Advances in Applied Mathematics 应用数学进展, 2023, 12(3), 1188-1194
Published Online March 2023 in Hans. https://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2023.123120

Hans汉斯

简述特征选择

赵锦芳

内蒙古大学数学科学学院，内蒙古 呼和浩特

收稿日期：2023年2月21日；录用日期：2023年3月20日；发布日期：2023年3月27日

摘　要

特征选择是通过寻找对于目标函数有突出贡献的特征来达到降维的效果，该方法希望可以尽可能多的去掉冗余特征，能够更加准确合理地解释这些数据。研究者们对于特征选择的研究历史也比较悠久，从而特征选择也变得越来越准确、有效。本文介绍特征选择概念之后，简单综述了特征选择在方法和理论上的发展，并且重点介绍了支持向量机在特征选择上的应用。

关键词

特征选择，模式识别，非凸优化

Brief Description of Feature Selection

Jinfang Zhao

School of Mathematical Sciences, Inner Mongolia University, Hohhot Inner Mongolia

Received: Feb. 21st, 2023; accepted: Mar. 20th, 2023; published: Mar. 27th, 2023

Abstract

Feature selection is to reduce dimension by finding features that contribute significantly to the objective function. This method hopes to remove redundant features as much as possible and interpret these data more accurately and reasonably. Researchers have a long history of research on feature selection, so feature selection is becoming more and more accurate and effective. After introducing the concept of feature selection, this paper briefly reviews the development of feature selection methods and theories, and focuses on the application of support vector machines in feature selection.

Keywords

Feature Selection, Pattern Recognition, The Convex Optimization

文章引用: 赵锦芳. 简述特征选择[J]. 应用数学进展, 2023, 12(3): 1188-1194.
DOI: 10.12677/aam.2023.123120

1. 引言

特征选择在模式识别领域的地位至关重要[1]。在模式识别系统中，一般都为模式获取，预处理，特征选择或特征提取，采用回归分析、分类研究或描述性研究处理数据，后处理这五个部分。其中，特征选择与特征提取也是人们经常混淆的概念。其实这两者有着本质的区别。

特征提取是在高维样本空间中，通过映射或变换的方法来降低样本维度，该映射或变换通过将原始特征作某种组合来形成新特征，该新特征被称为二次特征。特征提取的本质为一种变换，通过将测量空间中的变量通过变换转化为新特征空间中的变量，该变换称为特征提取器。常见的特征提取方法为主成分分析(PCA)，独立主成分分析(ICA)等。

特征选择则是在高维特征空间中挑选出一些最有效的特征从而实现降维的目的[2] [3]。在特征选择的过程中，一方面，在样本有限的前提下，不管是从开销成本还是从分类器性能来设计分类器都不是很理想；另一方面，线性关系在分类器性能和特征之间并不存在，当特征数量超出一定阈值时，将破坏分类器性能。因此，正确且有效地进行特征选择则在模式识别中显得尤为重要。

本文首先对文中涉及的标记和符号进行说明，其次简单介绍特征选择的基本知识及求解方法，接下来重点叙述特征选择的稀疏模型及其发展，最后对全文进行了总结。

2. 符号

$X \in R^{n\times p}$ 是标准化的数据矩阵，n 为样本数量，p 为特征数量，不失一般性，假设 X 的列均值为 0。单位矩阵记为 I。 $\alpha = \left(a_1, a_2, \cdots, a_p\right)^{\mathrm{T}} \in R^p$， $\|\alpha\|_2 = \left(\sum_{i=1}^{p} a_i^2\right)^{1/2}$， $\|\alpha\|_1 = \sum_{i=1}^{p} |a_i|$， $X = \begin{pmatrix} x_1^{\mathrm{T}} \\ x_2^{\mathrm{T}} \\ \vdots \\ x_n^{\mathrm{T}} \end{pmatrix}, x_i \in R^p$，矩阵 $A = \left(a_{ij}\right) \in R^{p\times k}$， $B \in R^{p\times n}$ 为数据矩阵， $e = \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix}^{\mathrm{T}} \in R^{n\times 1}$， $\|x\|' = \max_{\|y\|=1} x^{\mathrm{T}} y$。

3. 特征选择

特征选择的方法可如文献[4]中所述由特征子集形成方式和特征评价方法来分为两大类，本文将由这两大类为出发点介绍特征选择的算法。

3.1. 特征子集形成方式

3.1.1. 采用全局最优搜索策略的特征选择算法

分支定界算法策略运用到的两个分支规则[5]：

最低下界优先，即选择具有最低下界的子集进行分支，它能提供寻求最优值的最好机会，并且通过探测来排除大量非最优点。但如果问题规模较大，寻找最优子集则相当的耗费时间。

最新下界优先，即选择最新形成的子集进行分支，该方法避免了算法流程反复在对应的搜索树中跳转，并且缩短了定界的时间，从而可以提高算法的效率，如果当前考察的子问题对应的可行解与最优解差别很大时，那么将会进行大量不必要的界估计运算。

各类样本可以分开是因为它们位于特征空间中的不同区域，显然这些区域之间距离越大，类别可分性就越大[6]。

令 $x_k^{(i)}$、$x_l^{(j)}$ 分别为 w_i、w_j 类中的特征向量，$d\left(x_k^{(i)},x_l^{(j)}\right)$ 为它们之间的欧氏距离，则类别间的平均距离(类间距离)为

$$J_d(x)=\frac{1}{2}\sum_{i=1}^{C}P_i\sum_{i=1}^{C}P_j\frac{1}{n_i n_j}d\left(x_k^{(i)},x_k^{(j)}\right) \tag{1}$$

式中：C 为类别数；n_i、n_j 分别为 w_i、w_j 类的样本数；P_i、P_j 是相应的先验概率。若 m_i 表示第 i 类样本集特征的均值：

$$m_i=\frac{1}{n_i}\sum_{k=1}^{n_i}x_k^{(i)} \tag{2}$$

m 表示所有各类的样本集特征的总平均向量：

$$m=\sum_{i=1}^{C}P_i m_i \tag{3}$$

$$d\left(x_k^{(i)},x_l^{(j)}\right)=\left(x_k^{(i)}-x_l^{(j)}\right)^T\left(x_k^{(i)}-x_l^{(j)}\right) \tag{4}$$

将(2)、(3)、(4)代入(1)中得：

$$J_d'(x)=\sum_{i=1}^{C}P_i\left[\frac{1}{n_t}\sum_{k=1}^{n_i}\left(x_k^{(i)}-m_t\right)^{\mathrm{T}}\left(x_k^{(i)}-m_t\right)+\left(m_i-m\right)^{\mathrm{T}}\left(m_t-m\right)\right]=tr\left(S_w+S_b\right)$$

其中，$S_b=\sum_{i=1}^{C}P_i\left(m_i-m\right)\left(m_i-m\right)^{\mathrm{T}}$ 为类间散度矩阵；$S_w=\sum_{i=1}^{C}P_i\frac{1}{n_i}\sum_{k=1}^{n_i}\left(x_k^{(l)}-m_i\right)\left(x_k^{(l)}-m_i\right)^{\mathrm{T}}$ 为类内散度矩阵。易知，类间样本散度越大，类内样本散度越小，则类别之间越可分。

3.1.2. 采用随机搜索策略的特征选择算法

采用随机搜索策略进行特征选择其本质上是一个组合优化问题，它通常将特征选择与模拟退火算法、禁忌搜索算法、遗传算法等结合起来进行求解。其中，最常使用的，还是遗传算法(Genetic Algorithm，简称 GA) [7]。

通过对特征集 $F=\{f_1,f_2,\cdots,f_N\}$ 中的每一个特征利用二进制进行遗传编码，得到一个码长为 N 的二进码串：$H=\{h_1h_2\cdots h_N\mid h_i\in\{0,1\},i=1,\cdots,N\}$ 为所有特征子集的集合，称为个体空间。个体空间的大小可由组合数求得：$C_N^1+C_N^2+\cdots+C_N^N=2^N-1$。每一个个体都可表示为 $h_1,h_2,\cdots,h_N\in H$，同时将样本分为训练集和测试集两部分，用训练集训练模型，然后再测试其分类的正确率，最后计算平均分类正确率。易知，想要分类更准确，则需要让平均分类正确率达到最高，故可将平均分类正确率作为目标函数，并求其最大值。设 $r^1,r^2,\cdots,r^M$ 表示 M 类分类问题中的每一类正确分类率，则此类问题可归结为：$\max_{h\in H}f(h)$，其中，$f(h)=\frac{1}{M}\sum_{i=1}^{M}r_i$。在遗传算法中，经常使用适应度分配进行概率选择，一般情况下，我们都想利用使目标函数达到最大值的最优解来进行选择，但由于我们想要得到较少的特征，故同时想到了惩罚函数法：$f_1(h)=\frac{1}{M}\sum_{i=1}^{M}r_i-c\cdot l(h)$，其中，$l(h)$ 为个体 h 中 1 的位数。惩罚参数 c 的大小则根据需要进行选择。

3.1.3. 采用启发式搜索策略的特征选择算法

采用启发式搜索策略进行特征选择的方法有：单独最优特征选择(通过计算特征每个单独使用时的判

据值对特征进行排序)、序列前向选择方法(把需要的特征集合先初始化成一个空集，每一次向特征集合中增加一个特征直到达到最后的特征集。)、序列后向选择方法(首先假定整个特征集合就是所需要的优化特征集，然后在算法的每一步运行过程中删除一个对准则函数没有贡献的特征，直到剩余特征的个数符合问题要求。)、增 l 去 r 选择方法(其本质为序列前向以及后向的结合，但该方法比序列后向选择方法运算速度更快，比序列前向选择方法运算效果更好)。

3.2. 集合评价策略

3.2.1. 基于滤波(Filter)评价策略[8]的特征选择算法

在滤波评价策略中，其中一种为 Fscore 方法[9] [10]，首先对特征进行评分，其中利用到的是费舍尔准则，然后根据特征被评的分数来确定特征的好坏。类内距离越小，类间距离则越大，则特征分值越高，这意味着该特征有越强的分类特征。在选择子集时，将每个特征的得分由高到低的进行排序，最后根据问题的要求，选取分值最高的 k 个特征作为选择出的最优特征集。

Fscore 评估准则可利用以下公式：

$$SC_F\left(f_i\right)=\frac{\sum_{j=1}^{C} n_j\left(\mu_j^i-\mu^i\right)^2}{\sum_{j=1}^{C} n_j\sigma_j^{i2}}$$

其中，f_i 表示第 i 个特征，μ^i 是特征 f_i 的均值，n_j 是第 j 类样本的数目，μ_j^i 表示第 j 类的特征 f_i 的均值 σ_j^i 表示第 j 类的特征 f_i 的方差。Fscore 特征选择算法具体表示为以下步骤：

1) 计算每类训练样本集总体均值：μ_j^i；
2) 计算训练样本集总体均值：μ^i；
3) 计算每个特征的评分：$SC_F\left(f_i\right)$；
4) 根据特征评分，对特征进行降序排列；
5) 选取前 k 个特征作为最终的特征集。

3.2.2. 基于打包(Wrapper)评价策略的特征选择算法

打包方法和其所使用的分类器息息相关，在筛选特征的过程中，它直接利用所选的特征来训练分类器，并且根据这个分类器在验证集上的特点来评价所选择的特征。其中，有利用决策树进行特征选择、Fisher 判别分析结合遗传算法、结合极大似然模型进行特征选择、用遗传算法结合人工神经网络等进行特征选择的方法都取得了良好的效果[11] [12]。最值得注意的是，Wrapper 方法进行特征选择时都需要有良好的分类器作为基础，接下来，本文主要介绍目前的一项热点研究：用支持向量机进行特征选择。

4. 支持向量机(SVM)

支持向量机为一种二分类方法，目的是构造一个超平面能够使得样本点距离该超平面的距离能够最大化，超平面通常构造为线性方程：$\omega^{\mathrm{T}}x+b=0$。样本中距离超平面最近的一些点叫做支持向量，支持向量与超平面之间的距离为：$d=\frac{\left|\omega^{\mathrm{T}}x+b\right|}{\|\omega\|}$，变量 y 用来判断分类正确与否，正确时 $y=1$，错误时，则有 $y=-1$。我们想要最大化这个距离：

$$\max\frac{y\left(\omega^{\mathrm{T}}x+b\right)}{\|\omega\|}$$

通过转换，我们得到最优化问题：

$$\begin{aligned}&\min \frac{1}{2}\|\omega\|^2\\&\text{s.t. } y_i\left(\omega^{\mathrm{T}}x_i+b\right)\geq 1\end{aligned}。$$

一般情况下，我们用 Lagrange 乘子法进行求解。

5. 基于支持向量机(SVM)与凹函数构造的特征选择

前文中提到求解支持向量机模型通常采用 Lagrange 乘子法，在求对偶问题时，研究者们发现，稀疏变量 α 可利于样本稀疏化，达到提高运算效率，减少运算时间的目的。同样可得到：稀疏变量 ω 可得到特征稀疏化，Shao 等人[13]同时加入特征稀疏项 ω 和样本稀疏项 α，得到了稀疏的对偶最小二乘支持向量机(SPDLSSVM)：

$$\begin{aligned}&\min_{\omega,\alpha}\frac{1}{2}\|\omega\|_2^2+C_1\|\omega\|_1+\frac{C}{2}\left\|B^{\mathrm{T}}\omega-e\right\|_2^2+C_2\|\alpha\|_1+\frac{1}{2}\alpha^{\mathrm{T}}\left(YXX^{\mathrm{T}}Y^{\mathrm{T}}+\frac{1}{C}I\right)\alpha-e^{\mathrm{T}}\alpha\\&\text{s.t. } \omega-B\alpha=0\end{aligned}$$

其中，ω 和 α 都采用了 l_1 范数进行稀疏化，接着使用交替方向乘子法(ADMM)求出了模型的最优解，并且用 MATLAB 进行数值实验在多个数据集上 SPDLSSVM 模型虽然训练时间比其余模型长一点，但都筛选到了最少的特征以及样本，同时分类准确率也是最好的，所以 SPDLSSVM 模型在特征选择、样本选择，以及分类效果上都有着优异的表现。

在文章[14]中，作者介绍了用凹函数法和 SVM 进行特征选择，凹函数法将目标函数转化成了：

$$\begin{aligned}&\min_{\omega,\gamma,y,z,v}(1-\lambda)\left(\frac{e^{\mathrm{T}}y}{m}+\frac{e^{\mathrm{T}}z}{k}\right)+\lambda e^{\mathrm{T}}\left(e-\varepsilon^{-\alpha v}\right)\\&\text{s.t. } -A\omega+e\gamma+e\leq y,\\&\quad B\omega-e\gamma+e\leq z,\\&\quad y\geq 0,z\geq 0,\\&\quad -v\leq\omega\leq v\end{aligned}$$

基于这个出发点，作者用 SVM 的方法同样得到了特征选择的模型：

$$\begin{aligned}&\min_{\omega,\gamma,y,z,v}(1-\lambda)\left(e^{\mathrm{T}}y+e^{\mathrm{T}}z\right)+\frac{\lambda}{2}\|\omega\|'\\&\text{s.t. } -A\omega+e\gamma+e\leq y,\\&\quad B\omega-e\gamma+e\leq z,\\&\quad y\geq 0,z\geq 0\end{aligned}$$

其中，$\|\omega\|'$ 表示为对偶范数，以此来获得较少的特征。

该文用连续线性算法对凹函数模型进行了求解，并且与 l_1、l_2、l_∞ 范数下的 SVM 模型、序列化方法(RLP)在 6 个公开数据集上进行了实验对比，可以看出该文章提出的凹函数模型在 6 个公开数据集上都选择出了最少的特征并且在其中 3 个数据集上都取得了最高的测试分类准确率，由此可得出该凹函数模型不仅能够得到最少的特征个数，还能够保证一定优秀的分类效果。

在文章[15]中，通过建立非线性核支持向量机来进行特征选择，本质上是引入了一个核函数 $K(A,A')$，

将 $R^{m\times n}\times R^{n\times l}$ 映射到 $R^{m\times l}$ 中，从而将分类问题转化为以下的线性规划问题：

$$\begin{aligned}&\min_{u,\gamma,y,s} ve'y+e's\\ &\text{s.t.}\quad D\left(K\left(A,A'\right)u-e\gamma\right)+y\geq e\\ &\qquad -s\leq u\leq s,\\ &\qquad y\geq 0\end{aligned}$$

其中，v 是一个用来测量误分率的非负参数，易知：$e'y=\|y\|_1$，$e's=\|s\|_1$，为使输入空间中样本的维数尽可能得小，引入了一个对角矩阵 $E\in R^{n\times n}$，其中主对角线上元素为 0 或 1，0 对应的是被删除掉的样本特征，1 对应的是被保留的样本特征，则此时分类器变为：$K\left(x'E,EA'\right)u-\gamma=0$，上面的线性规划也就变成了下面的混合整数规划：

$$\begin{aligned}&\min_{u,\gamma,y,s,E} ve'y+e's+\sigma e'Ee\\ &\text{s.t.}\quad D\left(K\left(AE,EA'\right)u-e\gamma\right)+y\geq e,\\ &\qquad -s\leq u\leq s,\\ &\qquad y\geq 0,\\ &\qquad E=diag\left(1\text{ or }0\right)\end{aligned}$$

其中，σ 是一个对特征抑制项进行加权的非负参数，$e'Ee=\sum_{i=1}^{n}E_{ii}$。混合整数规划也是一个 NP-难问题，文章[15]也提出了一种减少特征支持向量机(RFSVM)来对模型进行求解。在此文中，作者在 UCI 数据集上对比了在(RFSVM)、回归特征消除法(RFE)和浮雕法(Relief) (一种用于特征筛选的方法)这三种算法下特征选择个数与分类结果，结果显示该模型都得到了更少的特征和更高的分类准确率。

文章[16]提出了 MICReliefF 算法并且以支持向量机模型分类的准确率作为评价指标，实现了 MICReliefF 算法和分类模型的交互优化，并且在多个 UCI 公开数据集上对该算法的性能做了实验。结果表现出 MICReliefF-SVM 自动特征选择算法不仅可以筛除更多的冗余特征，而且可以选择出具有良好稳定性和泛化能力的特征子集。与随机森林、最大相关最小冗余、相关性特征选择等特征选择算法相比，此算法具有更高的分类准确率。

6. 结论

本文主要介绍了特征选择的相关概念及模型，并且介绍了当前研究的重点内容：支持向量机模型[17]，然后将二者结合起来，即由支持向量机作为分类器进行特征选择的模型[18]，为了进行特征选择，研究者们构建了线性规划模型、混合整数规划模型、二次规划模型等[19]，但目前随着对于支持向量机模型对于特征选择应用的研究越来越深入[20] [21] [22] [23] [24]，研究者们发现可以用稀疏优化支持向量机的系数来进行特征选择，即构建出加入 l_0 范数的模型，但由于目标函数因此变得非凸非光滑，该问题也就成为了 NP-难问题。很多学者将 l_0 范数近似为 l_1 范数来进行求解[25]，而直接求解 l_0 范数问题取决于非凸优化的发展理论，同时此类问题的研究也成为了当前研究的热点问题。l_0 范数也具有较强的模型解释性以及实用性，因此在 l_0 范数上进行理论与计算创新也是本文作者今后的努力方向。

参考文献

[1] Guyon, I. and Elisseeff, A. (2003) An Introduction to Variable and Feature Selection. *The Journal of Machine Learning Research*, **3**, 1157-1182.

[2] 吴青, 付彦琳. 支持向量机特征选择方法综述[J]. 西安邮电大学学报, 25(5): 16-21.

[3] Geng, X., Liu, T.-Y.,Tao, Q. and Li, H. (2007) Feature Selection for Ranking. *Proceedings of the* 30*th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Amsterdam, 23-27 July 2007, 407-414. https://doi.org/10.1145/1277741.1277811

[4] 王娟, 慈林林, 姚康泽. 特征选择方法综述[J]. 计算机工程与科学, 2005, 27(12): 68-71.

[5] 汪祖柱, 程家兴. 求解组合优化问题的一种方法——分枝定界法[J]. 安徽大学学报(自然科学版), 2004, 28(1): 10-14.

[6] 闫相国, 明利强. 分支定界算法在白细胞特征选择中的应用研究[J]. 天津职业技术师范学院学报, 2004, 14(3): 9-12.

[7] 韦振中, 黄廷磊. 基于支持向量机和遗传算法的特征选择[J]. 广西科技大学学报, 2006, 17(2): 18-21.

[8] Lazar, C., *et al*. (2012) A Survey on Filter Techniques for Feature Selection in Gene Expression Microarray Analysis. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, **9**, 1106-1119. https://doi.org/10.1109/TCBB.2012.33

[9] 赵学华, 刘学艳, 杨欣斌, 等. 一种混合特征选择方法及应用研究[J]. 深圳信息职业技术学院学报, 2016, 14(3): 11-18.

[10] He, X., Cai, D. and Niyogi, P. (2005) Laplacian Score for Feature Selection. In: Weiss, Y., Schölkopf, B. and Platt, J., Eds., *Advances in Neural Information Processing Systems* 18, MIT Press, Cambridge.

[11] Zhu, Z., Ong, Y.-S. and Dash, M. (2007) Wrapper-Filter Feature Selection Algorithm Using a Memetic Framework. *IEEE Transactions on Systems, Man, and Cybernetics, Part B* (*Cybernetics*), **37**, 70-76. https://doi.org/10.1109/TSMCB.2006.883267

[12] Chandrashekar, G. and Sahin, F. (2014) A Survey on Feature Selection Methods. *Computers & Electrical Engineering*, **40**, 16-28. https://doi.org/10.1016/j.compeleceng.2013.11.024

[13] Shao, Y.-H., Li, C.-N., Huang, L.-W., *et al*. (2019) Joint Sample and Feature Selection via Sparse Primal and Dual LSSVM. *Knowledge-Based Systems*, **185**, Article ID: 104915. https://doi.org/10.1016/j.knosys.2019.104915

[14] Bradley, P.S. and Mangasarian, O.L. (1999) Feature Selection via Concave Minimization and Support Vector Machines. Morgan Kaufmann Publishers Inc.

[15] Mangasarian, O.L. and Gang, K. (2008) Feature Selection for Nonlinear Kernel Support Vector Machines. *Seventh IEEE International Conference on Data Mining Workshops* (*ICDMW* 2007), Omaha, 28-31 October 2007. https://doi.org/10.1109/ICDMW.2007.30

[16] 葛倩, 张光斌, 张小凤. 基于最大信息系数的 ReliefF 和支持向量机交互的自动特征选择算法[J]. 计算机应用, 2022, 42(10): 3046-3053.

[17] Zhang, X., Lu, X., Shi, Q., *et al*. (2006) Recursive SVM Feature Selection and Sample Classification for Mass-Spectrometry and Microarray Data. *BMC Bioinformatics*, **7**, Article No. 197. https://doi.org/10.1186/1471-2105-7-197

[18] Chen, Y.-W. and Lin, C.-J. (2008) Combining SVMs with Various Feature Selection Strategies. *Studies in Fuzziness and Soft Computing*, **207**, 315-324. https://doi.org/10.1007/978-3-540-35488-8_13

[19] Bradley, P.S., Mangasarian, O.L. and Street, W.N. (1998) Feature Selection via Mathematical Programming. *INFORMS Journal on Computing*, **10**, 121-260. https://doi.org/10.1287/ijoc.10.2.209

[20] Kalousis, A., Prados, J. and Hilario, M. (2007) Stability of Feature Selection Algorithms: A Study on High-Dimensional Spaces. *Knowledge & Information Systems*, **12**, 95-116. https://doi.org/10.1007/s10115-006-0040-8

[21] Fung, G. and Stoeckel, J. (2007) SVM Feature Selection for Classification of SPECT Images of Alzheimer's Disease Using Spatial Information. *Knowledge & Information Systems*, **11**, 243-258. https://doi.org/10.1007/s10115-006-0043-5

[22] Su, C.-T. and Yang, C.-H. (2008) Feature Selection for the SVM: An Application to Hypertension Diagnosis. *Expert Systems with Applications*, **34**, 754-763. https://doi.org/10.1016/j.eswa.2006.10.010

[23] Gupta, P., Doermann, D.S. and Dementhon, D. (2002) Beam Search for Feature Selection in Automatic SVM Defect Classification. 2002 *International Conference on Pattern Recognition*, Quebec City, 11-15 August 2002. https://doi.org/10.1109/ICPR.2002.1048275

[24] Li, G.-Z., Wang, Z.-X., Yang, J., *et al*. (2002) A SVM-Based Feature Selection Method and Its Applications. *Computers and Applied Chemistry*, **19**, 703-705.

[25] Yang, Y., Shen, H.T., Ma, Z., *et al*. (2011) L21-Norm Regularized Discriminative Feature Selection for Unsupervised learning. *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence-Volume Volume Two*, 16-22 July 2011, 1589-1594.