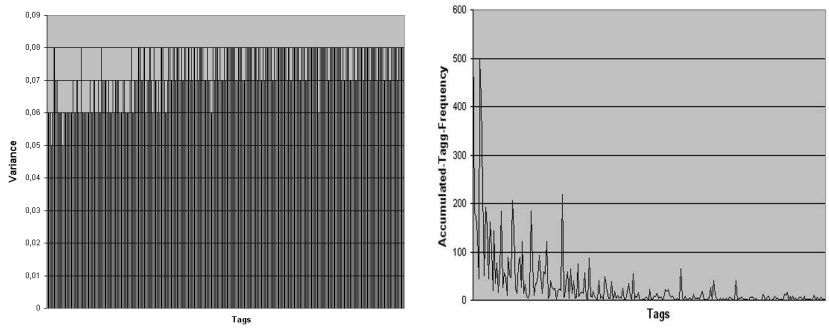
with $|D|$ being the total number of artists in the collection and $\sum_k t_{ik}$ being the accumulated frequencies of tag i over all artists. The resulting importance of tag i for artist j is given by

$$\mathrm{tfidf}_{ij} = \mathrm{tf}_{ij}\mathrm{idf}_i$$

As can be seen in figure 1 all the tags of the Last.fm dataset differ a lot in variance but for a meaningful comparison of the variables these variances have to be adjusted. For this purpose we used the *empirical cumulative distribution function (ECDF)*. The idea behind the ECDF is to assign a probability of $\frac{1}{n}$ to each of the n observations in the sample. The final tag frequencies are then given by

$$\text{tfidf}_{ij}^{ECDF} = \frac{|\text{tfidf}_{ik} \leq \text{tfidf}_{ij}|}{n}, k = 1..n$$

The adjusted variances after applying ECDF can be seen in figure 2. The accumulated tag frequencies of the Last.fm dataset can be seen in figure 3.



(a) The ECDF adjusted variances of the Last.fm tags

(b) Accumulated tag frequencies

Fig. 2. Tag variances

Finally we optain the feature vector w_j for artist j as

$$w_j = (\text{tfidf}_{1j}^{ECDF}, .., \text{tfidf}_{nj}^{ECDF})$$

In the context of self organizing maps, two different measures have been proposed to compute the similarity between two feature vectors w_i and w_j. The first method uses the familiar *euclidean distance*, while the second approach is based on the *cosine similarity*

$$cos(w_i, w_j) = \frac{w_i^t w_j}{\|w_i\| \, \|w_j\|}$$

This method emphasizes the relative values that each dimension has within each vector and not their overall length. Two vectors can have a value of zero even if their euclidean distance is arbitrarily large. A SOM model which uses the cosine similarity instead of euclidean distance has also been proposed by Kohonen (1982), introduced as *Dot-Product-SOM*, and has been succesfully used for document clustering problems. For the close analogy to tag spaces we decided to use this model rather than the standard model based on the euclidean distance.

Note, that the update function changes to

$$m_i(t+1) = \frac{m_i(t) + h_{c(x),i}(t)x(t)}{\left\|m_i(t) + h_{c(x),i}(t)x(t)\right\|}$$

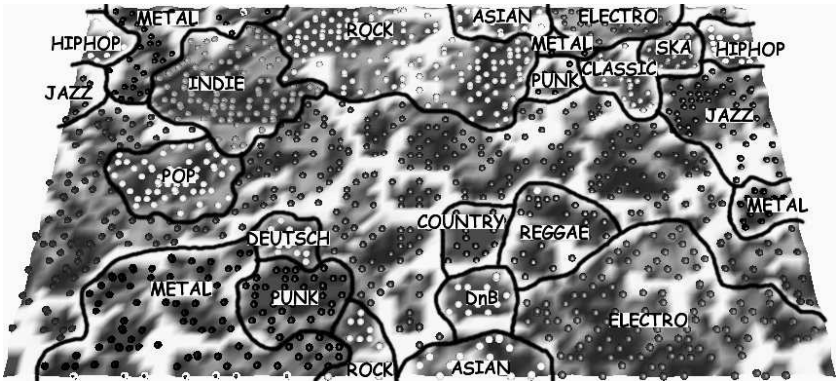


Fig. 3. Resulting U-Map. Note that the map is toroid, for example, the metal cluster is not split but spreaded over boundaries.

Although the training process slows down due to the normalization at each step, the search for the bestmatch is very fast and simple.

5 Experimental Results

We trained a 80x50 emergent self organizing map using 50 epochs with the preprocessed data using the *Databionics ESOM tool* (Ultsch and Mörchen, 2005). A *toroid topology* was used to avoid border effects.

The U-Map in figure 4 (visualized using *Spin3D*) can be interpreted as height values on top of the usually two dimensional grid of the ESOM, leading to an intuitive paradigm of a landscape. Clearly defined borders between clusters, where large distances in data space are present, are visualized in form of high mountains. Smaller intra cluster distances or borders of overlapping clusters form smaller hills. Homogeneous regions of data space are placed in valleys.

Detailed inspection of the map shows a very good conservation of the intercluster relations between the different music genres. One can observe smooth transitions between clusters like *metal*, *rock*, *indie* and *pop*.
In figure 5 we show a detailed view of the cluster *rock*.

The innercluster relations, e.g. the relations between genres like *hard rock*, *classic rock*, *rock and roll* and *modern rock* are very well preserved. This property also holds for the other clusters.

An interesting area is the little cluster *metal* next to the cluster *classic*. A precise examination revealed the reason for this cluster not being part of the big cluster *metal*. The cluster *classic* contains the former classic artists like Ludwig van Beethoven on the lower right edge with a transition to newer artists of the classical genre when moving to the upper left. The neighbouring

Fig. 4. Zoom of cluster *rock* to illustrate the good innercluster quality.

artists of the minicluster *metal* are bands like Apocalyptica and Therion which use a lot of classical elements in their songs.

6 Conclusion and future work

Our goal was to find a visualization method that fits the need and constraints of browsing collections of tagged data. A high dimensional feature vector of 250 dimensions is hard to grasp and clustering can reveal groups of similar objects based on their tags. The global organization of the tagged artists worked really well and in contrast to other clustering algorithms, soft transitions between the groups of similar tagged artist can be seen. The modified *Inverse Document Frequency* turned out to be a good preparation method when working with tagged data. It is however essential for the ESOM that the feature vectors are not to sparse and that the overlap between them is not to low. These problems occurred in experiments with the photo community *flickr* where information about tags is only binary (a tag occurs or not) without information about the tag frequencies.

We showed that the ESOM enables the user to navigate through the high dimensional space in an intuitive way. Future work could include combining the clustering of artists and their songs and an automatic playlist generation system from regions and paths on the map. The maps presented here can be seen in color and high resolution at www.indiji.com/musicsom.

References

BEGELMAN, G., KELLER, P., SMADJA, F. Automated Tag Clustering: Improving search and exploration in the tag space http://www.rawsugar.com/lab.

HASSAN-MONTERO, Y., HERRERO-SOLANA, V. Improving Tag-Clouds as Visual Information Retrieval Interfaces To appear: International Conference on Multidisciplinary Information Sciences and Technologies, InSciT2006, Merida, Spain, 2006.

KASKI, S., HONKELA, T. LAGUS, K., KOHONEN, T. WEBSOM–self-organizing maps of document collections. Neurocomputing, volume 21, pages 101-117, 1998

KASKI, S, KANGASZ, J., KOHONEN, T. Bibliography of self-organizing map (SOM) papers: 1981-1997

KOHONEN, T. Self-Organized Formation of Topological Correct Feature Maps Biological Kybernetics Vol. 43, pp59-69, 1982.

MATHES, A. Folksonomies Cooperative Classification and Communication Through Shared Metadata. http://www.adammathes.com/academic/ computermediatedcommunication/folksonomies.html.

MILLEN, D., FEINBERG, J. Using Social tagging to Improve Social Navigation Workshop on the Social Navigation and Community based Adaptation Technologies, 2006

MILLEN, D., FEINBERG, J., KERR, B. Social Bookmarking in the Enterprise. Social Computing, Vol. 3, No. 9. Nov. 2005.

MÖRCHEN, F., ULTSCH, A., NÖCKER, M., STAMM, C. Visual mining in music collections In Proceedings 29th Annual Conference of the German Classification Society (GfKl 2005), Magdeburg, Germany, Springer, Heidelberg, 2005

MÖRCHEN, F., ULTSCH, A., THIES, M., LÖHKEN, I., NÖCKER, M., STAMM, C., EFTHYMIOU, N., KÜMMERER, M. MusicMiner: Visualizing timbre distances of music as topographical maps Technical Report No. 47, Dept. of Mathematics and Computer Science, University of Marburg, Germany, 2005

ROBERTSON, S., Understanding Inverse Document Frequency: On theoretical arguments for IDF Journal of Documentation 60 no. 5, pp 503520

ULTSCH, A. Self-organizing neural networks for visualization and classification In Proc. GfKl, Dortmund, Germany, 1992.

ULTSCH, A. Maps for the Visualization of high dimensional Data Spaces In: Yamakawa T (eds) Proceedings of the 4th Workshop on Self-Organizing Maps, 225-230, 2003.

ULTSCH, A. U*-matrix: a tool to visualize clusters in high dimensional data. Technical report, Departement of Mathematics and Computer Science, Philipps-University Marburg, 2003.

ULTSCH, A. ,HERRMANN, L. The architecture of Emergent Self-Organizing Maps to reduce projection errors Proc ESANN, Brugges, pp. 1-6. 2005

ULTSCH, A., MÖRCHEN, F. ESOM-Maps. tools for clustering, visualization, and classification with Emergent SOM. Technical Report 46, CS Department, University Marburg, Germany, 2005.

ZHAO, Y., KARYPIS, G. Criterion Functions for Document Clustering Experiments and analysis. Machine Learning, in press, 2003.