基于 FCM 的文本迁移学习算法*

田宏泽，古　平
（重庆大学 计算机学院，重庆 400044）

摘　要：传统的机器学习方法是在训练数据和测试数据分布一致的前提下进行的，但在一些现实世界中的应用中，训练数据和测试数据是来自不同领域的。在不考虑数据分布的情况下，传统的机器学习算法可能会失效。针对这一问题，提出一种基于模糊 C-均值（FCM）的文本迁移学习算法。通过简单分类器对测试样本分类，利用自然邻算法构建样本初始模糊隶属度，再利用 FCM 算法通过迭代更新样本模糊隶属度，修正样本标签，对样本孤立点进行处理，得到最终的分类结果。实验结果表明，该算法具有较好的正确率，有效地解决了在训练数据和测试数据分布不一致的情况下的文本分类问题。

关键词：模糊 C-均值；自然邻；迁移学习；孤立点
中图分类号：TP301.6　　**文献标志码**：A　　**文章编号**：1001-3695(2018)07-1978-04
doi:10.3969/j.issn.1001-3695.2018.07.012

Text classification algorithm for transfer learning based on FCM

Tian Hongze, Gu Ping
(*College of Computer Science, Chongqing University, Chongqing* 400044, *China*)

Abstract: The traditional machine learning methods work under the assumption that the training data and test data are in the same distribution. However, in some real-world applications, training data and test data come from different domains. The traditional learning methods may fail without considering the shift of the data distribution. This paper proposed a text classification algorithm for transfer learning based on fuzzy C-means to solve this problem. First, it classified the test data with a simple classifier. Second, it initialized the fuzzy membership degree of each data based on natural nearest neighbor algorithm. Then, it updated the fuzzy membership degree based on FCM and refined the labels of test data. Finally, it classified the outliers in test data. In the experiment, it used 20 newsgroups data set and SRAA data set to evaluate the algorithm. The results indicate that the proposed algorithm make a great improvement in classification accuracy.

Key words: fuzzy C-means; natural nearest neighbor; transfer learning; outliers

0　引言

在传统的机器学习框架中，学习任务主要是在充足的训练数据情况下，训练一个良好的分类模型，接着利用该模型对测试数据进行分类。整个流程是建立在训练数据集和测试数据集在相同数据分布的假设下的。另外，监督学习需要有很多良好标记的训练数据集来得到较好的分类器。但是在一些特定领域，具有良好标注的训练数据很难得到。例如，在高速发展的互联网行业，会出现很多新颖的网络用语，尽管在个人微博或者其他网络媒介中这类文本有很多，但是针对网络用语，具有情感类别标注的文本却很少，这对于训练情感分类器来说是很棘手的。针对这些问题，迁移学习应运而生。

迁移学习的目标是利用某一领域学习的信息帮助提高目标领域的学习问题。在这一新兴领域中，有许多研究者作出了杰出的贡献。文献[1]中提出一种新颖的方法 kernel mean matching(KMM)来估计每一个实例在源领域的权重，进而估算出源领域实例的分布特点来应用于目标领域。另外，深度神经网络作为特征提取的方法，也被运用于迁移学习中。文献[2]提出一种新型模型，利用多层 Hubel-Wiesel 结构[3]来解决多重迁移学习任务。此外，最近几年出现了许多基于贝叶斯方法的迁移学习算法。例如，针对文档分类问题，文献[4]提出一种新型朴素贝叶斯迁移学习算法(NBTL)。

本文主要关注文本分类中的迁移学习问题。在很多情景中，训练数据和测试数据分别属于不同类别，但是它们共享许多相同的特征。例如，在标注的训练数据集 D_L 中，有两个子类数据集 D_L_POS 和 D_L_NEG，同时还有未标注的测试数据集 D_U。在测试数据集 D_U 中，许多文本的特征会与 D_L_POS 和 D_L_NEG 相关。在此情况下，迁移学习的任务就是利用训练数据集 D_L 中提取的信息，帮助改善测试集 D_U 的分类效果。

1　相关工作

本文提出一种基于模糊集修正的迁移学习方法。该方法利用简单的分类器得到测试数据的预测类标。该分类器是从训练数据集 D_L 训练得到，在测试数据集 D_U 进行测试。在这一步骤中，文本特征提取是在训练数据集上得到的。然而在测试集上的预测结果往往不会很好，因为测试集和训练集来自不同的数据源。本文提出的方法旨在改善关键特征，使其更接近于训练集和测试集共享的特征。

首先，在迁移过程中，计算每一个测试样本的最近邻居。本文使用了一种改进的自然最近邻居(3N)算法[5]。与传统的 K-近邻算法相比，3N 算法不需要设置额外的参数，并且对孤立点是十分敏感的。然后，根据测试集初始预测类标和最近邻居集，计算每一个测试文本相对于每一个类别的模糊隶属度。随后，利用模糊集理论知识对文本模糊隶属度进行修正，直到收敛为止。在每一次迭代过程中，关键特征集都会被更新，使其更接近测试数据集的特征分布。最后，单独考虑孤立点对分类的影响。

本文提出的算法可以表示为一个具有知识迁移功能的分类器和一个从源特征空间到目标特征空间的特征提取器。该方法主要贡献在以下几个方面：a)利用改进的自然最近邻方

收稿日期：2017-03-08；**修回日期**：2017-04-24　　**基金项目**：中央高校基本科研业务费专项基金资助项目(106112013CDJZR180014)
作者简介：田宏泽(1991-)，男，内蒙古包头人，硕士研究生，主要研究方向为机器学习、数据挖掘(thzcqu@163.com)；古平(1976-)，男，重庆人，副教授，主要研究方向为机器学习、模式分类.

法计算文本相似性,减少了参数设置,提高了效率;b)引入模糊 C-均值(FCM)方法,对测试文本模糊隶属度进行收敛计算,使其真正达到在测试集分类效率的提高;c)考虑到自然最近邻算法对孤立点是敏感的,本文将孤立点单独考虑,进行分类并提高了分类效果。

2　基本框架

2.1　问题描述

本文通过形式化定义来阐释相关问题。首先,假设 X_s 为源实例空间,在这一空间中,每一个实例 $x_s \in X_s$ 被表示成一个特征向量$(y_s(1),\cdots, y_s(n))$,其中 $y_s(i) \in Y_s$,Y_s 是源特征空间。假设 X_t 是目标实例空间,每一个实例 $x_t \in X_t$ 被表示成一个特征向量$(y_t(1),\cdots, y_t(n))$,其中 $y_t(i) \in Y_t$,Y_t 是目标特征空间。定义训练数据集和每一个训练实例的类标集合 L_s,$L_s = \{(x_s(i),c(i))\}$,$i=1,\cdots,n$,其中 $x_s(i) \in X_s$,$c(i) \in C = \{1,\cdots,|C|\}$ 表示实例 $x_s(i)$ 的真正类标。未被标注的测试集表示为 U,$D_U = \{x_u(i)\}$,$i=1,\cdots,k$,其中 $x_u(i) \in X_t$,k 则表示测试集实例的数量。值得注意的是,x_s 和 x_t、x_u 实例来自不同的实例空间,因而特征向量也不尽相同。但是因为源实例空间和目标实例空间具有相关性,所以不同实例会共享许多相同的特征。最终,该问题可以简化为对源实例空间进行学习,训练分类器,使其在目标实例空间中的分类效果得到提升。

对于源特征空间和目标特征空间的特征 $Y = Y_s \cup Y_t$,可以归纳为三个部分,其中:Y_+ 表示只在源特征空间中出现的特征;Y_{mix} 表示既在源特征空间中出现又在目标特征空间中出现的特征;Y_- 表示只在目标特征空间中出现的特征。本文主要关注 Y_{mix} 和 Y_-[6],因为这两部分会影响到 x_u 的分类准确性。

2.2　方法概述

在 bridge refinement 算法中[7],算法主要部分受到 PageRank算法的启发,该方法假设给定一个实例 d,该实例的类标 C 的条件概率是不变的,不管实例空间怎么变化。可以用如下的等式表示:$P_{D_u}(c|d) = P_{D_{\text{mix}}}(c|d) = P_{D_L}(c|d)$,其中 $P(d)$ 是变化的。如果一个实例既出现在训练数据集中又出现在测试数据集中,那么它们的类标应该是相同的,这一等式就基于这一事实。更进一步,如果在测试数据集中,对于越相似的实例,它们具有相同类标的概率就越高。这种情况下,实例之间具有相互强化的关系,因此可以凭借这一点可以来校正预测类标。本文不仅考虑到这一假设,同时运用了特征重提取和实例重表示的方法来校正预测类标。

该方法的描述如下:

a)从源实例空间 X_s 中提取出特征集合 Y_s。根据特征集合 Y_s 以及实例与真实类标,训练得到简单分类器。利用该分类器对目标数据集进行分类。

b)利用改进的自然最近邻算法计算目标数据集中每个实例的最近邻居。根据最近邻居信息计算每个目标数据集中实例的初始模糊隶属度。

c)受到 FCM 算法的启发,本文利用该方法来校正目标数据集实例的模糊隶属度。在每一次迭代校正中,特征集合 Y_s 会根据模糊隶属度的变化而更新。所有实例被重新表示为新的特征向量。

d)由于 3N 算法对于孤立点很敏感,根据新提取出的特征集合以及训练好的简单分类器对孤立点进行分类。

整个算法如算法 1 所示。

算法 1　基于模糊聚类的迁移学习算法

输入:文档集 S 与 T(S 表示源数据集,T 表示目标数据集,S 与 T 中文档均以 bag-of-words 形式表示),真实类标集 C(集合 C 中每个类标表示 S 数据集中某个文档的真实类标),相似函数 $\text{sim}(d_i,d_j)$(表示 S 与 T 文档集中每个文档之间的相似度),简单分类器 F 和特征提取器 Q。

输出:文档集 T 中每个文档的预测类标。

start
1　利用特征提取器 Q 以及 S 与 C 的信息,提取特征集合 L;
2　利用 L、S 与 C 的信息,训练分类器 F;
3　利用 F 计算 T 中每个文档初始类标集 C';
4　for each document $d_i \in T$ do
　　计算 d_i 的自然最近邻居 $NN_{nb(d_i)}(d_i)$
　end for
5　根据自然最近邻信息,将 T 分为孤立点集 O 和非孤立点集 N;
6　初始化 N 中每个文本的模糊隶属度 $u(i,j)$;
7　根据 FCM 算法校正 $u(i,j)$;
8　根据如下等式,计算校正类标集 C_{nr}:
$$C_{nr}(i) = \arg\max_j(u(i,j))$$
9　根据 $F^{(k)}$ 重新训练分类器 F;
10　利用 F 计算 O 中文档的校正类标集 C_{or};
return C_{nr}、C_{or}
end

2.2.1　自然最近邻算法流程

越相似的样本具有相同类标的概率越高,其在迁移过程中增强学习的效果越明显。本文拟采取一种新的无参近邻发现算法,即自然最近邻居(nature nearest neighbor,3N)算法,在不指定参数 K 的情况下,利用数据的自然分布特征,主动发现每个样本的可靠近邻集。在数据集中,孤立点具有更少的最近邻居和更低的能量,而中心点具有更多的最近邻居和更高能量。算法自动为每一个数据点设置不同的 k 值,这样能更准确地反映出数据分布情况。原始算法迭代结束条件十分苛刻,本文放宽了这一条件,在不严重影响结果的前提下能够更快地收敛到结果。

算法 2　改进的自然最近邻算法(3N)

输入:文档集合 D(每个文档以 bag-of-words 形式表示),D 中文档间相似度函数 $\text{sim}(d_i,d_j)$。

输出:每个文档的自然最近邻数 $nb(d_i)$,每个文档的自然最近邻集合 $NN_{nb(d_i)}(d_i)$。

start
1　initialize
2　　$r=1$, flag $=0$
3　　for each $d_i \in D$, $nb(d_i)=0$, $NN_r(d_i)=\varnothing$
4　end
5　while flag $=0$
6　　if all$(nb(d_i)) \neq 0$ then flag $=1$
7　　else
8　　　for each d_k in document collection D do
9　　　　计算 d_k 的第 r 个最近邻居:$NN_r(d_k)$
10　　　　$nb(NN_r(d_k)) = nb(NN_r(d_k)) + 1$
11　　　　$NN_r(d_k) = NN_r(d_k) \cup \{nn_r(d_k)\}$
12　　　end
13　　　$r = r+1$
14　　end if
15　end while
16　$\text{sup}_k = r-1$
return $nb(d_i)$, $NN_{nb(d_i)}(d_i)$
end

对于算法 2 来说,如果已知数据点之间的距离信息,则算法时间复杂度为 $O(N \times \text{sup}_k)$,其中 N 为数据点个数,最坏时间复杂度为 $O(N \times (N-1))$。

将每个文档实例转换成一个特征向量,利用 VSM 模型计算每个文档间的相似度 $\text{sim}(d_i,d_j)$;然后利用 3N 算法来计算每个实例的最近邻居,$NN_i = \{x(i_1), x(i_2),\cdots, x(i_k)\}$。自然最近邻算法如算法 2 所示。根据目标数据集最近邻居信息,通过如下等式来初始化目标数据集文档的模糊隶属度:

$$U_{x_t(c_j)} = \frac{\sum_{x_t(k) \in NN_i} \text{sign}(x_t(k))}{\text{sum}(NN_i)}$$

其中:sum(NN_i)是 NN_i 的总数;sign 函数如下所示。值得注意的是,如果 sum(NN_i) = 0,那么这些点被叫做孤立点。

$$\text{sign}(x_t(i)) = \begin{cases} 1 & \text{如果 } c_t(i) = 1 \\ 0 & \text{如果 } c_t(i) = 0 \end{cases}$$

2.2.2 模糊隶属度校正过程

本文受到 FCM 算法的启发,这一算法是由 Dunn[8] 在 1973 年提出,并由 Bezdek[9] 在 1981 年进行改进。FCM 是一种聚类算法,允许一个数据点属于两个或多个类簇。该方法主要基于最小化目标函数:

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^m \times D(i,j)^2 \quad 1 \leqslant m < +\infty \tag{1}$$

其中:m 是任意大于 1 的实数;u_{ij}是文档 x_i 属于类簇 j 的模糊隶属度;$D(i,j)^2$ 是文档 x_i 与第 j 个聚类中心的距离描述。

根据拉格朗日乘子法,对目标函数式(1)进行组合优化,可以得到模糊隶属度 u_{ij}和聚类中心 c_j 的更新函数:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{d(i,j)}{d(i,k)}\right)^{\frac{2}{m-1}}} \tag{2}$$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m \times x_i}{\sum_{i=1}^{N} u_{ij}^m} \tag{3}$$

当 $\max_{ij}\{|u_{ij}^{(k+1)} - u_{ij}^{(k)}|\} < \varepsilon$ 时,迭代停止。其中:ε 是迭代终止条件,$0 < \varepsilon < 1$。整个过程收敛于局部最小值或鞍点。在传统的 FCM 聚类过程中,初始化是随机的,迭代的收敛速度很慢,聚类结果并不总是好的。然而通过计算自然最近邻步骤,得到了初始模糊隶属度,使得该算法收敛速度变快,并对收敛结果进行了改进。

改进的校正算法如算法 3 所示。该方法将特征集的更新添加到了 FCM 算法迭代中。由于每次迭代过程中样本类标会发生变化,那么关键特征集特征提取也会发生变化。第 5 ~ 6 行计算特征集,并将实例表示成 bag-of-words 形式;第 7 ~ 9 行计算并更新了聚类中心、模糊隶属度以及预测的类标;迭代停止条件在第 10、11 行中进行更新;最后一行返回结果。在改进的 FCM 算法中,特征提取器可以利用各种文本特征提取的方法,本文选取了互信息方法[10] 作为特征提取器。

算法 3 改进的模糊 C-均值算法

输入:文档集 D(文档以 bag-of-words 表示),文档初始模糊隶属度 $U^{(0)} = \{u_{ij}\}$, $i \in D$, $j \in$ 类标集 C,特征提取器 φ,迭代终止参数 ε,迭代次数 K。

输出:模糊隶属度 $U^{(m)}$,校正的特征集 $F^{(m)}$。

```
start
1   initialize
2     k = 0, e = infinite
3   end
4   while k < K and e > ε do
5     利用特征提取器 φ 计算特征集 F^(k);
6     将每个文档重新表示成特征向量;
7     根据式(3)更新聚类中心向量 C^(k);
8     根据式(2)更新模糊隶属度矩阵 U^(k);
9     更新每个文档的类标
10      e = |U^(k) - U^(k-1)|
11      k = k + 1
12  end while
return U^(k), F^(k)
end
```

下一个步骤是将孤立点分类到更准确的类别中。特征集合是前边步骤中得到的更接近于 Y_{mix} 和 Y_- 的特征集合。因此,该特征集合可以训练出一个相对强大的分类器。对于离群点来说,初始预测类标是根据最初分类器得到的,而这一分类器的特征集合是源数据集得到的,预测结果并不准确。然而新的分类器是在混合特征集的基础上训练得到的,结果将更准确。

在时间复杂度上,对于模糊校正过程,每一次迭代需要计算每个样本与聚类中心的距离。若数据维度为 p,则计算距离时间复杂度为 $O(p)$。对于特征选择方法,假设其时间复杂度为 $O(T)$,数据集聚类数为 c 类,数据数为 n,迭代次数为 k,则整个模糊隶属度校正的时间复杂度为 $O(k(p+T)nc)$。由于特征选择算法复杂度远远大于距离计算时间复杂度,则最终校正过程时间复杂度为 $O(kTnc)$。

3 实验

本章主要验证基于迁移学习的模糊集校正算法的有效性。首先,将该算法应用于二分类问题;然后,构建四分类数据集,对该算法进行多分类验证。

3.1 数据集和数据预处理

本文主要应用的数据集是 20 newsgroups 和 SRAA。这两个数据集并不是为迁移学习设计的,所以需要改变数据分布,将源数据集和目标数据集数据分布分开。本文准备了两组不同的修正数据集,一个是二分类数据集(表 1),另一个是四分类数据集(表 2)。

表 1 二分类数据集

数据源	编号	类别	正样本	负样本	数量
20 newsgroups	1	训练集	comp. graphics	rec. motorcycles	2 000
		测试集	comp. windows. x	soc. religion. christian	1 997
	2	训练集	sci. electronics	talk. politics. misc	2 000
		测试集	sci. crypt	talk. politics. mideast	2 000
	3	训练集	rec. autos	talk. politics. guns	2 000
		测试集	rec. sport. baseball	talk. politics. mideast	2 000
	4	训练集	rec. sport. baseball	sci. med	2 000
		测试集	rec. sport. hockey	sci. space	1 999
SRAA	5	训练集	realauto	simauto	2824
		测试集	realaviation	simaviation	2173

表 2 四分类数据集

数据源	编号	类别	类一	类二	类三	类四
20 newsgroups	6	训练集	sci. electronics	talk. politics. misc	rec. autos	comp. graphics
		测试集	sci. crypt	talk. religion. misc	rec. motorcycles	comp. windows. x

原始数据需要作一些预处理,包括将所有文本处理成小写;移除非法字符;利用 WordNet[11] 工具包进行词干处理、停用词处理。对于特征提取器,本文选择了互信息方法。然后,将每个文本转换成 bag-of-words 形式。

3.2 实验效果

为了保证该算法在修正数据集上的有效性,本文设计了不同的对比实验:a)利用朴素贝叶斯分类器处理二分类问题,数据选用二分类数据集;b)针对四分类问题,本文对比了朴素贝叶斯分类器和支持向量机的效果。

在二分类实验中,首先,从训练样本中学习简单分类器,运用该分类器对测试样本集进行测试,得到初始类标。在训练样本中,正样本设置标签为 1,负样本设置标签为 0。这样,对于训练集和测试集中的样本,都有类标与之一一对应。接着,应用本文提出的算法,就会得到最终的结果。

图 1 显示了在二分类问题中,不同方法在各个数据集上的误差率。柱状图左边、中间和右边分别是没有应用本文提出的算法直接进行分类的结果、采用本文算法而没有考虑孤立点的分类结果和考虑孤立点情况下运用本文算法进行分类的结果。在没有考虑孤立点的情况下,应用本文算法进行分类时,错误率减少最多的是测试集 2 和 4,分别达到 87.5% 和 87.3%。在所有数据集上,平均错误率减少 74.7%。而在考虑孤立点的情况下,应用本文算法进行分类时,错误率减少最多的是测试集 1 和 2,分别为 96.5% 和 91.2%。在所有数据集上,平均错

误率减少 78.7%。由实验结果可以看出,本文提出的基于特征层面校正类标的迁移学习算法有效地提高了分类的准确率。

本文提出的算法主要包括两个参数,一个是最大迭代次数 K,另一个是迭代终止参数 ε。二分类中各数据集迭代次数如图 2 所示。从图 2 可以看出,在 K 取值适度的条件下,算法对 K 值并不敏感,每一组实验中,该算法在迭代 5 ~ 16 次间都得到收敛。所以,本文根据经验值将 K 设置为 20。而迭代终止条件参数 ε 用于算法 2 中校正模糊隶属度。理论上,ε 越小,得到的结果越准确。所以,根据经验本文选取 0.05 作为 ε 的值。

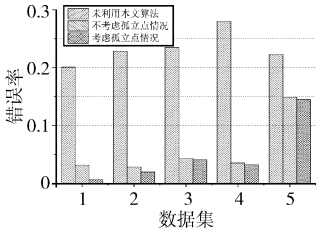

图 1　二分类数据集上的错误率

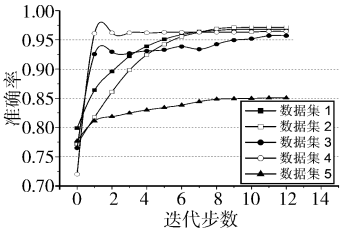

图 2　二分类中各数据集迭代次数

如图 3,4 所示,利用本文方法,在四分类数据集(表 2)上,分别对朴素贝叶斯和支持向量机的效果作了对比。对于朴素贝叶斯分类方法来说,未运用本文方法进行分类的正确率为 76.7%,而运用本文方法进行分类时,在第 3 次迭代后,结果达到稳定,分类的正确率为 83.1%,考虑孤立点后,分类的正确率为 85.5%;而对于支持向量机分类方法来说,未运用本文方法进行分类的正确率为 71.0%,而运用本文方法进行分类时,在第 6 次迭代后,结果达到稳定,分类的正确率为 94.8%,考虑孤立点后,分类的正确率为 95.3%。由结果看出,在处理该数据集时,不同的分类器会对本文方法产生影响,支持向量机效果明显好于朴素贝叶斯分类器。

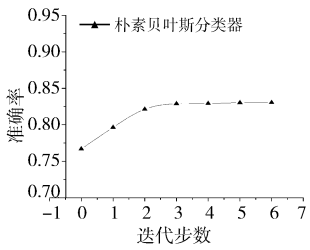

图 3　朴素贝叶斯作为分类器的四分类效果

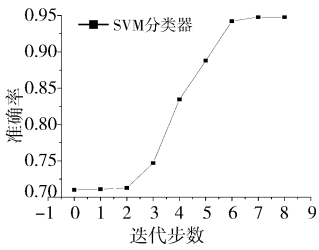

图 4　支持向量机作为分类器的四分类效果

4　结束语

本文提出了一种在训练集和测试集具有不同分布条件下通过迁移学习提高分类效果的算法。该方法利用最近自然邻居算法来计算测试集中样本的相似性,进而引用 FCM 算法对测试集类标进行校正。为了验证该算法的有效性,本文设计了两类实验,分别对二分类问题和四分类问题进行实验。结果表明,在处理二分类问题和多分类问题上,分类效果均有明显提高;同时,在四分类问题中,本文通过实验对比了朴素贝叶斯算法和支持向量机算法作为普通分类器对该算法的影响,结果表明,在处理四分类问题中,支持向量机表现的效果要优于朴素贝叶斯分类器。

在未来工作中,希望能够将该算法应用于网络大规模文本分类中,通过并行处理降低处理时间。同时,在迁移学习过程中,不仅仅考虑文本之间的相关性,而是将数据集扩展到图像数据与文本数据之间的联系,达到图像与文本之间迁移学习的目的。另外,在特征提取过程中,可以对比不同的方法对算法的影响。

参考文献:

[1] Huang Jiayuan, Smola A J, Gretton A, *et al.* Correcting sample selection bias by unlabeled data[C]//Proc of the 19th International Conference on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2007: 601-608.

[2] Ahmed A, Yu Kai, Xu Wei, *et al.* Training hierarchical feed-forward visual recognition models using transfer learning from pseudo-tasks [C]//Proc of the 10th European Conference on Computer Vision. Berlin: Springer, 2008: 69-82.

[3] Hubel D H, Wiesel T N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex [J]. Journal of Physiology, 1962, 160(1): 106-154.

[4] Dai Wenyuan, Xue Guiyong, Yang Qiang, *et al.* Transferring naive Bayes classifiers for text classification [C]//Proc of the 22nd National Conference on Artificial Intelligence. 2007:540-545.

[5] 张莹. 基于自然最近邻居的分类算法研究[D]. 重庆: 重庆大学, 2015.

[6] Yu Jianfei, Jiang Jing. A hassle-free unsupervised domain adaptation method using instance similarity features [C]//Proc of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. 2015: 168-173.

[7] Xing Dikan, Dai Weyuan, Xue Guiyong, *et al.* Bridged refinement for transfer learning [C]//Proc of the 11th European Conference on Principles and Practice of Knowledge Discovery in Databases. Berlin: Springer-Verlag, 2007:324-335.

[8] Dunn J C. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters [J]. Journal of Cybernetics, 1973, 3(3): 32-57.

[9] Bezdek J C. Pattern recognition with fuzzy objective function algoritms [M]. Norwell, MA: Kluwer Academic Publishers, 1981.

[10] Peng Hanchuan, Long Fuhui, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2005, 27(8): 1226-1238.

[11] Miller G A, Beckwith R, Fellbaum C D, *et al.* WordNet: an online lexical database [J]. International Journal of Lexicography, 1990,3(4): 235-244.

(上接第 1964 页)

[12] Zhang Jiang, Nie Guojiang, Li Shouju. Study on mathematical model of confined transportation of coal auger and simulation[J]. Meitan Xuebao/Journal of the China Coal Society,2013,38(S1):231-235.

[13] 马骊, 李阳, 樊锁海. 改进人工鱼群算法在外汇预测和投资组合中的应用[J]. 系统工程理论与实践, 2015, 35(5): 1256-1266.

[14] 龚波, 曾飞艳. 一种改进人工鱼群算法对 BP 神经网络的优化研究[J]. 湖南科技大学学报: 自然科学版, 2016, 31(1): 86-90.

[15] 郭海湘, 刘嫣然, 杨娟, 等. 煤矿物资配送车辆路径问题的人工鱼群算法[J]. 系统管理学报, 2012, 21(3): 341-351.

[16] Li Xiaohua. Parameter optimization of lowest secondary crushing rate for coal auger based on artificial fish school algorithm[J]. Meitan Xuebao/Journal of the China Coal Society,2011,36(2):346-350.

[17] Jiang Mingyan, Wang Yong, Pfletschinger S, *et al.* Optimal multiuser detection with artificial fish swarm algorithm [C]//Proc of the 3rd International Conference on Intelligent Computing. Berlin: Springer, 2007: 1084-1093.

[18] 杨淑莹, 张桦. 群体智能与仿生计算——MATLAB 技术实现[M]. 北京: 电子工业出版社, 2014: 208.

[19] 马昌凤. 最优化方法及其 MATLAB 程序设计[M]. 北京: 科学出版社, 2010: 42.

[20] 王联国, 施秋红. 人工鱼群算法[M]. 北京: 中国农业出版社, 2014: 22.

[21] Shi Yuhui, Eberhart R C. Fuzzy adaptive particle swarm optimization [C]//Proc of Congress on Evolutionary Computation. Piscataway, NJ: IEEE Press,2001: 101-106.