



# 基于文本挖掘与神经网络的音乐风格分类建模方法

张键锋,王 劲

(广东省电信规划设计院有限公司 广州 510630)

**摘 要:** 针对人工区分音乐风格会造成音乐风格关系不清以致混乱和某些歌曲难以人工划分其风格等问题,以歌曲的歌词数据为基础,分析歌曲所表达的情感,以划分其归属。运用机器学习算法的 BP 神经网络,建立一个音乐风格预测模型,对模型进行了合理的理论证明和推导。实验选用 MATLAB 作为建模工具,根据算法自身特点确定训练参数。随机从数据集中抽取 10% 的记录作为测试。该方法的结果显示,理论结果与数据模拟结果比较吻合,准确率达到 80%。

**关键词:** 神经网络;文本挖掘;网络爬虫;音乐分类

**doi:** 10.11959/j.issn.1000-0801.2015179

## A Classification Method of Music Style Based on Text Mining and Neural Network

Zhang Jianfeng, Wang Jin

(Guangdong Planning and Designing Institute of Telecommunications Co., Ltd., Guangzhou 510630, China)

**Abstract:** In terms of the confusion of music style ambiguous relationships caused by artificial classification and the problem that some songs cannot be classified into the proper style, the emotion of songs based on their lyrics to classify their belongings was analyzed. BP neural network was used to establish a music style forecast model and then a reasonable proof of the theory and derivation was made. This experiment use MATLAB as a modeling tool and determine the training parameters according to its characteristics, and then extract 10% from dataset to test. This approach shows that theoretical results and data simulation results agree well with the accuracy rate reached about 80%.

**Key words:** neural network, text mining, Web crawler, music classification

### 1 引言

随着互联网的发展,流行音乐的主要传播媒介从传统的电台和唱片逐渐过渡到网络下载和网络电台等。网络电台需要根据收听者的已知喜好,自动推荐并播放其他音乐。由于每个人喜好的音乐可能横跨若干种风格,区别甚大,需要分别对待。这就需要探讨如何区分音乐风格的问

题。在流行音乐中,传统的风格概念包括流行(pop)、乡村(country)、爵士(jazz)、摇滚(rock)、节奏布鲁斯(R&B)、新世纪(new age)等若干大类,它们分别可以细分成许多小类,有些小类甚至可以做更进一步的细分。而每首歌曲只能靠人工赋予风格标签。这样的做法有许多不足:有的类别之间关系不清楚,易造成混乱;有的类别过度粗略或精细;有的类别标签没有得到公认;有的音乐归属则存在争

收稿日期:2015-04-03;修回日期:2015-07-10

论文引用格式:张键锋,王劲. 基于文本挖掘与神经网络的音乐风格分类建模方法. 电信科学, 2015179

Zhang J F, Wang J. A classification method of music style based on text mining and neural network. Telecommunications Science, 2015179

议或者难以划归。因此,需要建立合理的数学模型,对流行音乐的风格给出合理的分类方法。

2 现状分析

给定一首歌曲根据其歌词或者音乐来预测它所属的风格是一个不容易的问题。歌词以文本作为载体,表达了歌曲的情感、主题、风格等,本文从音乐心情出发,利用文本挖掘的机器学习算法对歌词进行挖掘,推算新歌曲所属的音乐心情。BP(back propagation)神经网络能学习和存储大量输入—输出模式映射关系,而无需事先了解描述这种映射关系的数学方程。只要能提供足够多的样本模式供网络进行学习训练,它便能完成由  $n$  维输入空间到  $m$  维输出空间的非线性映射。本文利用已经标注好的歌曲、利用 BP 神经网络训练一个分类模型,预测新的歌曲其所属的音乐心情风格。

2.1 模型的假设

假设百度音乐网站 (<http://music.baidu.com/tag>) 所给的数据真实可靠;假设百度音乐网站的音乐已经代表大部分音乐的类型;假设百度音乐网站的音乐关于情感的分类符合大众的分类要求;假设百度音乐网站上每一首歌曲属于且仅属于一种情感<sup>[1]</sup>;假设百度音乐网站上每一首歌曲所属情感在建模阶段不会发生变化。

2.2 数据处理

用网页正文抽取技术从网页抽取给定类别下每首歌曲的歌词,在计算机对自然语言文本分类前,必须将自然语言转换成计算机可以识别的数值。首先对歌词进行预处理(如分词、去掉停用词等),结合词性规则抽取表示情感与主题的单词特征,用向量空间模型表示每首歌曲,单词特征的词频作为向量空间模型的属性特征。文本预处理逻辑过程如图 1 所示,具体介绍如下。

(1)分析百度音乐网页模板,网页中存放歌词的代码段为:`<div id="lyricCont" class="lyric-content lyric-hidden"> </div>`。

(2)利用 jsoup<sup>[2]</sup>开源类库,基于正则表达式抽取歌词正文,把每一首歌曲的歌词存进数据库。

(3)对歌词进行分词<sup>[3]</sup>标注词性,去掉停用词等噪音,用词频作为单词的权值,以词向量表示歌曲,作为后续 BP 神经网络的输入。

利用词性语义结合规则<sup>[4]</sup>进行特征选择后,向量的维数为 2 010 维,由于歌词中词语的重复数比较大,所以以词频作为向量属性的权值,并且取值量化到 0~5,语法规则如式(1):

$$R=\{<r_1,s_1>,<r_2,s_2>,\cdots,<r_m,s_m>\}$$
 (1)

其中  $<r_m,s_m>$  是一个二元组,表示当词  $word_{k_j}$  符合规则  $r_m$  时,得到  $s_m$  权重,即:

$$N(word)=\begin{cases} s_m, & word \text{ 符合规则 } r_m \\ \text{不抽取}, & word \text{ 不符合规则 } r_m \end{cases}$$
 (2)

其中,当  $s_m \geq 5$  时,  $s_m$  的取值均为 5。

2.3 定义与符号说明

定义与符号说明见表 1。

表 1 定义与符号说明

| 符号            | 说明                         |
|---------------|----------------------------|
| $X$           | 模型输入层向量,即各个歌名、歌词信息分词得到的统计值 |
| $h_i$         | 模型隐含层输入向量                  |
| $h_o$         | 模型隐含层输出向量                  |
| $W_{h_i}$     | 模型元素间连接权重,即每个倾向词的倾向度权重     |
| $Y_o$         | 模型隐含层输出数值,即最后的文本倾向值        |
| $y_i$         | 模型输出层输入向量                  |
| $y_o$         | 模型输出层输出向量                  |
| $d_o$         | 模型期望输出向量                   |
| $M$           | 模型最大学习次数                   |
| $e$           | 模型误差函数                     |
| $\varepsilon$ | 模型训练精度值                    |

3 模型建立与求解

3.1 BP 神经网络的应用背景

由于对流行音乐风格的定性分析需要复杂的乐理专

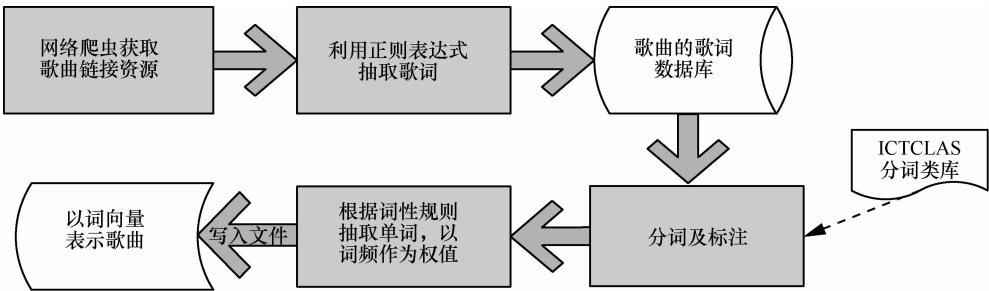


图 1 文本预处理逻辑



业知识,若通过分析音乐特性、提取数字音频特征等方法实现,需要耗费大量实验时间进行样本采集,同时专业知识的缺乏可能会在特征选择的过程中出现偏差,造成大量音乐风格的错误分类。一方面,歌曲信息的分类在信息处理领域日趋火热,而歌词文本中有着丰富内容,通过把歌名信息和歌词相结合解决歌曲分类的问题有一定的可行性;另一方面,神经网络具有较强的学习性、自动调整性以及非线性映射特性,已广泛应用于文本特征提取、模式识别、工业过程控制等方面<sup>[5]</sup>。在实际应用中,BP神经网络凭借非线性映射功能强大、泛化能力强等特点,成为目前应用最多的一种神经网络。

BP神经网络由神经元相互连接的一种按误差逆传播算法训练的多层前馈网络,通过反向传播,对连接神经元的权值不断调整,使模型输出结果不断逼近期望目标。BP神经网络通常由3层神经元组成,分别为输入层、隐含层和输出层。在函数应用上,本文采用自适应lrBP的梯度递减训练函数、基于梯度下降法的学习函数、均方误差规范化性能函数进行模型训练,并把S型函数作为激励函数,实现流行音乐风格的预测分类。

### 3.2 BP神经网络的模型和结构

BP神经网络模型<sup>[6]</sup>的总体结构与神经网络类似,简化后如图2所示。

$(X_1, X_2, \dots, X_n)$ 为输入层,是各个歌名、歌词信息分词得到的统计值。 $H_{on}$ 是隐含层输出, $(W_{h_{o1}}, W_{h_{o2}}, \dots, W_{h_{on}})$ 是每个倾向词的倾向度权重, $Y_o$ 是输出层输出,即最后的文本倾向值。 $f(\cdot)$ 为激活函数。BP神经网络流程如图3所示。

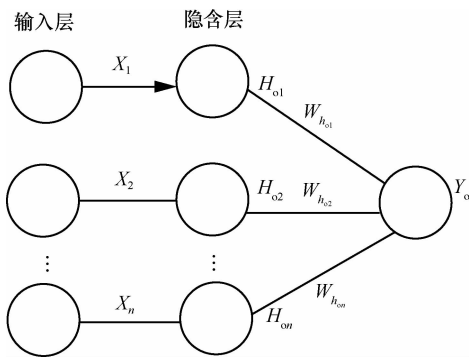


图2 BP神经网络模型结构

### 3.3 BP神经网络的数学论证与公式推导

设输入层有 $N$ 个神经元,隐含层有 $P$ 个神经元,输出层有 $Q$ 个神经元。

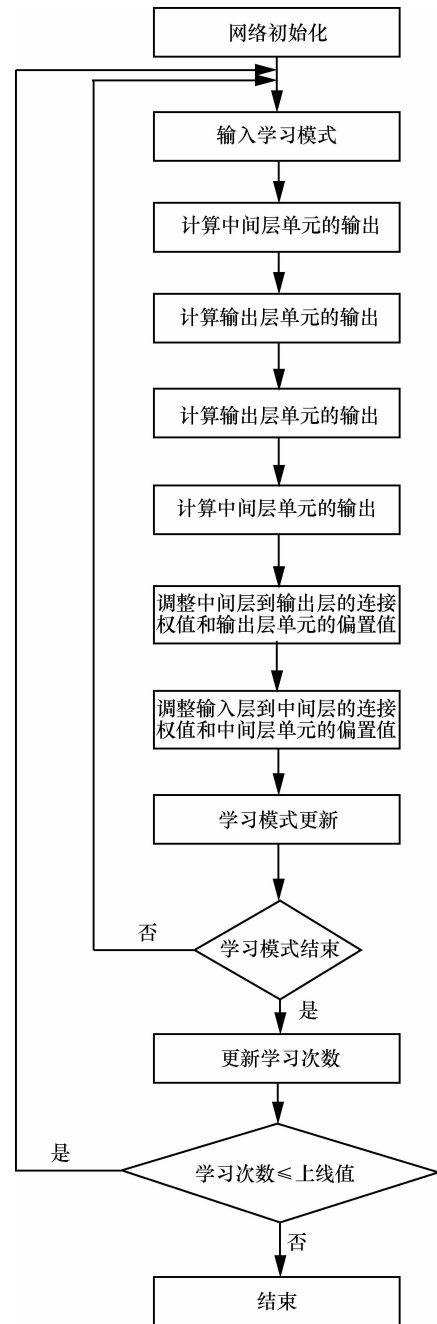


图3 BP神经网络流程

#### 3.3.1 变量定义

输入向量  $x=(x_1, x_2, \dots, x_N)$ , 隐含层输入向量  $h_i=(h_{i1}, h_{i2}, \dots, h_{ip})$ ; 隐含层输出向量  $h_o=(h_{o1}, h_{o2}, \dots, h_{op})$ ; 输出层输入向量  $y_i=(y_{i1}, y_{i2}, \dots, y_{iq})$ ; 输出层输出向量  $y_o=(y_{o1}, y_{o2}, \dots, y_{oq})$ ; 期望输出向量  $d_o=(d_1, d_2, \dots, d_q)$ ; 输入层与中间层的连接权值为  $w_{pq}$ ; 隐含层与输出层的连接权值为  $w_{pq}$ ; 隐含层各神经元的阈值为  $b_p$ ; 输出层各神经元的阈值为  $b_q$ ; 样本数据个数  $k=1, 2, \dots, m$ ; 激活函数为  $f(\cdot)$ ; 误差函数如式(3)所示。

$$e = \frac{1}{2} \sum_{q=1}^Q (d_{oq}(k) - y_{oq}(k))^2 \quad (3)$$

### 3.3.2 运算步骤

(1) 网络初始化。给各连接权值分别赋一个区间 $(-1, 1)$ 的随机数, 设定误差函数 $e$ , 给定计算精度值 $\varepsilon$ 和最大学习次数 $M$ 。

(2) 随机选取第 $k$ 个输入样本及对应期望输出, 如式(4)和式(5)所示:

$$x_n(k) = (x_1(k), x_2(k), \dots, x_N(k)) \quad (4)$$

$$d_{oq}(k) = (d_{o1}(k), d_{o2}(k), \dots, d_{oQ}(k)) \quad (5)$$

(3) 计算隐含层和输出层各神经元的输入和输出, 如式(6)~式(9)所示:

$$h_{ip}(k) = \sum_{n=1}^N w_{np} x_n(k) - b_p, \quad p=1, 2, \dots, P \quad (6)$$

$$h_{op}(k) = f(h_{ip}(k)), \quad p=1, 2, \dots, P \quad (7)$$

$$y_{iq}(k) = \sum_{p=1}^P w_{pq} h_{op}(k) - b_q, \quad q=1, 2, \dots, Q \quad (8)$$

$$y_{oq}(k) = f(y_{iq}(k)), \quad q=1, 2, \dots, Q \quad (9)$$

(4) 利用网络期望输出和实际输出, 计算误差函数对输出层的各神经元的偏导数 $\delta_{oq}(k)$ :

$$\frac{\partial e}{\partial w_{pq}} = \frac{\partial e}{\partial y_{iq}} \frac{\partial y_{iq}}{\partial w_{pq}} \quad (10)$$

$$\frac{\partial y_{iq}(k)}{\partial w_{pq}} = \frac{\partial (\sum_{p=1}^P w_{pq} h_{op}(k) - b_q)}{\partial w_{pq}} = h_{op}(k) \quad (11)$$

$$\frac{\partial e}{\partial y_{iq}} = \frac{\partial (\frac{1}{2} \sum_{q=1}^Q (d_{oq}(k) - y_{oq}(k))^2)}{\partial y_{iq}} = -(d_{oq}(k) - y_{oq}(k)) y'_{oq}(k) \quad (12)$$

(5) 利用隐含层到输出层的连接权值、输出层的 $\delta_{oq}(k)$ 和隐含层的输出计算误差函数对隐含层各神经元的偏导数 $\delta_{ip}(k)$ :

$$\frac{\partial e}{\partial w_{pq}} = \frac{\partial e}{\partial y_{iq}} \frac{\partial y_{iq}}{\partial w_{pq}} = -\delta_{oq}(k) h_{op}(k) \quad (13)$$

$$\frac{\partial e}{\partial h_{ip}(k)} = \frac{\partial (\frac{1}{2} \sum_{q=1}^Q ((d_{oq}(k) - f(\sum_{p=1}^P w_{pq} h_{op}(k) - b_q))^2)}{\partial h_{ip}(k)} \frac{\partial h_{op}(k)}{\partial h_{ip}(k)} \quad (14)$$

(6) 利用输出层各神经元的 $\delta_{oq}(k)$ 和隐含层各神经元的输出来修正连接权值 $w_{pq}(k)$ :

$$\Delta w_{pq}(k) = -\mu \frac{\partial e}{\partial w_{pq}} = \mu \delta_{oq}(k) h_{op}(k) \quad (15)$$

$$w_{pq}^{N+1} = w_{pq}^N + \eta \delta_{oq}(k) h_{op}(k) \quad (16)$$

(7) 利用隐含层各神经元的 $\delta_{ip}(k)$ 和输入层各神经元的输入修正连接权值:

$$\Delta w_{np}(k) = -\mu \frac{\partial e}{\partial w_{np}} = -\mu \frac{\partial e}{\partial h_{ip}(k)} \frac{\partial h_{ip}(k)}{\partial w_{np}} = \delta_{ip}(k) x_n(k) \quad (17)$$

$$w_{np}^{N+1} = w_{np}^N + \eta \delta_{ip}(k) x_n(k) \quad (18)$$

(8) 计算全局误差:

$$E = \frac{1}{2m} \sum_{k=1}^m \sum_{q=1}^Q (d_{oq}(k) - y_{oq}(k))^2 \quad (19)$$

(9) 判断网络误差是否满足要求。当误差达到预设精度或学习次数大于设定的最大次数, 则结束算法; 否则, 选取下一个学习样本及对应的期望输出, 返回到步骤(3), 进入下一轮学习。

### 3.4 BP 神经网络的建立

(1) 模型参数选择

本次建模共得到 12 414 组数据, 为方便进行统计计算, 把数据按类别进行分块, 每个数据块随机抽取 500 组数据, 共有  $500 \times 8 = 4\,000$  组。在抽样得到的 4 000 组记录中, 随机选择 90%, 即 3 600 组作为训练数据集, 剩下的 10%, 即 400 组作为检验数据集。使用 MATLAB 作为建模工具。选择的 BP 神经网络结构为 2 010-N-1, 其中 2 010 表示词向量的输入维数, 1 表示输出的维数 (只有 1 维: 歌曲类别),  $N$  为神经元隐含的个数。由于训练网络的输入层比较复杂, 考虑到模型不能过于复杂, 经过多次模型训练, 隐含层最终采用两层结构, 每层各为 30 个神经元, 即  $N = 30 \times 30 = 900$ 。

模型采用自适应 lrBP 的梯度递减训练函数 (traind), 基于梯度下降法的学习函数 (learnf)、均方误差规范化性能函数 (msef) 进行模型训练; 同大多数应用一样, S 型函数 (tansig) 作为模型的激励函数, 实现流行音乐风格的预测和分类。

BP 神经网络的学习速度对预测性能具有重大影响。学习速度过大可能节省训练时间和成本, 但极易造成网络不收敛, 影响预测和分类的准确性; 学习速度过小, 学习过程缓慢, 为保证训练精确度, 需要额外增加训练次数。考虑到训练维数较多, 最终决定学习速度为 0.01, 训练次数为 1 000 次。

(2) 建模结果输出

训练好的 BP 神经网络预测得到的结果是小数值, 可以方便地与真实结果进行对比。由于歌曲类别已规定为 1~8 的整数, 在进行分类准确度统计前, 在 MATLAB 中使用 if...else 的选择结构语句, 对预测的数值进行四舍五入的修正操作。



BP神经网络训练结束后得到训练误差(MSE)如图4所示。

可以看出,训练随着迭代次数的增加,训练误差不断减小;训练结束时达到最小,为0.044 291。利用训练得到的BP神经网络模型对歌曲类别进行预测,通过对预测值的修正,得到样本的整体分类准确率达到81%。由于数据集较大,图5显示了部分检验类别的预测结果及其与真实结果的对比。

如图5所示,从整体上看,预测的分类结果与真实结果相差不大,但在个别点上会存在较大误差,由于预测异常点占数据集整体比例很小,因此不会对模型的预测准确度造成太大影响。

## 4 结束语

利用本文训练得到的BP神经网络模型对歌曲类别进

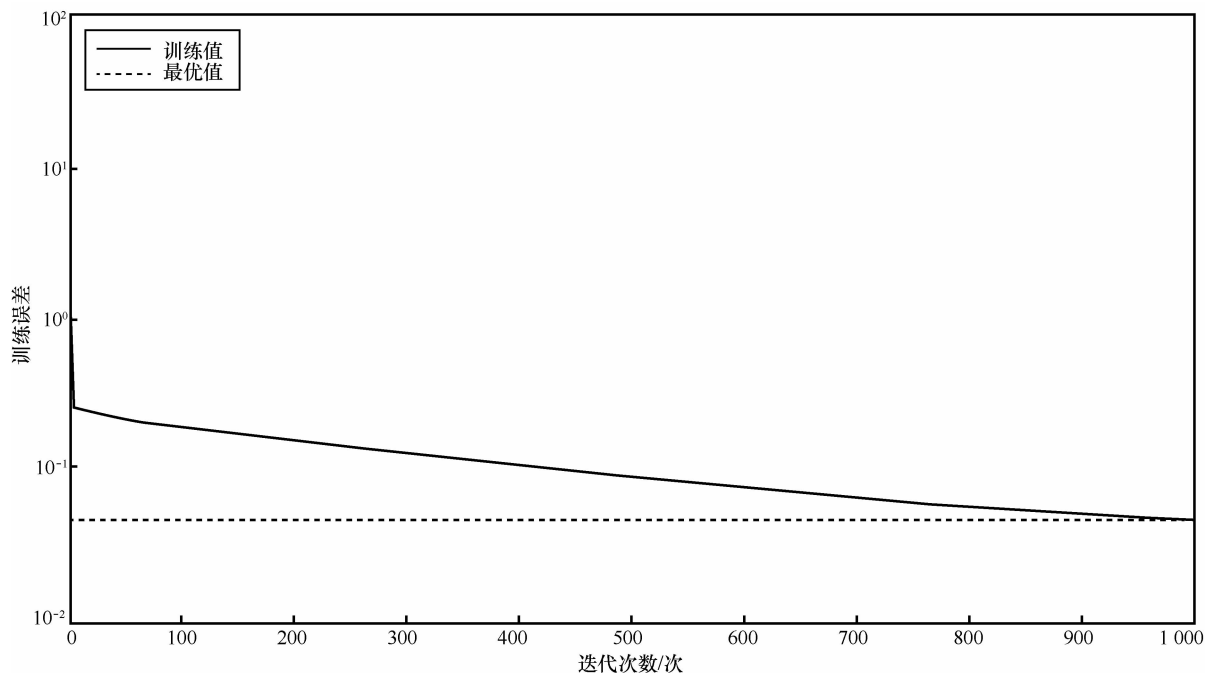


图4 BP神经网络误差训练

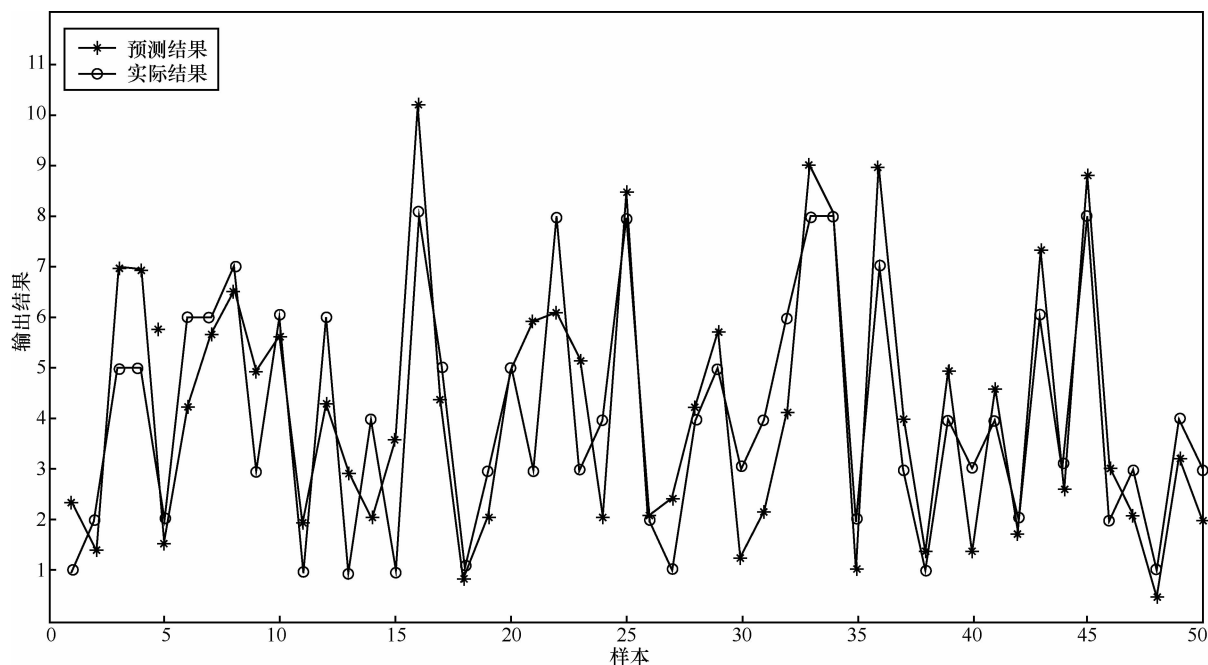


图5 BP神经网络预测结果与实际结果对比

行预测,通过对预测值的修正,得到样本的整体分类准确率为81%,证明本文中的模型对音乐风格的分类具有重要的应用价值。由于从网络采集到的语料有限,网络新词多变,使用情感语料库覆盖有限,使得神经网络的学习能力有一定限制,误判一些歌曲的音乐心情。如果有适合的网络情感新词发现算法,不断更新情感语料库,则可大大提高音乐心情的分类效果,一种新的识别情感生词算法,有待深入研究。目前标注的语料库普遍规模较小,涉及的领域和主题较少,大大降低了分类的效果。因此组织具有专业音乐知识的专家和计算机信息处理知识的专家共同制定规则,对领域和主题丰富的大规模语料库进行标注,建设一个标准的音乐风格语料库是具有广阔的发展前景。另外,此模型是以音乐心情的歌词为特征表示歌曲的,如果适当改变抽取规则针对主题去抽取关键词,则可以推广到以主题风格分类的音乐识别模型中。

### 参考文献

- 1 jsoup. <http://jsoup.org/>, 2015
- 2 ICTCLAS. <http://www.ictclas.org/>, 2015
- 3 潘文彬.基于情感词词典的中文句子情感倾向分析(硕士学位论文). 北京邮电大学, 2011
- 4 Lin P Y, Lin Z J, Kuang B Q, *et al.* A short chinese text incremental clustering algorithm based on weighted semantics and naive Bayes. *Journal of Computational Information Systems*, 2012, 8(10): 4257~4268
- 5 Hagan M T, Demuth H B, Beale M. *Neural Network Design*. PWS Pub Co, 2002
- 6 Mitchell T. *Machine Learning*. Boston: MIP Press, 2002
- 7 Baeza-Yates R, Ribeiro-Neto B. *Modern information retrieval*. Facet Publishing, 1999, 20(19): 305~306
- 8 Hu Y F, Koren Y, Volinsky C. Collaborative filtering for implicit feedback datasets. *Proceedings of IEEE 8th Conference on Data Mining, Pisa, Italy*, 2008: 263~272
- 9 Li B, Yang Q, Xue X. Transfer learning for collaborative filtering via a rating-matrix generative model. *Proceedings of the 26th International Conference on Machine Learning, Montreal, Canada*, 2009: 617~624
- 10 Singh A P, Gordon G J. Relational learning via collective matrix factorization. *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Las Vegas, Nevada, USA*, 2008: 650~658
- 11 Berkovsky S, Kuflik T, Ricci F. Mediation of user models for enhanced personalization in recommender systems. *User Modeling and User-Adapted Interaction*, 2008, 18(3): 345~352
- 12 Zhang Y, Koren J. Efficient bayesian hierarchical user modeling for recommendation systems. *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, New York, USA*, 2007: 47~54
- 13 Lee D D, Seung H S. Algorithms for non-negative matrix factorization. *Advances in Neural Information Processing Systems*, 2001, 13(3): 556~562
- 14 Luo S, Rong J. Flexible mixture model for collaborative filtering. *Proceedings of the 20th International Conference on Machine Learning, Washington DC, USA*, 2003(3): 704~711
- 15 张键锋,男,广东省电信规划设计院有限公司工程师,主要从事无线、数据网络规划设计工作。
- 16 王劲,男,广东省电信规划设计院有限公司工程师,主要从事无线、数据网络规划设计工作。
- 17 高升,男,博士,北京邮电大学副教授,主要研究方向为机器学习、信息推荐和社会网络分析等。
- 18 任思婷,女,北京邮电大学在读,主要研究方向为机器学习与信息推荐。
- 19 郭军,男,博士,北京邮电大学教授,主要研究方向为复杂网络与网络搜索等。

### [作者简介]



张键锋,男,广东省电信规划设计院有限公司工程师,主要从事无线、数据网络规划设计工作。

王劲,男,广东省电信规划设计院有限公司工程师,主要从事无线、数据网络规划设计工作。

2015179-6

(上接 2015188-5)

- 11 Li B, Yang Q, Xue X. Transfer learning for collaborative filtering via a rating-matrix generative model. *Proceedings of the 26th International Conference on Machine Learning, Montreal, Canada*, 2009: 617~624
- 12 Singh A P, Gordon G J. Relational learning via collective matrix factorization. *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Las Vegas, Nevada, USA*, 2008: 650~658
- 13 Berkovsky S, Kuflik T, Ricci F. Mediation of user models for enhanced personalization in recommender systems. *User Modeling and User-Adapted Interaction*, 2008, 18(3): 345~352
- 14 Zhang Y, Koren J. Efficient bayesian hierarchical user modeling for recommendation systems. *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, New York, USA*, 2007: 47~54
- 15 Lee D D, Seung H S. Algorithms for non-negative matrix factorization. *Advances in Neural Information Processing Systems*, 2001, 13(3): 556~562
- 16 Luo S, Rong J. Flexible mixture model for collaborative filtering. *Proceedings of the 20th International Conference on Machine Learning, Washington DC, USA*, 2003(3): 704~711
- 17 Baeza-Yates R, Ribeiro-Neto B. *Modern information retrieval*. Facet Publishing, 1999, 20(19): 305~306
- 18 Hu Y F, Koren Y, Volinsky C. Collaborative filtering for implicit feedback datasets. *Proceedings of IEEE 8th Conference on Data Mining, Pisa, Italy*, 2008: 263~272

### [作者简介]



高升,男,博士,北京邮电大学副教授,主要研究方向为机器学习、信息推荐和社会网络分析等。

任思婷,女,北京邮电大学在读,主要研究方向为机器学习与信息推荐。

郭军,男,博士,北京邮电大学教授,主要研究方向为复杂网络与网络搜索等。

2015188-6