Journal of Next-generation Convergence Information Services Technology Vol.4, No.2, December (2015), pp. 51-64
차세대컨버전스정보서비스기술논문지

Label Prediction of the Unlabeled Mood of a Music Genre Using Semi-Supervised Learning

Babu Kaji Baniya[1], Joonwhoan Lee[2]*

Abstract

Music genre and mood classifications are vital components of the field of multimedia retrieval and computational musicology. There is a growing interest in their development to address the difficulties of music categorization. The proposed method finds the unlabeled mood of a music genre with the help of labeled music mood datasets using different audio feature sets. Semi-supervised learning, which exploits huge amounts of unlabeled data, together with the limited labeled data for learning, has attracted a great deal of research interests. In this paper, we propose diverse audio features to precisely characterize music content. The feature sets belong to four groups: dynamic, rhythmic, spectral, and harmonic. A bin histogram was calculated from each feature to preserve all the important information associated with it. From the extracted audio features, we first tried to find the unlabeled mood of a music genre by using the labeled mood dataset. Harmonic and consistency (local and global) semi-supervised learning algorithms were considered to determine the unknown mood label of a music genre. In the next stage, we also evaluated whether unlabeled genre datasets would influence the mood classification accuracy. The unlabeled datasets were added to the training set in different proportions, so that the overall impact on classification accuracy could be analyzed. We improved the classification accuracy using an unlabeled music genre dataset in training. In the last section, we verified the classification accuracy by adding an unlabeled genre dataset to label mood with only the mood dataset (without adding a genre dataset).

Keyword : Music genre, mood classification, Harmonic, consistency, semi-supervised, learning musicology

1. Introduction

Music genre and mood are categorical labels created by musicians or composers to identify the content (style) of music. Music genre and mood research are growing areas in the field of information retrieval because the results can be applied for practical purposes, such as the efficient organization and categorization of online data collections. It is therefore essential to design automatic grouping tools that provide meaningful and efficient descriptions of music genre and mood classifications. Several well-known approaches have been proposed in this area. Efficient and accurate automatic music

1 Department Computer Science and Engineering, Chonbuk National Univ., Jeonju-si, Jeollabuk-do, 561-756, Korea
e-mail : bk.baniya@chonbuk.ac.kr
2 Department Computer Science and Engineering, Chonbuk National Univ., Jeonju-si, Jeollabuk-do, 561-756, Korea
e-mail: chlee@chonbuk.ac.kr (Corresponding author)

Received(July 29, 2015), Review(August 15, 2015), Accepted(December 02, 2015), Published(December 31, 2015)

ISSN: 2384-101X JNCIST

information processing remains the key issue; therefore, it has attracted the attention of researchers and musicologists. Here, we sought to find the unlabeled mood of a music genre with the help of a labeled mood dataset using semi-supervised learning.

Semi-supervised learning is a learning paradigm concerned with the study of how computers and natural systems such as humans learn in the presence of both labeled and unlabeled data. Generally, learning has been categorized into either the unsupervised paradigm (e.g., clustering, outlier detection), in which all the data are unlabeled, or the supervised paradigm (e.g., classification, regression), in which all the data are labeled. The goal of semi-supervised learning is to understand how combining labeled and unlabeled data may change learning behavior, and to design algorithms that take advantage of such combinations. Semi-supervised learning is of great interest in machine learning and data mining because it can use readily available unlabeled data to improve supervised learning tasks when labeled data are scarce or expensive.

Semi-supervised learning methods attempt to exploit the intrinsic data distribution information disclosed by unlabeled data, which usually is considered to be helpful for learning. To exploit unlabeled data, certain assumptions should be adopted for learning. There are two well-known assumptions in semi-supervised classification, i.e., those of the clustering and manifold approaches[1, 2]. The clustering approach assumes that similar instances are likely to share the same class label, which thus guides the classification boundary passing through the low-density region. The manifold approach assumes that data are distributed on some low-dimensional manifold represented by a Laplacian graph, and that similar instances should share similar classification outputs according to the graph. Almost all off-the-shelf semi-supervised classification methods adopt one or both of those assumptions explicitly or implicitly[3, 4]. For example, large-margin semi-supervised classification methods, such as those used by the transductive support vector machine (TSVM)[5], semi-supervised SVM (S3VM)[6], and their variants[7, 8], follow the cluster assumption. Graph-based semi-supervised classification methods, such as label propagation, graph cuts and Laplacian SVM (LapsSVM), refer to the manifold assumption. Semi-supervised classification methods are applied in diverse applications, e.g., text classification[7], bioinformatics[9], and spam email detection[10].

Here, we take the general problem of learning from labeled to unlabeled data.

Given a point set $X = \{x_1, \ldots\ldots, x_1, x_{l+1}, \ldots\ldots, x_n\}$ and a label set $L = \{1, \ldots\ldots, c\}$, the first l points

have labels $\{y_1,\cdots\cdots y_1\}\in L$ and the remaining points are unlabeled.

The goal is to predict the labels (moods) of the unlabeled points (genres). The performance of an algorithm is measured by the error rate of these unlabeled points. Such a learning problem is often called semi-supervised learning. This type of labeling often requires expensive human labor and a long time, whereas unlabeled data are far easier to obtain, and in such an environment, semi-supervised learning is very useful for many real-world problems. A typical application is web page categorization, in which manually classified web pages are always a very small portion of the entire web, and the number of unlabeled pages is large.

Features commonly exploited for music mood and genre classification can be roughly classified into timbral texture, rhythm, harmony, or their combinations[11, 12]. After the extraction of descriptive features, different pattern recognition algorithms are employed for the classification of the mood. In this proposed method, a large number of audio features were extracted, 59 in total, as shown in Table 1. They were divided into two categories, i.e., low- and high-level, based on the frame size. The frame lengths for the low-level and high-level features were 46 ms and 2 s, respectively, with both having 50% overlap. Each frame-level feature was used to make an eight-bin histogram, and was subsequently used to normalize it. Moreover, this approach makes it possible to preserve important frame-level information obtained from the feature extraction.

We implemented two different semi-supervised learning algorithms, i.e., harmonic and consistency (local and global) algorithms, to find the unlabeled mood of a music genre. For training, (mood-labeled) Coimbra[13] data were taken. Similarly, for testing (unlabeled mood dataset), the GTZAN[14] data set was used to determine the mood label. At first, we calculated the mood label of a music genre using both semi-supervised learning algorithms. Of the two algorithms, the harmonic semi-supervised algorithm yielded more discriminative results than the consistency algorithm. In the second stage, different experiments were performed to measure whether the unlabeled dataset influenced the classification accuracy. In the training phase, 25% of the unlabeled genre data was added to the mood dataset. Similarly, the training unlabeled genre data was added to the mood data in multiples of 25% and trained the system. We obtained better results as the number of unlabeled datasets increased in training. Finally, we tried to compare the classification accuracy of unlabeled genre-added music mood classification results with the results of mood-only classification. We found that unlabeled genre-added

music mood classification yielded better results.

The outline of the paper is as follows. Feature extraction for both music mood and genre is discussed in Section 2. Section 3 describes the details of semi-supervised learning algorithms, i.e., consistency and harmonic. Section 4 describes the data preparation for unlabeled mood determination using a labeled mood dataset. Section 5 briefly presents the experimental results and discussion of the prediction of the mood label of a music genre, and is followed by a conclusion in Section 6.

2. Feature Extraction

[Table 1] Diverse audio features for music mood classification

No.	Category	Feature	Acronyms
1	Dynamic	RMS energy	Histogram
2		Slope	〃
3		Attack	〃
4	Rhythm	Tempo	〃
5	Spectral	Spectral centroid	〃
6		Brightness	〃
7		Spread	〃
8		Rolloff85	〃
9		Rolloff95	〃
10		Spectral entropy	〃
11		Flatness	〃
12		Irregularity	〃
13		Roughness	〃
14		Zerocrossing	〃
15		Spectralflux	〃
16~28		MFCC(1~13)	〃
29~41		DMFCC(1~13)	〃
42~54		DDMFCC(1~13)	〃
55	Harmony	Chromagram peak	〃
56		Chromagram centroid	〃
57		Key clarity	〃
58		Key mode	〃
59		HCDF	〃

Audio feature extraction is the extraction of meaningful information from an audio sample in order to obtain a compact and concise description that is machine-readable. Features are usually selected in the context of specific tasks and domains, and are divided into two categories, i.e., low- and high-level, according to the frame size. The frame lengths for the low-level and high-level features were 46 ms and 2 s, respectively, with both having 50% overlap. Each frame-based feature provides a sequence of frame-wise values for audio data, which allows a set of a small number of meaningful scalar values to be determined. MIR[15] and MA[16] toolboxes are used for feature extraction. The MA toolbox is essentially a subset of the MIR toolbox.

3. Semi-Supervised Learning Algorithms

Two semi-supervised learning algorithms are implemented to determine the unlabeled mood of a music genre.

3.1 Local and Global Consistency

For a given point set $X=\{x_1,.....,x_1,x_{l+1},.....,x_n\}\subset R^m$ and a label set $L=\{1,.....c\}$, the first l points are labeled as , and the remaining points are unlabeled. The goal is to predict the label of the unlabeled points. Consider Ƒ to denote the set of metrics with nonnegative entries. A matrix $F=[F_1^T,.....,F_n^T]^T\in$Ƒ corresponds to classification of dataset X by labeling each point x as a label $y_i=\arg\max_j\leq cF_{ij}$.

We know that F is a vector function $F: X\rightarrow L^c$ which assigns a vector F_i to each point x_i.

Define an $n\times c$ matrix $Y\in$ Ƒ with $Y_{ij}=1$ if is labeled as $y_i=j$, and with $Y_{ij}=0$ otherwise. Clearly, Y is consistent with the initial labels according to the decision rule.

The algorithm is given below:

1. From the affinity matrix W defined by $W_{ij}=\exp(-\ x_i-x_j\ ^2/2\sigma^2)$ if $i\neq j$ and $W_{ii}=0$
2. Construct the matrix $S=D^{-1/2}WD^{-1/2}$ in which D is a diagonal matrix with its (i,i) element equal to the sum of the i-th of W.
3. Iterate $F(t+1)=aSF(t)+(1-a)Y$ until convergence, where is a parameter in (0, 1).

4. Let F^* denote the limit of the sequence $\{F(t)\}$.
5. Label each point x_i as a label $y_i = \arg\max_j \leq cF_{ij}^*$.

This algorithm can be understood intuitively in terms of spreading activation networks[17]. We first define a pairwise relationship W on dataset X with the diagonal elements being zero. We can assume that a graph $G=(V,E)$ is defined on X, where the vertex set V is just X, and the edges E are weighted by W. In the second step, the weight matrix W of G is normalized symmetrically, which is necessary for the convergence of the following iteration. The first two steps are exactly the same as in spectral clustering[18]. During each iteration of the third step, each point receives the information from its neighbors, i.e., first term, and also retains its initial information, i.e., second term. The parameter a specifies the relative amount of information from its neighbors and its initial label information. It is worth mentioning that self-reinforcement is avoided since the diagonal elements of an affinity matrix are set to zero in the first step. Moreover, the information is spread symmetrically, since S is a symmetric matrix. Finally, the label of each unlabeled point is set to be the class from which it has received most information during the iteration process.

(1) Consistency Algorithm-Convergence

Let us consider that the sequence $\{F(t)\}$ converges and $F^*=(1-a)(I-aS)^{-1}Y$. Without loss of generality, suppose $F(0)=Y$. By the iteration equation $F(t+1)=aSF(1-a)Y$ used in the algorithm, we have

$$F(t)=(aS)^{t-1}Y+(1-a)\sum_{i=0}^{t-1}(aS)^iY \tag{1}$$

Since $0<a<1$ and the eigenvalues of S in [-1, 1]

$$\lim_{t\to\infty}(aS)^{t-1}=0 \text{ and } \lim_{i\to\infty}\sum_{i=0}^{t-1}(aS)^i=(I-aS)^{-1} \tag{2}$$

Hence $F^* = \lim_{t\to\infty}F(t)=(1-a)(I-aS)-1Y$, for classification, which is equivalent to

$$F^* = (I-aS)^{-1}Y \tag{3}$$

Here, we can compute F^* directly without iteration. This also proves that the iteration result does not depend on the initial value for the iteration.

(2) Regularization Framework

The cost function associated with F is defined to be

$$Q(F) = \frac{1}{2}\left(\sum_{i,j=1}^{n} W_{ij}\left\|\frac{1}{\sqrt{D_{ij}}}F_i - \frac{1}{\sqrt{D_{ij}}}F_j\right\|^2 + u\|F_i - F_i\|^2\right) \quad (4)$$

where $\mu > 0$ is the regularization parameter. Then, the classifying function is

$$F^* = \arg\min_{F \to f} Q(F) \quad (5)$$

The first term on the right-hand side of the cost function is the smoothness constraint, which indicates a good classifying function that should not change much between nearby points. The second term is the fitting constraint, which means that a good classifying function should not change too much from the initial label assignment. The trade-off between these two competing constraints is captured by a positive parameter u.

The smoothness term essentially splits the function value at each point among the edges attached to it before computing the local changes, and the value assigned to each edge is proportional to its weight.

$\frac{aQ}{aF}\big|_{F=F^*} = F^* - SF^* + u(F^* - Y) = 0$, which can be transformed into

$F^* - \frac{1}{1+u}SF^* - \frac{1}{1+u}Y = 0.$

Let us introduce two new variables, $\alpha = \frac{1}{1+u}$ and $\beta = \frac{1}{1+u}$. Note that $\alpha + \beta = 1$.

Then $(I - aS)F^* = \beta Y$, and the next term is

$F^* = \beta(I - aS)^{-1}Y \quad (6)$

3.2 Harmonic Approach

There are l labeled points $(x_1, y_1), \ldots, (x_1, y_1)$, and u unlabeled points $x_{l+1}, \ldots, x_{l+u}$; generally l u.

Let $n = l + u$ be the total number of data points. To initialize, we assume the labels are binary:

$y \in \{0,1\}$. Again, consider a connected graph $G=(V,E)$ with nodes V corresponding to the n data points, and with nodes $L=\{1,....,l\}$ corresponding to labeled points. Our goal is to assign labels to nodes $U=\{l+1,....,l+u\}$. We assume that an $n \times n$ symmetric weight matrix W on the edges of the graph is given. For example, when , the weight matrix can be expressed as

$$w_{ij} = \exp\left(\sum_{n=1}^{m} \frac{(x_{\in} - x_{jn})^2}{\sigma_n^2}\right) \tag{7}$$

where xin is the n-th component of instance x_i, represented as a vector $x_i \in R^m$ and $\sigma_1,.....,\sigma_m$ are length-scale hyper parameters for each dimension. Thus, nearby points in Euclidean space are assigned large edge weights.

Other weightings are possible, and may be more appropriate when x is discrete. For this purpose, the matrix W fully specifies the data manifold structure.

This approach is to compute a real valued function $f: V \rightarrow R$ on G with certain special properties, and then to assign labels based on f. We constrain f to take values $f(i)=f_l(i)=y_i$ on the labeled data $i=1,...,l$. Intuitively, we need labeled points nearby in the graph to have similar labels. This emphasizes the selection of the quadratic energy function.

$$E(f) = \frac{1}{2}\sum_{i,j} w_{ij}(f(i)-f(j))^2 \tag{8}$$

It is not difficult to show that the minimum energy function $f = \arg\min_{f|L=f1} E(f)$ is harmonic; namely, it satisfies on unlabeled data point U, and is equal f_l to on the labeled data points L. Here, $\triangle$ is the computational Laplacian, given in matrix form as $\triangle = D - W$, where $D = diag(d_i)$ is the diagonal matrix with entries $d = \sum_j w_{ij}$, and $W = [w_{ij}]$ is the weight matrix.

The harmonic property means that the value of f at each unlabeled data point in the average of f at neighboring points:

$$f(i) = \frac{1}{d_j}\sum_{i\ j} w_{ij} f(i) \text{ for } j = l+1,....,l+v \tag{9}$$

which is consistent with our prior notation of the smoothness of f with respect to the graph. The expression is $f = pf$, where $P = D^{-1}W$. Because of the maximum principle of harmonic

functions[1], f is unique and is either a constant or satisfies $0 < f(j) < 1$ for $j \in U$.

To compute the harmonic solution explicitly in terms of matrix operations, we separate the weight matrix W into four blocks after the lth row and column:

$$W = \begin{bmatrix} W_{ll} & W_{lv} \\ W_{vl} & W_{vv} \end{bmatrix} \tag{10}$$

Letting $f = \begin{bmatrix} f_1 \\ f_v \end{bmatrix}$, where f denotes the values on the unlabeled data points, the harmonic solution $\triangle f = 0$ subject to is $f|_{L=} f_l$ given by

$$f_v = (D_{vv} - W_{vv})^{-1} W_{vl} f_l = (I - P_{vv})^{-1} P_{vl} f_l \tag{11}$$

4. Data preparation

Two different datasets were considered for this proposed method, i.e., Coimbra for music mood and GTZAN for genre. The Coimbra dataset was taken from the Informatics and System School of the University of Coimbra in Portugal. There are 903 mp3 audio files in total, which contain five different clusters with several emotional categories, given in [Table 2]. GTZAN consists of 1000 music pieces divided into ten different genres: Classical, Blues, Hip hop, Pop, Rock, Jazz, Reggae, Metal, Disco, and Country. Each class consists of 100 music pieces having a duration of 30 s. Each song in the database was stored as a 22050-Hz, 16-bit, and mono audio file.

[Table 2] Coimbra dataset having several mood adjectives, and the number of corresponding songs

Cluster	Adjective in class	Songs
C1	Passionate, Rousing, Confident, Boisterous, Rowdy	170
C2	Rollicking, cheerful, fun, sweet, amiable/goodnatured	164
C3	Literate, poignant, wisful, bittersweet, autumnal, brooding	215
C4	Humorous, silly, campy, quirky, whimsical, witty, wry	191
C5	Aggressive, fiery, tense/anxious, intense, volatile, visceral	163

5. Experimental results and discussion

First, we determined class labels of the unknown mood of a music genre using labeled music mood

datasets. [Table 3] shows the known mood label of a music genre using the consistency approach of semi-supervised learning. The mood labels of each genre belong to five different classes. The distribution of the mood labels of different genres is shown row-wise in [Table 3]. The rock genre contains diverse moods compared to other genres. Similarly, Metal, Disco, and Reggae also belong to different mood classes. However, they are less diverse than the Rock genre. Other genres belong to two or three moods, only varied in number. We also performed an experiment using the harmonic approach of semi-supervised learning to determine an unknown mood label of a music genre. The experimental result is shown in [Table 4].

Second, we performed an experiment to determine whether the size of the unlabeled genre dataset improves the classification accuracy of the music mood or not. The datasets were partitioned into training and testing sets. The mood data set was subjected to five-fold cross-validation using semi-supervised learning. Moreover, we added 25% of the unlabeled genre dataset without replacement in training and performed the validation. [Fig. 1] shows the data preparation for semi-supervised learning to verify whether unlabeled genre data improved the classification accuracy. Similarly, we performed the experiment by adding 50% of the unlabeled genre dataset to the training set and performing validation. This process was continued by adding 75% and 100% of the unlabeled genre dataset to the training set and performing the validation. The classification accuracy is shown in [Table 5]. The classification accuracy improved continually with the addition of greater proportions of the unlabeled genre dataset. The overall classification accuracy was 57.92% when all of the unlabeled genre dataset was added to the training set. Finally, we also performed five-fold cross-validation only considering the music mood data, assuming that the test data were unlabeled. The classification accuracy was 49.52%. This indicates that a semi-supervised classifier helps to improve the classification accuracy. The percentage of improvement may rely on knowledge of the unlabeled dataset for a given scenario.

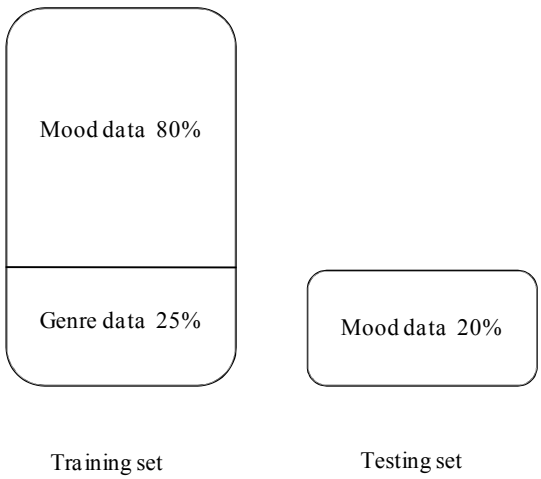
[Table 3] Class label determination using the consistency approach of semi-supervised learning (903 known mood-labeled songs having five different classes, and 1000 unlabeled songs belong to ten different classes)

Genre / Mood	C1	C2	C3	C4	C5
Blues	18.0	22.0	43.0	15.0	2.0
Classical	11.0	5.0	78.0	0.0	6.0
Country	16.0	10.0	68.0	6.0	0.0

Disco	52.0	6.0	18.0	20.0	4.0
Hiphop	23.0	28.0	3.0	45.0	1.0
Jazz	18.0	14.0	61.0	3.0	4.0
Metal	62.0	3.0	8.0	10.0	17.0
Pop	52.0	10.0	19.0	19.0	0.0
Reggae	14.0	15.0	19.0	49.0	3.0
Rock	32.0	8.0	38.0	11.0	11.0

[Table 4] Class label determination using the harmonic approach of semi-supervised learning (903 known mood-labeled songs having five different classes, and 1000 unlabeled songs belong to ten different classes)

Genre / Mood	C1	C2	C3	C4	C5
Blues	17.0	24.0	50.0	7.0	2.0
Classical	7.0	6.0	80.0	0.0	7.0
Country	16.0	11.0	66.0	7.0	0.0
Disco	51.0	12.0	18.0	16.0	3.0
Hiphop	24.0	49.0	3.0	23.0	1.0
Jazz	12.0	18.0	61.0	2.0	7.0
Metal	62.0	2.0	6.0	6.0	21.0
Pop	44.0	24.0	19.0	13.0	0.0
Reggae	13.0	27.0	21.0	35.0	4.0
Rock	33.0	9.0	35.0	11.0	12.0



[Fig. 1] Training and testing dataset preparation for music mood classification using an unlabeled music genre

[Table 5] Classification accuracy of music mood with the addition of increasing proportions of an unlabeled music genre dataset

No. of experiment	% of unlabeled genre dataset mixed with mood	Classification accuracy of music mood
1	25.00	51.12%
2	50.00	53.22%
3	75.00	56.31%
4	100.00	57.92%

6. Conclusion

In this paper, we focused on how to acquire the unlabeled mood of a music genre using semi-supervised learning. Before determining the mood label of a music genre, we extracted different audio features belonging to low- and high-level features. From each feature, an eight-bin histogram was calculated (the number of histograms depends on the number of extracted audio features in the proposed method). Three important conclusions have been drawn from this experiment. At first, the unlabeled mood of music genre was determined using consistency and harmonic approaches to semi-supervised learning. The experiments revealed that harmonic semi-supervised learning gives more discriminative results than the consistency approach, as shown in Tables 3 and 4. In the second phase, we performed several experiments by adding different proportions of the unlabeled genre dataset to the mood dataset in the training stage. The unlabeled genre dataset was added in 25% portions without replacement. The classification accuracy of the music mood was 51.12% with 25% of the unlabeled genre dataset, and continually increased with the addition of unlabeled genre data during the mood training stage. We achieved 57.92% classification accuracy when all the unlabeled genre data were added to the mood dataset in the training stage. Finally, we sought to assess the semi-supervised music mood classification by adding unlabeled genre data to only the music mood classification, assuming that the testing set mood datasets were unlabeled. The classification accuracy of genre-based music mood classification was higher than that of music mood-only classification.

References

[1] X. Zhu, "Semi-supervised learning literature survey", Wisconsin-Madison, University Doctoral Dissertation, **(2008)**.

[2] P. K. Mallapragada, R. Jin, A. K. Jain and Y. Liu, "Semiboost: Boosting for semi-supervised learning", Pattern Analysis and Machine Intelligence, IEEE Transactions, vol. 31, no. 11, **(2009)**, pp. 2000-2014.

[3] Z. -H. Zhou and M. Li, "Semi-supervised learning by disagreement", Knowl. Inf. Syst., vol. 24, no. 3, **(2010)**, pp. 415-439.

[4] O. Chapelle, B. Scholkopf and A. Zien, "Semi-Supervised Learning", MA, USA: MIT Press, Cambridge, **(2006)**.

[5] T. Joachims, "Transductive inference for text classification using support vector machines", in Proc. 16th Int. Conf. Mach. Learn., **(1999)**, pp. 200-209.

[6] G. Fung and O. L. Mangasarian, "Semi-supervised support vector machine for unlabeled data classification", Optim. Methods Softw., vol. 15, no. 1, **(2001)**, pp. 99-105.

[7] R. Collobert, F. Sinz, J. Weston and L. Bottou, "Large scale transductive SVMs", J. Mach. Learn. Res., vol. 7, **(2006)**, pp. 1687-1712.

[8] Y. -F. Li, J. Kwok and Z. -H. Zhou, "Semi-supervised learning using label mean", in Proc. 26th Int. Conf. Mach. Learn., **(2009)**, pp. 633-640.

[9] N. Kasabov and S. Pang, "Transductive support vector machines and applications in bioinformatics for promoter recognition", Neural Inf. Process. Lett. Rev. vol. 3, **(2004)**, pp. 31-38.

[10] B. Pfahringer, "A semi-supervised spam mail detector", In Proceedings of ECML/PKDD Discovery Challenge Workshop, **(2006)**.

[11] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals", IEEE Trans. Speech Audio Process., vol. 10, no. 5, **(2002)**, pp. 293-302.

[12] P. Saari, T. Eerola and O. Lartillot, "Generalizability and Simplicity as Criteria in Feature Selection: Application to Mood Classification in Music", IEEE Trans. On Audio, Speech, and Lang. Process., vol. 19, no. 6, **(2011)**, pp. 1802-1812.

[13] http://mir.dei.uc.pt/downloads.html, **(2015)**.

[14] Marasys, “Data sets”, http://marsysas.info/download/data, **(2015)**.

[15] O. Lartillot and P. Toiviainen, “MIR in Matlab (II): A toolbox for musical feature extraction from audio”, in Proc. Int. Conf. Music Inf. Retrieval, **(2007)**, pp. 127-130 [Online]. Available: http://users.jyu.fi/lartillo/mirtoolbox/

[16] E. Pampalk, “A Matlab toolbox to compute music similarity from audio”, in Proc. Int. Conf. Music Inf. Retrieval, **(2004)**, [Online]. Available:http://www.ofai.at/elias.pampalk/ma/

[17] J. Shrager, T. Hogg and B. A. Huberman, "Observation of phase transitions in spreading activation

networks", Science, vol. 236, **(1987)**, pp. 1092-1094.

[18] A. Y. Ng, M. I. Jordan and Y. Weiss, "On spectral clustering: analysis and an algorithm", In NIPS, 2001. Ng, A. Y., Jordan, M. I., & Weiss, Y. **(2002)**. On spectral clustering: Analysis and an algorithm. Advances in neural information processing systems, vol. 2, 849-856.