基于改进贝叶斯的书目自动分类算法

杨晓花[1] 高海云[2]

(福州大学至诚学院 福州 350002)[1] (福州大学物理与信息工程学院 福州 350016)[2]

摘 要 贝叶斯算法被广泛应用于书目自动分类领域。该算法常使用差分进化算法来评估概率项,但是传统的差分进化算法容易陷入局部最优解,使得贝叶斯分类精度较低。针对该问题,提出了基于改进贝叶斯的书目自动分类方法。该方法通过多父突变和交叉操作估计概率项的最优解,提高贝叶斯分类精度;在进行书目自动分类时,先采用ICTCLAS系统进行文本预处理,再提取文本的词频-逆向文件频率特征,接着采用改进的贝叶斯估计方法对特征进行训练与分类,最终实现书目的自动分类。仿真结果表明,该方法具有较高的分类准确率。

关键词 贝叶斯算法,书目自动分类,差分进化,特征提取

中图法分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2018.08.036

Improved Bayesian Algorithm Based Automatic Classification Method for Bibliography

YANG Xiao-hua[1] GAO Hai-yun[2]

(Zhicheng College, Fuzhou University, Fuzhou 350002, China)[1]

(College of Physics and Information Engineering, Fuzhou University, Fuzhou 350016, China)[2]

Abstract Bayesian algorithm is widely used in the field of automatic classification for bibliography. This method usually adopts differential evolution method to estimate the probability items. However, the traditional differential evolution method is easy to fall into the local optimum when estimating the probability items, which reduces the accuracy of Bayesian classifcation. In view of this problem, this paper proposed an improved Bayesian algorithm based automatic classification method for bibliography. In this method, the optimal solution of probability items is estimated through multi-parent mutation and crossover operation, which improves the accuracy of Bayesian classification. In the process of automatic classification for bibliography, the ICTCLAS system is used to preprocess the text and then extract the term frequency-inverse document frequency features of texts. Then, the improved Bayesian estimation method is utilized to train and classify the features. Finally, the automatic classification for bibliography is achieved. Simulation results show that this method has a high classification accuracy.

Keywords Bayesian algorithm, Automatic classification for bibliography, Differential evolution, Feature extraction

书目分类[1-3]是一种文本分类技术,依据文本之间的差异性特征以及机器学习方法来实现不同类别文本的分类。文本分类主要包含两项关键技术,一是文本特征的提取,二是分类器的设计[4-10]。在文本特征提取方面,常采用向量空间模型来对文本进行表示和描述。向量空间模型各个维的特征可以直接使用分词或者词频统计算法得到词汇,也可以在词汇的基础上进一步提取更具鉴别能力的特征,如信息增益、词频、互信息、TFIDF(Term Frequency Inverse Document Frequency)等[11-13]。由于直接使用词汇作为特征所得到的向量空间模型的维数太大,不利于求解,因此通常在词汇的基础上进一步提取特征来构建向量空间模型。其中,TFIDF特征的应用最为广泛。在分类器设计方面,随着机器学习技术的迅猛发展,可选的分类器很多。但针对文本分类的特性,目前文本分类领域常用的分类器包括K近邻、支持向量机和朴素贝叶斯[14-16]。K近邻和支持向量机分类方法在处理小样本分类时效果较好,而朴素贝叶斯分类方法对文本分类的适应性更强,适合不同尺寸的样本集的文本分类。因此,朴素贝叶斯分类方法在文本分类领域应用广泛。朴素贝叶斯方法的分类性能主要依赖于估计的条件概率项的准确性。概率项的估计通常使用差分进化方法。现有的差分进化方法依据单亲进行变异和交叉操作,当概率项差异较小时容易陷入局部最优。

为了解决差分进化方法在估计朴素贝叶斯概率项时可能陷入局部最优的问题,本文提出一种改进的贝叶斯估计方法。该方法通过多父突变和交叉操作来估计朴素贝叶斯分类器的概率项,以避免概率项估计陷入局部最优,提高朴素贝叶斯分类的精度,进而提高图书馆书目分类的准确率。

1 中文书目自动分类框架

基于机器学习的中文书目自动分类框架主要包括两个阶段:1)在训练阶段,对于参与训练的书目数据集,先将非结构

到稿日期:2018-02-28 返修日期:2018-04-13 本文受福建省中青年教师教育科研项目(JAT160658)资助。

杨晓花(1979-),女,硕士,高级工程师,主要研究领域为大数据分析、图像处理,E-mail:45665192@qq.com(通信作者);**高海云**(1979-),女,硕士生,助理研究员,主要研究领域为非线性整数规划、图像处理、系统建模算法优化设计。

化的文本数据进行预处理操作，并将之转换为便于计算机处理的结构化数据。预处理操作之后，再进行特征提取操作，建立向量空间模型。对于不同书目所得到的向量空间模型，采用机器学习方法（如 K 近邻、支持向量机和朴素贝叶斯）进行训练，构建分类器。2）在书目自动分类阶段，对于待分类的书目采用同样的预处理和特征提取操作，构建向量空间模型，然后使用已训练好的分类器对向量空间模型进行分类，实现书目的自动分类。

结合上述分析，基于机器学习的中文书目自动分类方法主要包括 3 个关键环节：预处理、特征提取和机器学习。

1）预处理

本文采用中国科学院计算机研究所开发的 ICTCLAS 系统对文本进行预处理操作，得到文本所包含的词条信息，将非结构化的文本信息转换为结构化的词条信息。

2）特征提取

本文使用 TFIDF 特征提取方法，将词条信息描述为向量空间模型。TFIDF 特征用两个项的乘积来表示：

$$TFIDF(t_i,d_j)=TF(t_i,d_j)\times IDF(t_i) \tag{1}$$

其中，$TF(t_i,d_j)$项表示词条 t_i 在文档 d_j 中出现的频度，该值越大说明词条 t_i 在文档 d_j 的相关性越强，从而说明该词条 t_i 对文档 d_j 越重要。$IDF(t_i)$项表示逆文档频度，可以表示为：

$$IDF(t_i)=\log_{10}\left[\frac{N}{DF(t_i)}\right] \tag{2}$$

其中，N 表示训练文档总数；$DF(t_i)$表示词条 t_i 在所有训练文档中出现的总次数。可见，在所有文档中，词条 t_i 出现的频度越大，说明该词条 t_i 对文档的区分能力越弱。

在书目分类时，词条的 TFIDF 值越高，表明该词条的区分能力越强。

3）机器学习

本文使用朴素贝叶斯分类器进行文本的学习与分类。朴素贝叶斯是一种基于概率统计的分类器，依据先验概率和条件概率来计算后验概率，可以表示为：

$$P(C|d)=P(d|C)P(C) \tag{3}$$

其中，$P(C|d)$表示文本 d 属于类别 C 的后验概率；$P(C)$表示类别 C 的先验概率；$P(d|C)$表示类别 C 下文本 d 的条件概率，可以用文本 d 中各个词条的条件概率的乘积来表示：

$$P(d|C)=P(w_1,w_2,\cdots,w_n|C)=\prod_{i=1}^{n}P(w_i|C) \tag{4}$$

其中，$w_1,w_2,\cdots,w_n$ 表示文本 d 中的所有词条。

朴素贝叶斯方法详见文献[16]，本文不再赘述。其中，朴素贝叶斯分类器的条件概率项估计使用差分进化方法。考虑到现有差分进化方法依据单亲进行变异和交叉操作，当概率项差异较小时容易陷入局部最优的问题，本文提出一种改进的贝叶斯估计方法，通过多父突变和交叉操作来估计朴素贝叶斯分类器的概率项。

2 改进朴素贝叶斯概率估计

朴素贝叶斯学习算法是一种简单、高效的文本分类方法，本文选择该方法对文本进行分类。朴素贝叶斯方法的分类性能主要取决于估计的条件概率项的准确性。然而，当训练数据稀缺时，这些条件概率项的估计精度较低，从而导致朴素贝叶斯方法的分类性能下降。条件概率项的估计问题可以看作是一个最优化问题，常用 3 种元启发方法进行求解，包括遗传算法、模拟退火和差分进化方法。差分进化方法是一种随机分析方法，基于种群模型来优化实参数或者实数函数，在朴素贝叶斯概率项估计领域应用广泛。采用差分进化方法可以微调朴素贝叶斯分类器，通过为已使用的概率项寻找更好的估计来调整朴素贝叶斯分类器，其目标是寻找使得分类精度最高的概率估计。然而，经典的差分进化方法容易陷入局部最优，可能导致估计的概率项并不是全局最优的概率项，从而间接地降低了朴素贝叶斯的分类精度。为此，本文提出一种改进的贝叶斯估计方法，通过多父突变和交叉操作来估计朴素贝叶斯分类器的概率项，避免概率项估计陷入局部最优，以提高朴素贝叶斯的分类精度。

2.1 经典差分进化方法

经典差分进化方法的基本思想是：基于临时种群来发现当前种群的个体差异。该算法从下一代中选择拥有最高适应度值的个体，与其父亲生成新的下一代。该过程主要包括 3 个步骤：选择、变异和交叉[17]。差分进化算法通过保存最优个体和消除略弱个体来寻找最优解。

在选择阶段，从当前代 g 中随机选择 3 个个体，记为 $X_{r1,g}$，$X_{r2,g}$ 和 $X_{r3,g}$。

在变异阶段，根据随机选择的两个个体的差分以及另一个个体向量生成变异向量，表示为：

$$V_{i,j,g+1}=X_{r1,j,g}+F\cdot(X_{r2,j,g}-X_{r3,j,g}) \tag{5}$$

其中，$V_{i,j,g+1}$表示第 i 个个体的变异向量的第 j 个元素的值。$X_{r1,j,g}$，$X_{r2,j,g}$ 和 $X_{r3,j,g}$ 分别表示个体 $X_{r1,g}$，$X_{r2,g}$ 和 $X_{r3,g}$ 的第 j 个元素的值。F 是一个常数项，表示变异率，范围在(0，2)之间。

在交叉阶段，变异向量与目标向量基于交叉率进行混合，生成一个试验向量，表示为：

$$U_{i,j,g+1}=\begin{cases}V_{i,j,g+1}, & r\leqslant CR \text{ 或 } j\leqslant I_{\text{rand}}\\ X_{i,j,g}, & \text{其他}\end{cases} \tag{6}$$

其中，CR 表示交叉率；I_{rand} 是一个整形随机数，取值范围为[1，N]，N 为种群尺寸；r 是一个随机数，取值范围为[0，1]。

然后比较目标向量和试验向量的适应度值，并选择适应度值最大的向量传递给下一代。

变异、重组和选择操作迭代进行，直到到达迭代终止条件。这里，迭代终止条件是指代数达到最大。算法的具体实现如下。

输入：适应度函数、初始种群尺寸 N、最大代数 g_{max}、F、CR

输出：适应度函数参数、最大适应度值

1. g=0；
2. while(g<g_{max})
3. for(种群中的任意个体 i)
4. 随机选择 3 个个体 $X_{r1,g}$，$X_{r2,g}$ 和 $X_{r3,g}$；
5. 计算变异向量；
6. 生成随机数 I_{rand} 和 r，计算试验向量；
7. 选择适应度值最大的向量替换 $X_{i,j,g}$。
8. end for
9. g=g+1；
10. end while

采用朴素贝叶斯方法为种群中的每一个个体解决方案的每一个条件概率项分配一个可能的值。经典差分进化算法使用选择、变异和交叉操作来评价由多个个体解决方案组成的种群，寻找朴素贝叶斯分类器的最优概率项。其中，初始种群生成的创建步骤为：首先，在训练样本集中，依据文献[16]所述方法估计朴素贝叶斯的各个概率项，得到最优的解决方案；其次，使用已求解的解决方案作为父向量，生成初始种群的 N 个解决方案。为了保障初始种群的多样性，本文引入一个[0,1]之间的随机数 t，依据该随机数对初始解决方案的概率项进行微调，表示为：

$$K'=\begin{cases}K+t, & J<\dfrac{N}{2}\\ K-t, & \text{其他}\end{cases} \tag{7}$$

其中，K 和 K' 分别表示微调前后概率项的值，J 表示种群中概率项的序号。由式(7)可见，对于排在前半段的概率项，本文在其原始概率项的基础上增加了一个随机数；而对于排在后半段的概率项，在其原始概率项的基础上减去一个随机数。本文通过这种方式增加初始种群的多样性。

单个个体解决方案 X 可以描述为一个 d 维的向量，记为 $X=[x_1,x_2,\cdots,x_d]$，其中，$x_i(i=1,2,\cdots,d)$ 表示待估计的第 i 个概率项，每一代由 N 个这样的向量组成。在采用经典差分进化方法寻找最佳解决方案的过程中，遍历所有变异变量，利用目标向量和变异向量构建试验向量，并选择适应度值最高的向量传递给下一代。该过程采用贪婪搜索策略，一个个体的适应度采用相应朴素贝叶斯分类器的分类精度来度量。重复这一进化过程，直到其收敛到最优。

2.2　改进差分进化方法

在采用差分进化方法寻找最优解决方案的过程中，当最近 5 代的最高适应度所对应的所有概率项的差异小于一个很小的常数值 ε(本文取 0.001)时，最优化过程容易陷入局部最优。为了解决这一问题，本文改进差分进化方法，使用多父差分进化策略挖掘上一代更多的额外信息，提高朴素贝叶斯的分类精度，以便获取全局最优的解决方案。

多父差分进化策略的核心是通过使用多父变异和交叉操作充分挖掘有价值的信息，以便于选择最优的解决方案。选择更好的双亲有助于算法更好地寻找最佳解决方案，同时还需要保持种群的多样性，以便尽可能地避免得到局部最优解。多父差分进化策略允许变异向量从 3 个最好的父母中继承最佳概率项，最好的父母在训练样本集中拥有最好的分类精度值。多父差分进化策略给每个变异向量的元素提供同样的概率来构建试验向量。

多父差分进化策略采用与经典差分进化方法相同的初始种群创建步骤、适应度函数和终止条件。不同的是，多父差分进化策略依据训练数据集中所计算的各个个体对应的朴素贝叶斯分类准确率对初始种群中的各个个体进行降序排列，然后选择排在前 100 位的个体组成候选个体集合。种群中的每个个体经历一个多父变异操作，选出最好的 3 个个体作为变异向量的父亲，记为 $X_{1,g}$，$X_{2,g}$ 和 $X_{3,g}$。其中，g 表示当前代。

然后随机选择 3 个其他个体，记为 $X_{r1,g}$，$X_{r2,g}$ 和 $X_{r3,g}$，并依据适应度值对其进行排序。假定适应度值排序结果为 $f(X_{r1,g})>f(X_{r2,g})>f(X_{r3,g})$，其中，$f$ 表示适应度。那么，多父差分进化策略得到的 3 个变异向量可以表示为：

$$V_{1,j,g+1}=X_{1,j,g}+r_F\cdot(X_{r2,j,g}-X_{r3,j,g}) \tag{8}$$

$$V_{2,j,g+1}=X_{2,j,g}+r_F\cdot(X_{r3,j,g}-X_{r1,j,g}) \tag{9}$$

$$V_{3,j,g+1}=X_{3,j,g}+r_F\cdot(X_{r1,j,g}-X_{r2,j,g}) \tag{10}$$

其中，r_F 是[0,F]之间的一个随机数。

从上述 3 个方程来看，式(8)和式(10)使用两个适应度值相邻的个体来生成解决方案，它们应该可以在相邻的区域内生成一个适合的解决方案，这有助于帮助多父差分进化策略寻找一个更适合的解决方案。另一方面，式(9)使用最大适应度值和第三适应度值对应的个体来生成变异向量，所用的这两个个体并不是很相似，这有助于提高种群的多样性，防止求解过程陷于局部最优。

在多父差分进化策略中，试验向量($U_{i,g+1}$)由目标向量($X_{i,g}$)和 3 个变异向量($V_{1,g+1}$，$V_{2,g+1}$ 和 $V_{3,g+1}$)之间的交叉操作生成，其定义如下：

$$U_{i,j,g+1}=\begin{cases}V_{1,j,g+1}, & 0\leqslant r<\dfrac{CR}{3}\text{或 } j=I_{\text{rand}}\\ V_{2,j,g+1}, & \dfrac{CR}{3}\leqslant r<\dfrac{2CR}{3}\text{或 } j=I_{\text{rand}}\\ V_{3,j,g+1}, & \dfrac{2CR}{3}\leqslant r<CR\text{ 或 } j=I_{\text{rand}}\\ X_{i,j,g}, & \text{其他}\end{cases} \tag{11}$$

其中，交叉范围为[0,CR]，该区间被等分成 3 个子区间，子区间长度为(CR/3)；j 是属性数；$V_{k,j,g+1}$ 表示第 k 个变异向量的第 j 个属性值；$X_{i,j,g}$ 表示第 g 代中第 i 个目标向量的第 j 个属性值；对于试验向量 $U_{i,j,g+1}$，至少有一个属性值来自变异向量；I_{rand} 是一个整形随机数，范围为[1,N]；r 是一个随机数，范围在[0,1]之间。在本文中，参数 $CR=0.8$，种群尺寸 $N=100$，最大 $g_{\max}=30$，变异因子 $F=0.4$。

与经典差分进化方法相同，多父差分进化策略的迭代终止条件仍为繁衍代数到达最大。通过迭代进行选择、变异和交叉操作，寻找最优的解决方案，估计朴素贝叶斯分类器的概率项，构建最终的朴素贝叶斯分类器。

3　实验与分析

3.1　实验数据

本文选用的实验数据与文献[18]相同，选取五大类书目，相关信息如表 1 所列。

表 1　实验数据信息

Table 1　Experimental data information

类别	书目数量
D 类	2042
F 类	4485
I 类	2291
K 类	2012
T 类	6817

书目文本数据中包含了书号、书名、作者、出版社、出版时间、内容摘要、读者对象、分类号等字段，其中，书名和内容摘要两个字段能有效反映图书的主题，本文选取这两个字段作为实验的测试语料。将这两个字段提取的特征取平均值，并

将该值作为书目文本数据的最终特征。

3.2 贝叶斯估计的对比分析

下面以 TFIDF 作为文本特征，选用朴素贝叶斯机器学习方法进行训练和分类，测试不同概率项估计方法得到的分类准确率指标，测试结果如图 1 所示。

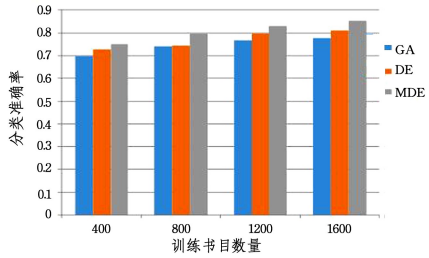

图 1 不同概率项估计方法的测试结果

Fig. 1 Test results of different probability item estimation methods

在图 1 中，GA 表示遗传算法(Genetic Algorithms)[19]，DE 表示经典差分进化(Differential Evolution)算法[17]，MDE 表示本文改进的差分进化(Modified Differential Evolution)算法。分别采用这 3 种方法来估计朴素贝叶斯的概率项，并在不同训练书目数量条件下测试书目的分类准确率指标。可见，在相同的训练书目数量条件下，本文提出的 MDE 方法得到的分类准确率指标高于 GA 和 DE 两种方法；而且随着训练书目数量的增加，尽管 3 种方法对应的分类准确率都有提升，但是 MDE 方法提升的幅度更大，尤其是在训练书目数量从 400 增加到 800 的过程中。这说明，本文改进的贝叶斯估计方法能够有效提高书目分类的准确率。

3.3 特征选择的对比分析

下面采用本文所述的训练与分类方法，对比不同文本特征在不同训练书目数量条件下的书目分类准确率，测试结果如图 2 所示。

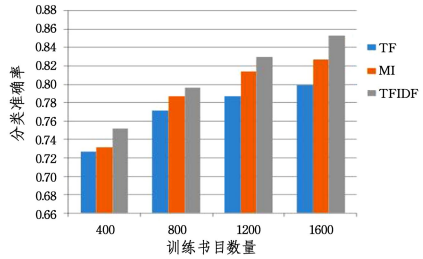

图 2 不同特征选择方法的测试结果

Fig. 2 Test results of different feature selection methods

在图 2 中，TF 表示词条频度(Term Frequency)特征[18]，MI 表示互信息量(Mutual Information)特征[20]。很明显，在训练书目数量相同的条件下，TFIDF 特征对应的书目分类准确率指标高于其他两种特征对应的分类准确率，因此本文选择 TFIDF 特征作为文本分类所使用的特征。

3.4 不同方法的对比分析

下面对比不同书目分类方法对于本文实验数据的分类准确率指标，结果如图 3 所示。其中，TFIDF＋SVM 方法出自文献[18]；MI＋DT 方法出自文献[20]，这里 DT(Decision Tree)表示决策树。

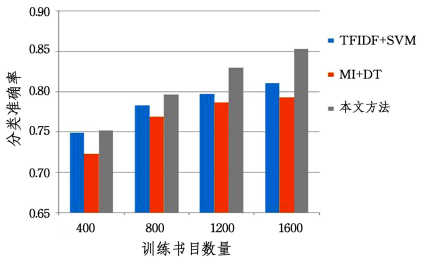

图 3 不同概率项估计方法的准确率

Fig. 3 Accuracy of different probability item estimation methods

由图 3 看出，在相同的训练书目数量条件下，本文方法的分类准确率指标高于其他两种书目分类方法的分类准确率，且训练书目数量越多，该优势越明显。仔细分析图 3 可知，当训练书目数量为 400 时，TFIDF＋SVM 方法与本文方法所得到的分类准确率指标相近；但随着训练书目数量的增加，本文方法所得的分类准确率增加非常明显，而 TFIDF＋SVM 方法所得的分类准确率增加却比较缓慢。以上说明，在特征提取方面，相同的条件下，SVM 方法更适用于小样本的训练与分类；而本文提出的结合改进差分进化的朴素贝叶斯方法既适用于小样本数据，也适用于大样本数据。另外，从图 3 中还可以看出，MI 特征＋DT 分类器对书目的分类效果明显不如 TFIDF 特征＋SVM 或朴素贝叶斯分类器的分类效果。

综上所述，本文采用 TFIDF 特征，结合改进差分进化优化的朴素贝叶斯学习方法，可以有效地进行书目的自动分类，分类准确率达 85％以上。

结束语 本文设计了一种改进贝叶斯估计的图书馆书目自动分类方法，实现了书目的自动分类。本文的核心是提出了改进的贝叶斯估计方法，该方法通过多父突变和交叉操作来估计朴素贝叶斯分类器的概率项，避免了概率项估计陷入局部最优，从而提高了朴素贝叶斯分类的精度。通过与现有的 TFIDF＋SVM 方法和 MI＋DT 方法进行对比可以发现，本文方法对中文书目的分类准确率更高，是一种有效的图书馆书目自动分类方法。然而，本文的所有参数目前都是按照经验进行赋值的，如果能够根据训练数据自适应选取，将能进一步提高书目分类的准确率，这是值得进一步研究的方向。

参 考 文 献

[1] MURTAGH F, KURTZ M J. The Classification Society ' s Bibliography Over Four Decades: History and Content Analysis[J]. Journal of Classification, 2016, 33(1): 6-29.

[2] KLEIN K. A Review of Bibliography Complex: Fundamentals of Librarianship and Knowledge Management[J]. Cataloging & Classification Quarterly, 2014, 52(3): 341-342.

[3] WELDON S P. Organizing knowledge in the Isis bibliography from Sarton to the early twenty-first century[J]. Isis, 2013, 104(3): 540-550.

[4] ARGAMON S, WHITELAW C, CHASE P, et al. Stylistic text classification using functional lexical features[J]. Journal of the Association for Information Science & Technology, 2014, 58(6): 802-822.

[5] LIN Y S, JIANG J Y, LEE S J. A Similarity Measure for Text Classification and Clustering[J]. IEEE Transactions on Knowledge & Data Engineering, 2015, 26(7): 1575-1590.

[6] UYSAL A K, GUNAL S. The impact of preprocessing on text

classification[J]. Information Processing & Management，2014，50(1)：104-112.

[7] D'ASPREMONT A. Predicting abnormal returns from news using text classification[J]. Quantitative Finance，2015，15(6)：999-1012.

[8] SHANG C，LI M，FENG S，et al. Feature selection via maximizing global information gain for text classification[J]. Knowledge-Based Systems，2013，54(4)：298-309.

[9] KANAAN G，AL-SHALABI R，GHWANMEH S，et al. A comparison of text-classification techniques applied to Arabic text [J]. Journal of the American Society for Information Science & Technology，2014，60(9)：1836-1844.

[10] KHORSHEED M S. Comparative evaluation of text classification techniques using a large diverse Arabic dataset[J]. Language Resources & Evaluation，2013，47(2)：513-538.

[11] ABUERRUB A. Arabic Text Classification Algorithm using TFIDF and Chi Square Measurements[J]. International Journal of Computer Applications，2014，93(6)：40-45.

[12] HU J，YAO Y. Research on the Application of an Improved TFIDF Algorithm in Text Classification[J]. Journal of Convergence Information Technology，2013，8(7)：639-646.

[13] GHAG K，SHAH K. SentiTFIDF-Sentiment Classification using Relative Term Frequency Inverse Document Frequency[J]. International Journal of Advanced Computer Science & Applications，2014，5(2)：36-43.

[14] BILAL M，ISRAR H，SHAHID M，et al. Sentiment classification of Roman-Urdu opinions using Navie Baysian，Decision Tree and KNN classification techniques[J]. Journal of King Saud University-Computer and Information Sciences，2016，28(3)：330-344.

[15] CHEN R，CHEN F，SUN Y. Research on Automatic Text Classification Algorithm Based on ITF-IDF and KNN[J]. Applied Mechanics & Materials，2015，713-715：1830-1834.

[16] FENG G，WANG H，SUN T，et al. A Term Frequency Based Weighting Scheme Using Naïve Bayes for Text Classification [J]. Journal of Computational & Theoretical Nanoscience，2016，13(1)：319-326.

[17] GONG W，CAI Z. Differential evolution with ranking-based mutation operators[J]. IEEE Transactions on Cybernetics，2013，43(6)：2066-2081.

[18] YANG M，GU J. Study and Apply of Chinese Bibliographies Automatic Classification Based on Support Vector Machine[J]. Library and Information Service，2012，56(9)：114-119. (in Chinese)
杨敏，谷俊. 基于 SVM 的中文书目自动分类及应用研究[J]. 图书情报工作，2012，56(9)：114-119.

[19] PAULINAS M. A survey of genetic algorithms applications for image enhancement and segmentation[J]. Information Technology & Control，2015，36(3)：278-284.

[20] JIN X R，QI J D，WANG L C，et al. Approach of classification mapping between international patent--classification and chinese library classification based on machine learning[J]. Journal of Computer Applications，2011，31(7)：1781-1784. (in Chinese)
靳雪茹，齐建东，王立臣，等. 基于机器学习的类目映射方法——国际专利分类法与中国图书馆分类法[J]. 计算机应用，2011，31(7)：1781-1784.

[21] YANG B，HAN Q W，LEI M，et al. Short Text Clssification Algorithm Based on Improved TF-IDF Weight [J]. Journal of Chongqing University of Technology (Natural Science)，2016，30(12)：103-113. (in Chinese)
杨彬，韩庆文，雷敏，等. 基于改进 TF-IDF 权重的短文本分类算法[J]. 重庆理工大学学报(自然科学)，2016，30(12)：103-113.

(上接第 185 页)

[12] JAGADEESAN R，JEFFREY A，RIELY J. A calculus of untyped aspect-oriented programs[C]// European Conference on Object-Oriented Programming. Springer Berlin Heidelberg，2003：54-73.

[13] LÄMMEL R. A semantical approach to method-call inte rception[C]// Proceedings of the 1st International Conference on Aspect-oriented Software Development. ACM，2002：41-55.

[14] WALKER D，ZDANCEWIC S，LIGATTI J. A theory of aspects [J]. Acm Sigplan Notices，2003，38(9)：127-139.

[15] TUCKER D B，KRISHNAMURTHI S. Pointcuts and advice in higher-order languages[C]// Proceedings of the 2nd International Conference on Aspect-oriented Software Development. ACM，2003：158-167.

[16] MASUHARA H，KICZALES G. Modeling crosscutting in aspect-oriented mechanisms[C]// European Conference on Object-Oriented Programming. Springer Berlin Heidelberg，2003：2-28.

[17] TABAREAU N. Aspect Oriented Programming：a language for 2-categories[C]// Proceedings of the 10th International Workshop on Foundations of Aspect-oriented Languages. ACM，2011：13-17.

[18] MOLDEREZ T，JANSSENS D. Modular Reasoning in Aspect-Oriented Languages from a Substitution Perspective[C]// Transactions on Aspect-Oriented Software Development XII. Springer Berlin Heidelberg，2015：3-59.

[19] ZHANG Q，KHEDRI R. On the weaving process of aspect-oriented product family algebra[J]. Journal of Logical and Algebraic Methods in Programming，2016，85(1)：146-172.

[20] GANG X，BO Y，MINGYI Z. A Semantics of Poincuts in AspectJ[J]. IERI Procedia，2013，4：323-330.

[21] XIE G，ZHANG M Y，YANG B. A Static Sema Ntic For Aspectj [J]. Journal of Computational Information Systems，2012，8(16)：6951-6962.

[22] XIE G，WEI L，WU X. static semantics of aspect-oriented programming[J]. Computer Science，2017，44(9)：184-189. (in Chinese)
谢刚，韦立，吴祥. 面向方面程序的静态语义研究[J]. 计算机科学，2017，44(9)：184-189.

[23] 陆钟万. 面向计算机科学中的数理逻辑(第 2 版)[M]. 北京：科学出版社，2002：117-118.

[24] HOARE A R C，HE J. Unifying theories of program ming[M]. Englewood Cliffs：Prentice Hall，1998.