Review-Driven Multi-Label Music Style Classification by Exploiting Style Correlations

Guangxiang Zhao[1]*, **Jingjing Xu**[2]*, **Qi Zeng**[2], **Xuancheng Ren**[2], **Xu Sun**[1,2]
[1]Center for Data Science, Beijing Institute of Big Data Research, Peking University
[2]MOE Key Lab of Computational Linguistics, School of EECS, Peking University
{zhaoguangxiang,jingjingxu,pkuzengqi,renxc,xusun}@pku.edu.cn

Abstract

This paper explores a new natural language processing task, review-driven multi-label music style classification. This task requires systems to identify multiple styles of music based on its reviews on websites. The biggest challenge lies in the complicated relations of music styles. To tackle this problem, we propose a novel deep learning approach to automatically learn and exploit style correlations. Experiment results show that our approach achieves large improvements over baselines on the proposed dataset. Furthermore, the visualized analysis shows that our approach performs well in capturing style correlations.[1]

1 Introduction

As music style (e.g., Jazz, Pop, and Rock) is one of the most frequently used labels for music, music style classification is an important task for applications of music recommendation, music information retrieval, etc. Several methods have been proposed for automatic music style classification (Qin and Ma, 2005; Zhou et al., 2006; Wang et al., 2009; Choi et al., 2017). Most of them mainly focus on using audio information to identify styles. Motivated by the fact that a pieces of music could has different styles, several studies (Wang et al., 2009; Oramas et al., 2017) also aim at multi-label music style classification.

Although these methods make promising progress, they are limited in two aspects. First, not all audio data is available in real-world applications because of copyright restrictions, which limits the generalization ability. Second, some of them are based on a strong assumption that a piece of music should be assigned with only one style. Different from these studies, we focus on using easily obtained reviews in conjunction with multi-label music style classification. The motivation comes from the fact that lots of user reviews contain rich style-related information, which can be used for music style classification.

The major challenge of this task lies in the complicated correlations of music styles. For example, Soul Music[2] contains elements of R&B and Jazz. These three labels can be used alone or in combination. Traditional multi-label classification methods may mistake the true label [Soul Music, R&B, Jazz] for the false label [R&B, Jazz]. If well learned, style relations are useful knowledge for improving the performance, e.g., increasing the probability of Soul Music if we find that it is heavily linked with two high probability labels: R&B and Jazz. Therefore, to better exploit style correlations, we propose a novel deep learning approach with two parts: a label-graph based neural network, and a soft training mechanism with correlation based continuous label representation.

Our contributions are listed as follows:

- To the best of our knowledge, this work is the first to explore review-driven multi-label music style classification.
- To learn the relations among music styles, we propose a label-graph based neural network and a soft training mechanism with correlation-based label representation.

2 Related work

This paper is related with music style classification and multi-label classification. In this section, we give a detailed introduction about the related studies.

*Equal Contribution

[1] The code and the dataset are available at https://github.com/lancopku/RMSC

[2] Soul Music is a popular music genre that originated in the United States in the late 1950s and early 1960s. It contains elements of African-American Gospel Music, R&B and Jazz.

Proceedings of NAACL-HLT 2019, pages 2884–2891
Minneapolis, Minnesota, June 2 - June 7, 2019.

Music Title	Mozart: The Great Piano Concertos, Vol.1
Styles	Classical Music, Piano Music
Reviews	*(1) I've been listening to **classical** music all the time.* *(2) Mozart is always good. There is a reason he is ranked in the top 3 of lists of greatest **classical** composers.* *(3) The sound of **piano** brings me peace and relaxation.* *(4) This volume of Mozart concertos is superb.*

Table 1: An illustration of review-driven multi-label music style classification. For easy interpretation, we select a simple and clear example where styles can be easily inferred from reviews. In practice, the correlation between styles and associated reviews is relatively complicated.

2.1 Music Style Classification

Previous work mainly focuses on using audio information to identify music styles. Traditional machine learning algorithms are adopted in these studies, such as Support Vector Machine (SVM) (Xu et al., 2003), Hidden Markov Model (HMM) (Chai and Vercoe, 2001; Pikrakis et al., 2006), and Decision Tree (DT) (Zhou et al., 2006). In addition to audio information, Fell and Sporleder (2014) also propose to classify music by statistical analysis of lyrics. Motivated by the fact that a piece of music could has different styles, several studies (Wang et al., 2009; Oramas et al., 2017) also aim at multi-label music style classification. Different from these studies, we focus on using easily obtained reviews in conjunction with multi-label music style classification.

2.2 Multi-Label Classification

Multi-label classification has been widely applied to diverse problems, including image classification (Qi et al., 2007; Wang et al., 2008), audio classification (Boutell et al., 2004; Sanden and Zhang, 2011), web mining (Kazawa et al., 2004), information retrieval (Zhu et al., 2005; Gopal and Yang, 2010), etc. Compared with the existing multi-label learning methods (Wei et al., 2018; Li et al., 2018b,a; Yang et al., 2018; Lin et al., 2018), our method has the following novelties: a label graph that explicitly models the relations of styles; a soft training mechanism that introduces correlation-based continuous label representation.

3 Review-Driven Multi-Label Music Style Classification

3.1 Task Definition

Given several reviews from a piece of music, this task requires models to predict a set of music styles. Assume that $X = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_i, \dots, \boldsymbol{x}_K\}$ denotes the input K reviews, and $\boldsymbol{x}_i = x_{i,1}, \dots, x_{i,J}$ represents the i^{th} review with J words. The term $Y = \{y_1, y_2, \dots, y_M\}$ denotes the gold set with M labels, and M varies in different samples. The target of review-driven multi-label music style classification is to learn the mapping from input reviews to style labels.

3.2 Dataset

The dataset is collected from a popular Chinese music review website,[3] where registered users are allowed to comment on all released music albums. Each sample includes a music title, a set of human annotated styles, and associated reviews. An example is shown in Table 1.

In order to build a high-quality dataset, we refer to the literature about music styles. We merge similar music styles and delete music styles that violate the music classification list. 22 styles are defined in our dataset.[4] For user reviews, we first delete reviews with too little information by rule-based methods and then select top 40 voted reviews. Music samples with too few reviews are also deleted. The constructed datataset contains over 7.1k samples, 288K reviews, and 3.6M words.

4 Proposed Approach

The proposed approach contains two parts: a label-graph based neural network and a soft training mechanism with continuous label representation. An illustration of the proposed method is shown in Figure 1.

4.1 Label-Graph Based Neural Network

The first layer is a hierarchical attention layer (Yang et al., 2016) that lets the model to pay more or less attention to individual words

[3] https://music.douban.com

[4] Alternative Music, Britpop, Classical Music, Country Music, Dark Wave, Electronic Music, Folk Music, Heavy Metal Music, Hip-Hop, Independent Music, Jazz, J-Pop, New-Age Music, OST, Piano Music, Pop, Post-Punk, Post-Rock, Punk, R&B, Rock, and Soul Music.

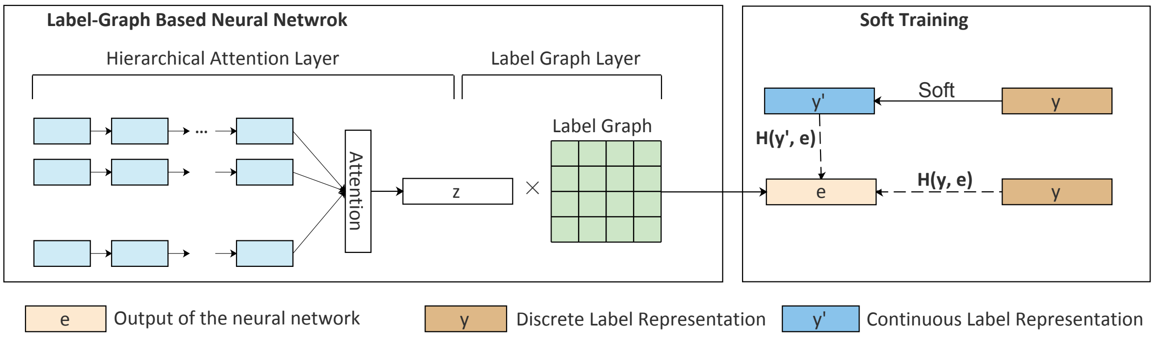
Label-Graph Based Neural Netwrok
Hierarchical Attention Layer
Label Graph Layer
Label Graph
Attention
z
×
Soft Training
y'
Soft
y
H(y', e)
e
H(y, e)
y
e
Output of the neural network
y
Discrete Label Representation
y'
Continuous Label Representation

Figure 1: An illustration of the proposed approach. Left: The label-graph based neural network. Right: The soft training method. The label graph defines the relations of labels. $\boldsymbol{e}$ is the label probability distribution. Soft training means that we combine the continuous label representation $\boldsymbol{y}'$ and the discrete label representation $\boldsymbol{y}$ together to train the model. The hierarchical attention layer is responsible for extracting style-related information. The label graph layer and soft training are used for exploiting label correlations.

and reviews when constructing "raw" label probability distribution $\boldsymbol{z}$.

Label Graph. To explicitly take advantage of the label correlations when classifying music styles, we add a label graph layer to the network. This layer takes $\boldsymbol{z}$ as input and generates a "soft" label probability distribution $\boldsymbol{e}$.

Formally, we denote $\mathcal{G} \in R_{m \times m}$ as the label graph, where m is the number of labels, $\mathcal{G}$ is initialized by an identity matrix. An element $\mathcal{G}[l_i, l_j]$ is a real-value score indicating how likely label l_i and label l_j are related. The graph $\mathcal{G}$ is a part of parameters and can be learned by back-propagation.

Then, given the "raw" label probability distribution $\boldsymbol{z}$ and the label graph $\mathcal{G}$, the output of this layer is:

$$\boldsymbol{e} = \mathrm{sigmoid}(\boldsymbol{z} \cdot \mathcal{G}). \tag{1}$$

The probability of l_j is not only determined by the current classification result, but also determined by other labels probabilities and their correlations to l_j. For example, the probability of a label heavily linked with many high-probability labels will be increased.

4.2 Soft Training

Given a predicted label probability distribution $\boldsymbol{e}$ and a target discrete label representation $\boldsymbol{y}$, the typical loss function is computed as

$$L(\theta) = \mathcal{H}(\boldsymbol{y}, \boldsymbol{e}) = -\sum_{i=1}^{m} y_i \log e_i, \tag{2}$$

where θ denotes all parameters, and m is the number of the labels. The function $\mathcal{H}$ denotes the cross entropy between two distributions.

However, the widely used discrete label representation does not apply to the task of music style classification, because the music styles are not mutually exclusive and highly related to each other. The discrete distribution without label relations makes the model over-distinguish the related labels. Therefore, it is hard for the model to learn the label correlations.

Instead, we propose a soft training method by combining a discrete label representation $\boldsymbol{y}$ with a correlated-based continuous label representation $\boldsymbol{y}'$. The probability gap between two similar labels in $\boldsymbol{y}'$ should not be large.

A straight-forward approach to produce the continuous label representation is to use the label graph matrix $\mathcal{G}$ to transform the discrete representation $\boldsymbol{y}$ into a continuous form:

$$\boldsymbol{y}' = \boldsymbol{y} \cdot \mathcal{G}. \tag{3}$$

Based on the discrete label representation $\boldsymbol{y}$ and continuous label representation $\boldsymbol{y}'$, we define the loss function as

$$Loss(\theta) = \mathcal{H}(\boldsymbol{e}, \boldsymbol{y}) + \mathcal{H}(\boldsymbol{e}, \boldsymbol{y}'), \tag{4}$$

where the loss $\mathcal{H}(\boldsymbol{e}, \boldsymbol{y})$ aims to correctly classify labels, and the loss $\mathcal{H}(\boldsymbol{e}, \boldsymbol{y}')$ aims to avoid the over-distinguishing problem and to better learn label correlations.

5 Experiment

In this section, we evaluate our approach on the proposed dataset. We first introduce evaluation metrics, then show experiment results and give a detailed analysis. The training details can be found at Appendices.

Models	OE(-)	HL(-)	Macro F1(+)	Micro F1(+)
ML-KNN	77.3	0.094	23.6	38.1
Binary Relevance	74.4	0.083	24.7	41.8
Classifier Chains	67.5	0.107	29.9	44.3
Label Powerset	56.2	0.096	37.7	50.3
MLP	71.5	0.081	29.8	45.8
CNN	37.9	0.099	32.5	49.3
LSTM	30.5	0.089	33.0	53.9
HAN	25.9	0.079	52.1	61.0
+Label Graph	23.4	0.077	54.2	62.8
+ Soft Training	**22.6**	**0.074**	**54.4**	**64.5**

Table 2: Comparisons between our approach and the baselines on the test set. OE and HL denote one-error and hamming loss respectively. HAN denotes the hierarchical attention network. "(+)" represents that higher scores are better and "(-)" represents that lower scores are better. It can be seen that the proposed approach significantly outperforms the baselines.

5.1 Evaluation Metrics

Multi-label classification requires different evaluation metrics compared with traditional single-label classification. In this paper, we use the following widely-used evaluation metrics for multi-label classification.

- F1-score: We calculate macro F1 and micro F1, respectively. Macro F1 computes the metric independently for each label and then takes the average as the final score. Micro F1 aggregates the contributions of all labels to compute the average score.
- One-Error: One-error evaluates the fraction of examples whose top-ranked label is not in the gold label set.
- Hamming Loss: Hamming loss counts the fraction of the wrong labels to the total number of labels.

5.2 Baselines

We implement several widely-used multi-label classification methods as baselines, such as ML-KNN (Zhang and Zhou, 2007), Binary Relevance (Tsoumakas et al., 2010), Classifier Chains (Read et al., 2011), Label Powerset (Tsoumakas and Vlahavas, 2007). The details of baselines can be found at Appendices.

5.3 Results

The results on the test set are summarized in Table 2. The proposed approach significantly outperforms the baselines, with micro F1 of 64.5, macro F1 of 54.4, and one-error of 22.6, improving the metrics by 10.6, 21.4, and 7.9 respectively. The improvements are attributed to two parts, a hierarchical attention network and a label correlation mechanism. Only using the hierarchical attention network outperforms the baselines, which shows the effectiveness of hierarchically paying attention to different words and sentences. The greater F1-score is achieved by adding the proposed label graph, which demonstrates that the proposed label graph helps a lot by taking advantage of label correlations.

It can be clearly seen that with the help of soft training, the proposed method achieves the best performance. Especially, the micro F-score is improved from 62.8 to 64.5, and the one-error is reduced from 23.4 to 22.6. With the new loss function, the model not only knows how to distinguish the right labels from the wrong ones, but also can learn the label correlations that are useful knowledge, especially when the input data contains too much style unrelated words for the model to extract all necessary information.

5.4 Visualization Analysis

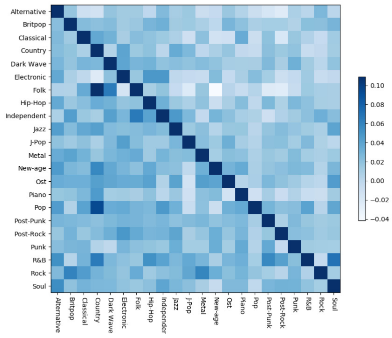

Figure 2: A heatmap of the automatically learned label graph. Deeper color represents closer relation. We can see that some obvious relations are well captured by the model, e.g., "Heavy Metal Music (Metal)" and "Rock", "Country Music (Country)" and "Folk".

Figure 2 shows a whole heatmap of the automatically learned label graph, for convenience of display, we have subtracted the label graph by an identity matrix. We can see from the picture that some obvious music style relations are well captured. For "Country Music", the most related label is "Folk Music". In reality, these two music styles are highly similar and the boundary between them is not well-defined. For three kinds of rock music, "Heavy Metal Music", "Britpop Music", and "Alternative Music", the label graph correctly captures that the most related label for them is “Rock”. For a more complicated relation where “Soul Music” is highly linked with two different labels, “R&B” and “Jazz”, the label graph also correctly capture such relation. These examples demonstrate that the proposed approach performs well in capturing relations among music styles.

5.5 Selected Exmaples

Ground Truth	**Without LCM**	**With LCM**
Britpop[5], Rock	Britpop	Britpop, Rock
Hip-Hop[6], Pop, R&B[7]	Electronic Music, Pop	Pop, R&B
Pop, R&B	Pop, Rock, Britpop	Pop, R&B
Country Music, Folk, Pop	Country Music, Pop	Country Music, Pop, Folk
Classical Music, New-Age Music[8], Piano Music	Piano Music, Classical Music	Piano Music, New-Age Music, Classical Music

Table 3: Examples generated by the methods with and without the label correlation mechanism (LCM). The labels correctly predicted by two methods are shown in blue. The labels correctly predicted by the method with the label correlation mechanism are shown in orange. We can see that the method with the label correlation mechanism classifies music styles more precisely.

For clearer understanding, we compare several examples generated with and without the label correlation mechanism in Table 3. By comparing gold labels and predicted labels generated by different methods, we find that the proposed label correlation mechanism identifies the related styles more precisely. This is mainly attributed to the learned label correlations. For example, the correct prediction in the first example shows that the label correlation mechanism captures the close relation between “Britpop” and “Rock”, which helps the model to generate an appropriate prediction.

5.6 Error Analysis

Although the proposed method has achieved significant improvements, we also notice that there are some failure cases. In this section, we give the detailed error analysis.

First, the proposed method performs worse on the styles with low frequency in the training set. Table 4 compares the performance on the top 5 music styles of highest and lowest frequencies. As we can see, the top 5 fewest music styles get much worse results than top 5 most music styles. This is because the label distribution is highly imbalanced where unpopular music styles have too little training data.

Second, we find that some music items are wrongly classified into the styles that are similar with the gold styles. For example, a sample with a gold set [Country Music] is wrongly classified into [Folk] by the model. The reason is that some music styles share many common elements and only subtly differ from each other. It poses a great challenge for the model to distinguish them. For future work, we would like to research how to effectively address this problem.

Most Styles	**% of Samples**	**Micro F1**
Rock	30.4	75.8
Independent Music	30.0	64.8
Pop	26.2	67.1
Folk Music	21.9	73.7
Electronic Music	13.9	61.8
Fewest Styles	**% of Samples**	**Micro F1**
Jazz	4.3	37.5
Heavy Metal Music	3.9	55.6
Hip-Hop	3.1	7.5
Post-punk	2.5	17.1
Dark Wave	1.3	17.4

Table 4: The performance of the proposed method on the most and fewest styles.

6 Conclusions

In this paper, we focus on classifying multi-label music styles with user reviews. To meet the challenge of label correlations, we propose a label-graph neural network and a soft training mechanism. Experiment results have showed the effectiveness of the proposed approach. The visualization of label graph also shows that our method performs well in capturing label correlations.

Acknowledgments

We thank all reviewers for providing the thoughtful and constructive suggestions. This work was supported in part by National Natural Science Foundation of China (No. 61673028). Email correspondence to Xu Sun.

[5] Britpop is a style of British Rock.

[6] Hip-Hop is a mainstream Pop style.

[7] Rhythm and Blues, often abbreviated as R&B, is a genre of popular music.

[8] New-Age Music is a genre of music intended to create artistic inspiration, relaxation, and optimism. It is used by listeners for yoga, massage, and meditation.

References

Matthew R. Boutell, Jiebo Luo, Xipeng Shen, and Christopher M. Brown. 2004. Learning multi-label scene classification. *Pattern Recognition*, 37(9):1757–1771.

Wei Chai and Barry Vercoe. 2001. Folk music classification using hidden markov models. In *Proceedings of International Conference on Artificial Intelligence*, volume 6.

Keunwoo Choi, György Fazekas, Mark Sandler, and Kyunghyun Cho. 2017. Convolutional recurrent neural networks for music classification. In *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*, pages 2392–2396.

Michael Fell and Caroline Sporleder. 2014. Lyrics-based analysis and classification of music. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 620–631, Dublin, Ireland. Dublin City University and Association for Computational Linguistics.

Siddharth Gopal and Yiming Yang. 2010. Multilabel classification with meta-level features. In *Proceeding of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2010, Geneva, Switzerland, July 19-23, 2010*, pages 315–322.

Geoffrey E. Hinton, Oriol Vinyals, and Jeffrey Dean. 2015. Distilling the knowledge in a neural network. *CoRR*, abs/1503.02531.

Hideto Kazawa, Tomonori Izumitani, Hirotoshi Taira, and Eisaku Maeda. 2004. Maximal margin labeling for multi-topic text categorization. In *Advances in Neural Information Processing Systems 17 [Neural Information Processing Systems, NIPS 2004, December 13-18, 2004, Vancouver, British Columbia, Canada]*, pages 649–656.

Diederik P. Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980.

Wei Li, Xuancheng Ren, Damai Dai, Yunfang Wu, Houfeng Wang, and Xu Sun. 2018a. Sememe prediction: Learning semantic knowledge from unstructured textual wiki descriptions. *CoRR*, abs/1808.05437.

Wei Li, Zheng Yang, and Xu Sun. 2018b. Exploration on generating traditional chinese medicine prescription from symptoms with an end-to-end method. *CoRR*, abs/1801.09030.

Junyang Lin, Qi Su, Pengcheng Yang, Shuming Ma, and Xu Sun. 2018. Semantic-unit-based dilated convolution for multi-label text classification. *arXiv preprint arXiv:1808.08561*.

Tianyu Liu, Kexiang Wang, Baobao Chang, and Zhifang Sui. 2017. A soft-label method for noise-tolerant distantly supervised relation extraction. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1790–1795, Copenhagen, Denmark. Association for Computational Linguistics.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

Sergio Oramas, Oriol Nieto, Francesco Barbieri, and Xavier Serra. 2017. Multi-label music genre classification from audio, text, and images using deep features. *arXiv preprint arXiv:1707.04916*.

Aggelos Pikrakis, Sergios Theodoridis, and Dimitris Kamarotos. 2006. Classification of musical patterns using variable duration hidden markov models. *IEEE Trans. Audio, Speech & Language Processing*, 14(5):1795–1807.

Guo-Jun Qi, Xian-Sheng Hua, Yong Rui, Jinhui Tang, Tao Mei, and Hong-Jiang Zhang. 2007. Correlative multi-label video annotation. In *Proceedings of the 15th International Conference on Multimedia 2007, Augsburg, Germany, September 24-29, 2007*, pages 17–26.

Dan Qin and GZ Ma. 2005. Music style identification system based on mining technology. *Computer Engineering and Design*, 26:3094–3096.

Jesse Read, Bernhard Pfahringer, Geoff Holmes, and Eibe Frank. 2011. Classifier chains for multi-label classification. *Machine learning*, 85(3):333.

Chris Sanden and John Z. Zhang. 2011. Enhancing multi-label music genre classification through ensemble techniques. In *Proceeding of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2011, Beijing, China, July 25-29, 2011*, pages 705–714.

Xu Sun, Bingzhen Wei, Xuancheng Ren, and Shuming Ma. 2017. Label embedding network: Learning label representation for soft training of deep networks. *CoRR*, abs/1710.10393.

Grigorios Tsoumakas, Ioannis Katakis, and Ioannis P. Vlahavas. 2010. Mining multi-label data. In *Data Mining and Knowledge Discovery Handbook, 2nd ed.*, pages 667–685.

Grigorios Tsoumakas and Ioannis P. Vlahavas. 2007. Random k -labelsets: An ensemble method for multilabel classification. In *Machine Learning: ECML 2007, 18th European Conference on Machine Learning, Warsaw, Poland, September 17-21, 2007, Proceedings*, pages 406–417.

Fei Wang, Xin Wang, Bo Shao, Tao Li, and Mitsunori Ogihara. 2009. Tag integrated multi-label music style classification with hypergraph. In *Proceedings of the 10th International Society for Music Information Retrieval Conference, ISMIR 2009, Kobe International Conference Center, Kobe, Japan, October 26-30, 2009*, pages 363–368.

Mei Wang, Xiangdong Zhou, and Tat-Seng Chua. 2008. Automatic image annotation via local multi-label classification. In *Proceedings of the 7th ACM International Conference on Image and Video Retrieval, CIVR 2008, Niagara Falls, Canada, July 7-9, 2008*, pages 17–26.

Bingzhen Wei, Xuancheng Ren, Xu Sun, Yi Zhang, Xiaoyan Cai, and Qi Su. 2018. Regularizing output distribution of abstractive chinese social media text summarization for improved semantic consistency. *CoRR*, abs/1805.04033.

Changsheng Xu, Namunu C Maddage, Xi Shao, Fang Cao, and Qi Tian. 2003. Musical genre classification using support vector machines. In *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, volume 5, pages V–429.

Pengcheng Yang, Xu Sun, Wei Li, Shuming Ma, Wei Wu, and Houfeng Wang. 2018. Sgm: Sequence generation model for multi-label classification. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3915–3926, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical attention networks for document classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1480–1489, San Diego, California. Association for Computational Linguistics.

Min-Ling Zhang and Zhi-Hua Zhou. 2007. ML-KNN: A lazy learning approach to multi-label learning. *Pattern Recognition*, 40(7):2038–2048.

Yatong Zhou, Taiyi Zhang, and Jiancheng Sun. 2006. Music style classification with a novel bayesian model. In *Advanced Data Mining and Applications, Second International Conference, ADMA 2006, Xi'an, China, August 14-16, 2006, Proceedings*, pages 150–156.

Shenghuo Zhu, Xiang Ji, Wei Xu, and Yihong Gong. 2005. Multi-labelled classification using maximum entropy method. In *SIGIR 2005: Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Salvador, Brazil, August 15-19, 2005*, pages 274–281.

A Appendices

A.1 Training Details

In the proposed method and baselines, we use skip gram (Mikolov et al., 2013) to get pre-trained word embeddings of reviews. The Jieba toolkit is used to split sentences into words. To help with the training of the label graph, we use soft target mechanism on the continuous label representations (Hinton et al., 2015; Sun et al., 2017; Liu et al., 2017) and add the negative of the l2 loss of the difference between the label graph and the identity matrix to the loss function. For evaluation, we introduce a hyper-parameter p. If the probability of a style is greater than p, we consider it as one of the final music styles. We tune hyper-parameters based on the performance on the validation set. We set p to 0.2, hidden size to 128, embedding size to 128, vocabulary size to 135K, learning rate to 0.001, and batch size to 128. The optimizer is Adam (Kingma and Ba, 2014) and the maximum training epoch is set to 100. We choose parameters with the best Micro F1 scores on the validation set and then use the selected parameters on the test set.

A.2 Baselines

We implement the following widely-used multi-label classification methods for comparison. Their inputs are the music representations which are produced by averaging review representations. The review representation is obtained by averaging word embeddings.

- ML-KNN (Zhang and Zhou, 2007): It is a multi-label learning approach derived from the traditional K-Nearest Neighbor (KNN) algorithm.

- Binary Relevance (Tsoumakas et al., 2010): It decomposes a multi-label learning task into a number of independent binary learning tasks. It learns several single binary models without considering the dependencies among labels.

- Classifier Chains (Read et al., 2011): It takes label dependencies into account and keeps the computational efficiency of the binary relevance method.

- Label Powerset (Tsoumakas and Vlahavas, 2007): All classes assigned to an example are combined into a new and unique class in this method.

- MLP: It feeds the music representations into a multi-layer perceptron, and generates the probability of music styles through a sigmoid layer.

Different from the above baselines, the following two methods only take word embeddings as inputs. Similar to MLP, they produce label probability distribution by a sigmoid function.

- CNN: It includes two layers of CNN which has multiple convolution kernels.

- LSTM: It includes two layers of LSTM, which processes words and sentences separately to get the music representations.