

doi:10.16055/j.issn.1672-058X.2020.0005.015

基于因子分析和 SVM 的网络舆情危机预警研究*

朱光婷, 潘晓琳**

(重庆师范大学 数学科学学院, 重庆 401331)

摘要:针对网络舆情的指标冗余和复杂度高不利于监管,提出了因子分析和 SVM 建立综合评价模型;首先利用因子分析将网络舆情的 14 个指标进行降维为 3 个公因子,其次在简化的指标体系中用遗传算法的 5-折交叉优化 SVM 参数,建立遗传算法优化 SVM 的网络舆情危机预警模型,最后将两类的 SVM 改进为一对多算法对 4 种情况进行分类,得出网络舆情的预警;对 2019 年的 10 个网络舆情事件进行实证分析表明,低于 0.51% 的误差预警充分说明了其可行性,达到了强化网络舆情的监管,而因子分析降低了指标体系的复杂性,遗传算法的 5-折交叉提高了 SVM 分类器的学习能力,能更准确地预测训练集,并用一对多算法使得分类速度较快,对网络舆情的监管提供了帮助。

关键词:网络舆情;因子分析;遗传算法;SVM;一对多算法

中图分类号: O212.5

文献标志码: A

文章编号: 1672-058X(2020)05-0094-07

0 引言

在互联网时代,网络成为人们生活中不可缺少的一部分,为社会带来巨大的进步,但是在进步的背后发现其中隐藏着网络舆情的危机。人人言论自由,一个事件的多条言论就会对人民,对社会,甚至对国家都有着或多或少的影响,控制舆情就是目前一个紧张而重要的工作。尽管网络舆情的控制是一个难点,但如何提高网络舆情预警仍然得到政府机关和企业单位的高度重视,受到国内外学者的密切关注。孙玲芳等提出利用遗传算法优化 BP 神经网络的初始权值和阈值,构建基于 BP 神经网络和遗传算法的网络舆情危机预警模型;周晓琳等人提出一种基于 AHP 的网络舆情安全与检测模型;张

巍提出基于 Logistic 模型的网络舆情预警系统,Gil-Garcia 提出使用动态层次文本聚类法找出网络舆情预警过程中的热点话题^[1],国内外对网络舆情这一问题展开多种研究。在如今也做出了许多的成果来,但是总的来讲,国内的发展还不成熟需要做出更大改变,在各个方面、各种方法上做出一定努力。目前,支持向量机无论是理论研究还是实际应用在国际上都飞速发展,而国内对于此领域的研究还有待提升^[2],而且在网络舆情方面利用支持向量机建立预警模型的屈指可数。

针对网络舆情的指标冗余和复杂度高不利于监管,提出了因子分析和 SVM 建立综合评价模型。对于网络舆情烦琐的大数据利用因子分析进行简化,能够很好地涵盖原始数据的信息,将原始数据的分析过程简化为主因子的分析,为数据处理降低

收稿日期:2019-11-02;修回日期:2019-12-09.

* 基金项目:重庆市教委科学技术研究基金项目资助(KJ130658);重庆市自然科学基金项目资助(CSTC2019JCYJ-MSXM0801);重庆市教委科学技术研究基金项目资助(KJ1400521)。

作者简介:朱光婷(1996—),女,重庆开州区人,研究生,从事概率论与数理统计研究。

** 通讯作者:潘晓琳(1972—),女,副教授,硕士生导师,从事系统分析研究. Email:1943376596@qq.com.

难度。利用得到的主因子的数据建立遗传优化支持向量机的网络舆情模型,其中为了精确得到最优参数,将数据集等分为5份,进行5-折交叉运算,进一步得到两类SVM最优分类函数。对于多类别分类需要用一对多的算法训练4个svm二类分类器,分别计算出测试数据中的决策函数值,选取最大的决策函数值其所对应的类别决定为方案的类别。

1 因子分析简化指标体系

在网络舆情危机预警的研究中,海量的信息不同程度的反映对网络舆情带来的危机。有的指标带来不同的影响,有的可能是相同的影响,有的指标对网络舆情的危机微乎其微,就据实际情况而言,大多数的指标数据是冗余复杂的,可能对网络舆情的危机影响不同,也可能各指标间有一定的相关性。如果直接用原始指标进行预警,不仅计算起来是困难的而且往往会降低评价模型的精确度,因此用因子分析将错综复杂的指标降维为少数的公共因子,为数据处理降低难度。

1.1 因子分析方法的计算模型

设有 $n \times p$ 的一组数据 X , n 为样本数,每个样本数有 p 个变量,因子分析方法就是把 p 个变量表示成 $m(m < p)$ 个公共因子 $F = (F_1, F_2, \dots, F_m)$ 和特殊因子 $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p)$ 的线性加权和。因子分析模型: $X = AF + \varepsilon$,其中为 A 因子载荷矩阵, $A = (a_{ij})$, a_{ij} 为第 i 个变量在第 j 个因子上的载荷,反映了 X_i 和 F_j 之间的相关程度, a_{ij} 绝对值越大,表明 X_i 和 F_j 的相关程度越大。由于因子分析每个相应的系数不是唯一的,可知因子载荷矩阵也不是唯一的。

1.2 因子分析的基本步骤

首先需要看分析变量之间是否具有较强的相关性,并检验待分析的原始变量是不是适合用因子分析方法,可根据巴特利特球体检验、相关矩阵检验和KMO检验来确定。选用比较变量间简单系数和偏相关系数来确定是否因子分析的KMO检验^[3]。

$$F_{\text{KMO}} = \frac{\sum_{j \neq i} \sum_{i \neq j} r_{ij}^2}{\sum_{j \neq i} r_{ij}^2 + \sum_{j \neq i} p_{ij}^2} \quad (1)$$

其中 r_{ij} 是变量 x_i 和 x_j 的简单相关系数, p_{ij} 是变量 x_i 和 x_j 的偏相关系数。 $F_{\text{KMO}} > 0.6$ 才适合因子分析,表示变量间的相关性较强。

提取公因子,可以用主成分分析法提取公因子,再选择适合公因子的数量。公因子的数量一般可以根据特征值 $\lambda_i (i=1, 2, \dots, p)$ 的大小确定,选取大于1的特征值,也可以根据因子累计方差贡献率来确定^[4],累计方差贡献率($F_{(k)}$):

$$F_{(k)} = \frac{\sum_{k=1}^i \lambda_k}{\sum_{k=1}^p \lambda_k} \quad (2)$$

应达到70%,选出合适数量的公因子能合理地描述原始变量的信息。

因子旋转,首先构建因子载荷矩阵,计算因子载荷矩阵 A 的方法有3种,主成分法、极大似然法和主因子法,选取主成分法。由于因子载荷矩阵实际意义不明显,通过正交旋转中的最大方差法给出旋转后的因子载荷矩阵,因子旋转不改变变量共同度,只改变公因子的方差贡献率,使得分析旋转之后的公因子意义更加清楚。

最后进行因子命名,计算因子得分,进行结果解释。根据因子载荷较大对应的几个原始变量的含义对因子进行命名,用命名后的公因子能反映原始变量的关系,用公共因子表示变量的线性组合称为得分函数:

$$F_j = b_{j1}x_1 + b_{j2}x_2 + \dots + b_{jp}x_p, j=1, 2, \dots, m \quad (3)$$

对于每一样本数据,得到公共因子在不同因子的具体数值。

2 建立SVM模型确定危机等级

2.1 网络舆情预警危机等级

网络舆情危机预警模型中,输出参数为网络舆情预警危机等级类别^[5],根据公共事件危机等级分为4级,A级(特别严重)(1,0,0,0),B级(比较严重)(0,1,0,0),C级(一般严重)(0,0,1,0),D级(比较安全)(0,0,0,1),因此网络舆情危机预警模型的输出参数为预警危机等级值的类别。以下将根据网络舆情危机等级值与实际情况做比较,确定是否符合实际情况;也会用网络舆情危机等级值与

期望值间的误差大小。

2.2 网络舆情预警危机模型建立

首先进行归一化数据预处理^[6]

$$\frac{X_i - \min(X)}{\max(X) - \min(X)}$$

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i)\}$$

$i=1, 2, \dots, n; x \in R^d; y \in \{+1, -1\}$ 为训练集, 对于非线性可分问题, 需要通过非线性变换将其转化为高维空间的线性问题^[7], 在变换空间中求最优分类面, 其目标函数为

$$\min \varphi(\omega) = \frac{1}{2}(\omega \cdot \omega) + C \sum_{i=1}^n \xi_i$$

约束条件为

$$y_i(\omega \cdot x_i + b) - 1 + \xi_i \geq 0, (i=1, 2, \dots, n)$$

其中 $\varphi(x)$ 是非线性映射函数, w 为权重系数, C 为惩罚系数, 作用于对错分样本惩罚程度控制, ξ_i 为松弛变量进行调整, 转化为对偶问题^[8]。

目标函数为

$$\max Q(a) = \sum_{i=1}^n a_i - \frac{1}{2} \sum_{j=1}^n a_i a_j y_i y_j K(x_i, y_j)$$

约束条件

$$\sum_i y_i a_i = 0$$

$$0 \leq a_i \leq C, (i=1, 2, \dots, n)$$

其中 a_i 为格朗日系数, 原来的 (x_i, y_i) 用 $K(x_i, y_i)$ 替代, 选取的核函数为径向基函数^[9]:

$$K(x_i, x_j) = \exp\left\{-\frac{|x_i - x_j|^2}{\sigma^2}\right\} \quad (4)$$

σ 为核参数, 为了更精确、简便的求出最优参数, 采用遗传算法寻找最优参数组合, 利用 5-折交叉验证法将训练集等分为不相交的 5 个子集, 轮流选取 1 个子集做验证数据, 其余 4 个子集做训练数据, 根据训练出的模型把模型放在测试集上得到分类率, 共进行 5 次。计算每次的验证结果的平均值, 设置初始最优组合, 通过训练集训练学习, 从而求解得到最优分类函数^[10]:

$$f(x) = \text{sgn}\left(\sum_{i=1}^n y_i a_i^* K(x, x_i) + b^*\right) \quad (5)$$

2.3 SVM 一对多分类

对于多分类问题, 传统的一对一的算法将不能完成, 可以间接的利用一对一, 即采取一对多的算

法(也称一对其余算法), 每一类样本和对应其余样本之间产生一个最优决策面, 属于该类的样本点作为正类点, 对应剩余样本作为负类点。因此对 4 个类别需要构造 4 个 SVM 将该类的数据与其余类区分开^[11], 详见图 1。假定将第 $j(j=1, 2, 3, 4)$ 类样本看成正类, 其余类别均看成负类, 通过两类 SVM 方法求出一个决策函数:

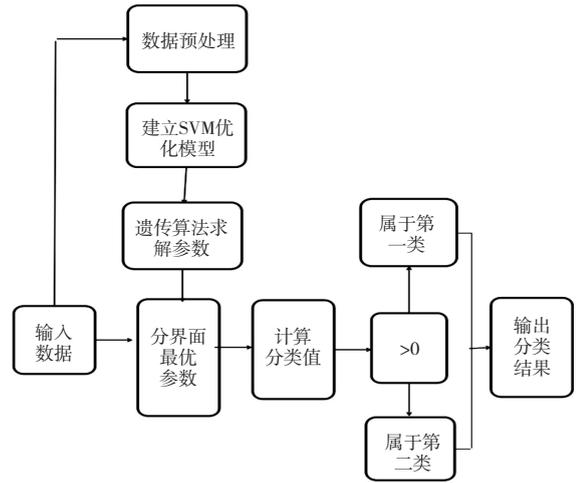


图 1 遗传算法优化 SVM 模型程序

Fig. 1 Optimization of SVM Model Program by genetic algorithm

$$f(x)_j = \text{sgn}\left(\sum_{i=1}^n y_i a_i^j K(x, x_i) + b^j\right) \quad (6)$$

共 4 个, 给定一个测试数据 x 输入, 当 $f_k(x) = \max_{j=1, 2, 3, 4} (f_j(x))$ 时, 则 x 属于 K 类。

3 实证分析

3.1 相关指标数据的选择

对于网络舆情危机选取 2019 年 9、10 月份的 10 个近期影响程度不同的事件, 分别为无锡高架坍塌、中通快递双十一涨价、波音延长停飞计划、NBA 总裁肖华发声、98 岁老人持刀杀妻、成都理工大学化学系研究生虐狗、本科生体育不合格不能毕业、新中国成立 70 周年大阅兵、李心草事件、新奥尔良在建酒店突然倒塌。下面将其记作事件 1, 事件 2, ..., 事件 10, 参照了相关文献, 充分考虑网络舆情的危机从舆情受众关注度、舆情的状态、传播过程的变化率、传播媒介的特征, 选取具有代表性的网络舆情指标 14 个。

3.2 利用因子分析对指标体系的整理

选择14个分析变量^[12],分别为 X_1 网络搜索量、 X_2 原创微博发布量、 X_3 转发量、 X_4 评论量、 X_5 观点倾向度、 X_6 舆情真实度、 X_7 内容直观度、 X_8 搜索变化率、 X_9 原创微博发布量变化率、 X_{10} 转发量变化率、 X_{11} 评论量变化率、 X_{12} 媒体知名度、 X_{13} 媒体权威度、 X_{14} 媒体参与数,相关数据可根据百度指数和微博获取,用KMO(Kaiser-Meyer-Olkin)检验统计

量)进行检验分析原始变量是否可以使用因子分析,经式(1)检验的测度值为0.715,表明变量间的相关性较强,适合做因子分析。进一步提取公因子,采用主成分分析法,选择合适公因子的数量来达到简化数据的目的,并用方差最大法(Varimax)进行正交旋转^[13],得出旋转因子使得公因子具有可解释性,见表1。

表1 特征根与方差贡献率表

Table 1 Feature root and variance contribution rate table

Component	Total Variance Explained								
	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	Variance/%	Cumulative/%	Total	Variance/%	Cumulative/%	Total	Variance/%	Cumulative/%
1	8.144	58.172	58.172	8.144	58.172	58.172	6.232	44.513	44.513
2	2.714	19.385	77.557	2.714	19.385	77.557	3.391	24.220	68.733
3	1.670	11.927	89.484	1.670	11.927	89.484	2.905	20.752	89.484
4	0.723	5.165	94.649	—	—	—	—	—	—
5	0.359	2.564	97.213	—	—	—	—	—	—
6	0.214	1.528	98.742	—	—	—	—	—	—
7	0.090	0.643	99.385	—	—	—	—	—	—
8	0.066	0.471	99.856	—	—	—	—	—	—
9	0.020	0.144	100.000	—	—	—	—	—	—
10	3.79E-16	2.707E-15	100.000	—	—	—	—	—	—
11	2.32E-16	1.659E-15	100.000	—	—	—	—	—	—
12	-6.90E-17	-4.934E-16	100.000	—	—	—	—	—	—
13	-3.60E-16	-2.575E-15	100.000	—	—	—	—	—	—
14	-4.21E-16	-3.010E-15	100.000	—	—	—	—	—	—

根据特征值的大小和因子的累计方差贡献率来确定公因子数量,一般取大于1的特征值,累计方差贡献率应达到70%以上。由式(2)显示结果前3个主成分方差累计率已达到89.484%(大于85%),对比旋转前和旋转后的特征根、方差贡献率及累计方差贡献率,前3个公因子的累计贡献率相同。分析可知选取3个公因子比较合适,可以很好地反映原指标近90%的信息,从而减少了网络舆情预警指标的数量,降低了预警成本,提高了预警科学性和准确性。

可以很清晰地知道3个公因子和14个原始变量之间的关系和变量间的关系。可得出: X_2 原创微博发布量、 X_3 转发量、 X_4 评论量、 X_9 原创微博发布量变化率、 X_{10} 转发量变化率、 X_{11} 评论量变化率这6

个变量距离较近,且在第一个公因子轴上的投影坐标较大,反映的是群众对事件的所思、所想、所说,可以命名为舆情主体因子; X_5 观点倾向度、 X_6 舆情真实度、 X_{12} 媒体知名度、 X_{13} 媒体权威度这4个变量距离较近,且在第二个公因子轴上的投影坐标较大,反映了与事件的相关度,可以命名为舆情客体因子; X_1 网络搜索量、 X_7 内容直观度、 X_8 搜索变化率、 X_{14} 媒体参与数这4个变量距离较近,且在第一个公因子轴上的投影坐标较大,反映了媒体对事件的转播,可以命名为舆情载体因子。因此指标体系中的14个指标可以降低成舆情主体因子、舆情客体因子、舆情载体因子就能充分体现原始数据的信息。由公式(3)计算出所选取的10个事件对于3个主因子的得分数据情况如表2:

表 2 因子得分情况

Table 2 Factor score

	1	2	3	4	5	6	7	8	9	10
舆情主体	-636.84	-153.94	-50.73	-684.48	-25.32	-27.55	-114.07	-881.91	-797.72	-65.95
舆情客体	-25 244.4	-5 224.17	-1 022.31	-26 349.92	-587.21	-767.10	-3 596.29	-33 020.93	-31 323.25	-2 069.71
舆情载体	17 596.90	3 525.70	524.08	18 481.73	28.20	145.24	2 263.22	23 561.95	22 119.20	1 185.90

3.3 网络舆情危机 SVM 模型等级分类辨识

对简化处理后的指标建立支持向量机模型,通过遗传算法利用 5-折交叉验证法将训练集等分为 5 份,轮流选取 1 份做验证数据,其余 4 份做训练数据,共进行 5 次运算得出核函数参数 $\sigma = 4.322$,惩罚系数 $C = 3.598 2^{[14]}$,利用 SVM 找出各类别的特征

样本与其余的特征样本的最优分类超平面式(6),将 1~7 事件数据作为训练样本集,8~10 事件数据作为测试样本集。以公因子舆情主体因子、舆情客体因子、舆情载体因子作为模型输入即表 2,进行网络舆情危机等级分类辨识如表 3 所示。

表 3 svm 训练集数据

Table 3 Svm training set data

序号	期望输出值	实际输出值($\times 100$)	误差/%	风险等级
1	(0,1,0,0)	(0.135,98.962,0.001,0.074)	0.45	B
2	(0,0,1,0)	(0.021,0.00,97.125,0.173)	0.41	C
3	(0,0,0,1)	(0.003,0.215,0.027,99.361)	0.52	D
4	(1,0,0,0)	(97.497,0.266,0.001,0.014)	0.48	A
5	(0,0,0,1)	(0.039,0.047,0.026,94.279)	0.57	D
6	(0,0,1,0)	(0.003,0.013,99.963,0.019)	0.42	C
7	(0,0,0,1)	(0.170,0.000,0.002,98.795)	0.46	D

由训练集的数据显示,以网络舆情的危机期望输出作为网络舆情危机辨识模型的输出,其中为了更清晰地知道实际输出值,对实际输出值扩大 100 倍的处理^[12],事件 1~7 的结果将训练好的网络舆情危机预警模型得到测试集的期望输出和实际输出,如表 4 所示,显然实际输出值的网络舆情危机风险等级与实际情况相符合,而发现期望输出值与实际输出值最大的相对误差为 0.51%^[15],具有较高的精度。

表 4 svm 测试集数据

Table 4 Svm test set data

序号	期望输出值	实际输出值($\times 100$)	误差/%	风险等级
8	(0,0,0,1)	(0.000,0.144,0.013,95.167)	0.51	D
9	(1,0,0,0)	(1.000,0.130,0.010,0.005)	0.35	A
10	(0,0,1,0)	(0.003,0.000,98.328,0.274)	0.47	C

分析发现网络舆情模型危机预警模型的训练和测试可以得到:在因子分析简化后的指标体系中建立支持向量机模型,将多余的网络舆情指标去除后会使模型更加简单。在接下来的支持向量积的模型中对参数的求解采用遗传算法得到的参数精度增加,且用一对其余的算法使得结果清晰明显。能确定每个事件在发生后对网络舆情所带来的危机等级程度,测试集的 3 个样本都得到了正确预警结果,说明模型的预警正确率很高,并可以得到其相对误差小于 0.51%。实验表明:所研究的因子分析对指标降维,遗传算法优化 SVM 的参数,用一对多的算法处理网络舆情危机等级分类的网络舆情危机预警模型是合理的、可行的、精准的。

4 结 语

通过对网络舆情危机的发展趋势进行预警,对

2019年的10个网络舆情事件进行实证分析得出基于因子分析和SVM的网络舆情危机预警模型在真实事件的数据反映下有小于0.51%相对误差。因子分析将网络舆情的14个指标进行降维为3个主因子,其次在简化的指标体系中用遗传算法的5-折交叉优化SVM参数,建立遗传算法优化SVM的网络舆情危机预警模型,最后将两类的SVM改进为一对多算法对4种情况进行分类,得出网络舆情的预警。

一是因子分析对指标的降维处理,将冗余的指标剔除,指标体系的简化,为模型的指标处理降低了很大的复杂度。二是在模型的算法中又利用遗传交叉算法求解SVM最优参数,提高SVM的学习能力,使模型的精确度得到保证。三是将两类SVM分类器推广为一对多的分类,训练出4个分类器,且分类速度相对较快。低于0.51%的误差预警充分说明了可行性,达到强化对网络舆情的监管。因子分析降低了指标体系的复杂性,遗传算法的5-折交叉提高了SVMS分类器的学习能力,能更准确地预测训练集,并用一对多算法使得分类速度较快。在简单、精准、迅速的基础上建立模型使预警的错误率能够很低,加强了网络舆情的监管,同时在化解和应对危机方面提供了帮助。

参考文献(References):

- [1] 王兰成,陈立富.国内外网络舆情演化、预警和应对理论研究综述[J].图书馆杂志,2018,37(12):6—15
WANG L C, CHEN L F. A Review of the Theoretical Research on the Evolution, Early Warning and Response of Network Public Opinion at Home and Abroad [J]. Library Magazine, 2018, 37 (12): 6—15 (in Chinese)
- [2] 郝楠,冯晶.基于模糊综合评价的网络舆情预警方法研究[J].重庆理工大学学报(自然科学),2019,12(8):49—56
HAO N, FENG J. Research on Early Warning Method of Network Public Opinion Based on Fuzzy Comprehensive Evaluation[J]. Journal of Chongqing University of Technology (Natural Science), 2019, 12 (8): 49—56 (in Chinese)
- [3] 汪东华.多元统计分析 with SPSS 应用[M].上海:华东理工大学出版社,2010
WANG D H. The Multivariate Statistical Analysis and the Application of SPSS [M]. Shanghai: East China

University of Technology Press,2010(in Chinese)

- [4] 王丹丹,晓兰.基于因子分析和聚类分析的内蒙古各盟市经济发展水平的研究[J].内蒙古统计,2016,417(4):16—17
WANG D D, XIAO L. The Study of the Economic Development Level of Each City in Inner Mongolia Based on Factor Analysis and Cluster Analysis [J]. Inner Mongolia Statistics,2016,417(4):16—17 (in Chinese)
- [5] 刘立宪.基于模糊支持向量机的火灾预警信号危险性辨识研究[J].轻工科技,2017,367(9):85—86
LIN L X. Research on Risk Identification of Fire Early Warning Signal Based on Fuzzy Support Vector Machine [J]. Light Industrial Technology,2017,367(9):85—86 (in Chinese)
- [6] 茆诗松,王静龙.高等数理统计[M].北京:高等教育出版社,2006
MAO S S, WANG J L. Higher Mathematical Statistics [M]. Beijing:Higher Education Press,2006 (in Chinese)
- [7] 赵丹丹,丁健臣.中国银行业系统性风险预警研究:基于SVM模型的建模分析[J].国际商务对外经济贸易大学学报,2019,273(4):97—101
ZHAO D D, DING J C. Research on Systematic Risk Early Warning of China's Banking Industry: Modeling and Analysis Based on SVM Model [J]. Journal of University of International Business and Foreign Economics and Trade,2019,273(4):97—101 (in Chinese)
- [8] 杜栋,庞庆华.现代综合评价方法与案例精选[M].北京:清华大学出版社,2015
DU D, PANG Q H. Modern Comprehensive Evaluation Methods and Case Selection [M]. Beijing: Tsinghua University Press, 2015 (in Chinese)
- [9] 毛志勇,黄春娟.砂土地震液化预测的GA-SVM-Adaboost模型[J].煤田地质与勘探,2019,512(3):19—21
MAO Z Y, HUANG C J. GA-SVM-Adaboost Model for Prediction of Seismic Liquefaction of Sand [J]. Coalfield Geology and Exploration, 2019, 512 (3): 19—21 (in Chinese)
- [10] 李锦绣.基于Logistic回归模型和支持向量机模型的多分类研究[M].武汉:华中师范大学,2014
LI J X. Multi-Classification Research Based on Logistic Regression Model and Support Vector Machine Model [M]. Wuhan: Huazhong Normal University, 2014 (in Chinese)

- [11] 刘雨康,张正阳. 基于 KNN 算法的改进一对多 SVM 多分类器 [J]. 计算工程与应用, 2015, 847 (24): 130—135
LIU Y K, ZHANG Z Y. Improved One-to-Many SVM Multi-Classifer Based on KNN Algorithm [J]. 2015, 847 (24): 130—135 (in Chinese)
- [12] 呼雨,陈新杰. 网络舆情检测及预警指标体系研究综述 [J]. 情报探索, 2012, 181(11): 7—10
HU Y, CHEN X J. A Summary of the Research on Network Public Opinion Detection and Early Warning Index System [J]. Information Exploration, 2012, 181 (11): 7—10 (in Chinese)
- [13] 汤永峰,卫志. 基于因子分析和支持向量机的电网故障风险评估 [J]. 电网技术, 2013, 101(4): 167—172
TANG Y F, WEI Z. Fault Risk Assessment of Power Grid Based on Factor Analysis and Support Vector Machine [J]. Power Grid Technology, 2013, 101 (4): 167—172 (in Chinese)
- [14] 孟金龙,丁超洋. 基于 SVM 的图像分类算法研究 [J]. 数字技术与应用, 2017, 725(10): 133—134
MENG J L, DING C Y. Research on Image Classification Algorithm Based on SVM [J]. Digital Technology and Applications, 2017, 725(10): 133—134 (in Chinese)

Research on the Early Warning for Online Public Opinion Crisis Based on Factor Analysis and SVM

ZHU Guang-ting, PAN Xiao-lin

(School of Mathematical Science, Chongqing Normal University, Chongqing 401331, China)

Abstract: According to index redundancy and high complexity of online public opinions which are not conducive to supervision, a comprehensive evaluation model based on factor analysis and SVM is proposed, in this method, 14 indicators for online public opinions are reduced into three common factors by factor analysis, then 5-fold cross of genetic algorithm is used to optimize SVM parameters in the simplified index system, the early warning model for online public opinion crisis by using genetic algorithm to optimize SVM is set up, finally, two kinds of SVM are improved into one-to-many algorithm to classify four cases, as a result, the early warning on online public opinions is obtained. The empirical analysis of 10 online public opinion events in 2019 shows that the early warning error is lower than 0.51 percent, which reveal that the model is feasible and which strengthen the supervision on online public opinions. Factor analysis reduces the complexity of index system, 5-fold cross of the genetic algorithm improves the learning ability of SVM classifier, thus, the model can more accurately predict training set, one-to-many algorithm makes classification speed more quickly, which provide the help for the supervision on online public opinions.

Key words: online public opinion; factor analysis; genetic algorithm; SVM; one-to-many

责任编辑:田 静

引用本文/Cite this paper:

朱光婷,潘晓琳. 基于因子分析和 SVM 的网络舆情危机预警研究 [J]. 重庆工商大学学报(自然科学版), 2020, 37(5): 94—100

ZHU G T, PAN X L. Research on the Early Warning for Online Public Opinion Crisis Based on Factor Analysis and SVM [J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2020, 37(5): 94—100