

文章编号: 2095-2163(2023)06-0072-06

中图分类号: TP181;C913

文献标志码: A

# 基于 XGBoost 的网络舆情文章预警模型建立

方茜

(贵州师范学院 数学与大数据学院, 贵阳 550018)

**摘要:** 现有的舆情预警模型主要是以指标体系为依据来综合评判预警的等级。研究中指标体系的建立及量化是通过专家经验知识获得,具有较强的主观性,缺乏以数据驱动的指标验证研究。针对以上问题,本文为网络舆情新闻文章设计了一套全面量化的预警指标,并使用特征筛选的方法,来验证所提指标对网络舆情文章预警的重要性,建立基于 XGBoost 的网络舆情文章预警模型,从而实现了对舆情文章预警判定。该模型与现有用于舆情预警建模的 SVM 和决策树两个方法进行实验对比,结果证明该模型在各项性能指标上均高于其他两种模型,对网络舆情文章预警等级判定更准确。

**关键词:** 网络舆情; 指标; XGBoost; 预警

## The development of early warning model of network public opinion articles based on XGBoost

FANG Xi

(School of Mathematics and Big Data, Guizhou Normal University, Guiyang 550018, China)

**[Abstract]** The existing public opinion early warning model is mainly based on the index system to comprehensively evaluate the level of early warning. However, in the current research, the establishment and quantification of the indicator system are highly subjective. The indicators are obtained through expert experience and knowledge, and there is a lack of data-driven indicator verification research. In view of the above problems, this paper designs a set of comprehensive and quantifiable early warning indicators for network public opinion news articles, and uses the method of feature screening to verify the importance of the proposed indicators to the early warning of network public opinion articles. Besides, an early warning model of network public opinion articles is established based on XGBoost to realize the early warning of public opinion articles. We compare our model with SVM and decision tree methods for public opinion early warning modeling. The results show that the model performs better than the other two models in various performance indicators, and the early warning level of network public opinion articles is more accurate.

**[Key words]** network public opinion; indicators; XGBoost; early warning

## 0 引言

随着互联网的快速发展,越来越多的网络应用成为广大民众畅所欲言的平台。绝大部分的网民主要从手机新闻 app、微博、论坛、微信公众号等平台获得资讯,而这些平台在提供资讯的同时也提供用户对文章进行评论、转发、点赞等互动功能。网络的开放性及自由性使得人们通过网络表达对文章报道事件的看法及心声,而这些看法及心声反映着民众的态度,同时也反映着社会舆情。然而当负面的舆情占比较大时,舆情危机产生,从而对社会造成重大影响,扰乱社会秩序,导致相关部门的管理失衡。

网络舆情预警是在预警对象传播的整个生命周期采取不间断的监测工作。通过实时采集预警对象的数据并对其分析,当达到监测阈值时,应使用预警模型及时反应预警对象的预警等级。预警使得一些可能产生舆情危机或已成为危机的信息能够被及时发现,由相关部门做出相应的引导决策,及时控制不良信息及不好舆论泛滥成灾。

## 1 相关工作

目前,国内的研究中提出了多种类型的舆情指标体系。其中,曾润喜等人<sup>[1]</sup>通过文献调研,将网络舆情监测与预警指标体系分为 4 类:

**基金项目:** 贵州省教育厅青年科技人才项目(黔教技(2022)258号)。

**作者简介:** 方茜(1993-),女,硕士,助教,主要研究方向:区块链技术、网络舆情、自然语言处理。

**通讯作者:** 方茜 Email: fqpaper@163.com

**收稿日期:** 2023-03-02

(1) 基于社会预警启示的指标, 从警源、警兆和警情 3 个一级指标来构建指标体系, 该方法的主观性较强;

(2) 基于主题分类的指标体系<sup>[2]</sup>, 从舆情事件产生背景及指向性设计指标体系, 然而该方法缺少指标量化研究;

(3) 基于舆情不同发生主体的指标体系, 从特定领域的角度设计指标体系(如高校)<sup>[3]</sup>等;

(4) 基于网络舆情内在机理的指标体系<sup>[4]</sup>, 使用舆情发展周期中出现的因素建立指标体系。国内已有舆情预警研究中, 有的将网民的情感极性<sup>[5]</sup>作为其中一个预警指标, 也有直接通过分析文本的情感极性<sup>[6]</sup>对舆情进行预警。

研究中发现, 基于指标体系的预警研究大部分还处于理论研究阶段, 实际应用价值不高, 并存在以下问题:

(1) 指标体系建立依赖专家经验知识, 主观因素较强, 缺乏数据驱动的客观实证性研究。

(2) 一些定性指标的量化依赖专家打分, 导致舆情预警综合研判滞后, 难以实现对预警对象的及时预警。

(3) 已有的基于指标体系的预警研究还处于理论期, 实用性较差。因此, 本文提出一套基于文章预警的特征, 通过相关性分析和特征筛选的方法, 对预定义文章的预警特征进行筛选, 并使用 XGBoost 方法建立文章预警模型。

## 2 网络舆情文章预警指标体系的建立

### 2.1 预选指标特征

本文以新闻文章为预警对象, 通过分析造成文章预警因素, 并结合以往舆情预警研究的指标体系综合分析, 从文章特征、受众特征两方面考虑, 列出对文章预警具有影响的特征作为预选指标。

#### 2.1.1 文章特征

文章特征从文章内容出发, 通常描述某一事件内容, 不同类型事件其影响程度不同。文章特征主要有以下几点:

##### (1) 文章主题重要性(topic\_importance)

文章主题重要性用于衡量这篇文章报道事件内容的重要程度。不同的主题所造成的影响程度不同, 通过提取文章的关键主题词, 关键主题词出现频率越高, 对文章的主题表现就越重要。因此, 统计该关键主题词频占文章中总词的频率作为文章重要性判断。

##### (2) 文章情绪(mood\_emotion)

文章情绪表示文本内容特征信息所表达出的情绪, 文章本身的情绪对阅读文章的网民情绪具有一定的影响。文章情绪可以使用 Lei J<sup>[7]</sup>提出的方法, 计算出文章的情绪标签, 并取概率值最大的情绪作为文章的情绪, 该值为离散值。

#### 2.1.2 受众特征

受众特征表示除文章本身信息外的其它信息, 其来源于网民在阅读完文章后对文章的一系列行为动作产生的数据。主要包括以下特征:

##### (1) 参与人数 $P$ (participants\_num)

参与人数用于表示阅读该文章的网民中参与评论总人数  $C$  及对评论点赞的人数  $Z$  之和, 也即是评论区参与总人数:  $P = C + Z$ 。文章的参与人数反映网民对这件事的关注程度, 可以通过爬虫直接从新闻页面获取得到。

##### (2) 评论总数 $C$ (comments\_num)

评论总数表示对这篇文章进行评论的总人数, 可以通过爬虫直接从新闻页面获取。评论越多, 表示网民对这件事的反映越强烈, 同时也会使该文章成为热门文章, 从而受到更多人的关注。

##### (3) 评论情感倾向 CE(comment\_emotion)

文章评论反映网民对文章报道事件的态度倾向, 情感倾向的正负极表示网民对这件事是支持还是反对。通过现有的情感分析技术判断评论情感得分, 用  $F\_score$  表示每条评论情感分值, 文章所有评论的情感倾向值:  $CE = (\sum_{i=1}^n F\_score) / C$ , 使用均值来表示。当负面评论较多时,  $CE < 0$ , 表示为 -1,  $CE \geq 0$  时, 表示为 1。

##### (4) 评论变化拐度

评论变化率是指单位时间内评论数的变化, 评论变化率  $C_{fre} = (C_{t_2} - C_{t_1}) / \Delta t$ ,  $\Delta t = t_2 - t_1$ 。其中,  $C_{t_1}$  表示单位时间  $t_1$  的评论数,  $C_{t_2}$  表示在单位时间  $t_2$  的评论数。通过一段时间内评论数的变化, 能够了解到在这段时间内网民对这篇文章的关注趋势变化。若评论数在某一时段内持续上升时, 应引起相关部门关注, 预防舆情危机产生。评论变化拐度用于描述评论数在前一时间段转变到后一时间段趋势变化的情况, 分别用前一时间段评论数变化率和后一时间段评论数变化率进行比较得出。通过分析, 拐度的 7 种情况见表 1。

从 G1~G7 穷举所有可能的拐度情况, G1~G3 表示相对于前一段时间, 下一时间段往评论变大的

方向拐;当两时间段内容评论变化率相等时,G4表示持平状态,下一时间段评论往持平方向;G5~G6表示相对于前一段时间,下一时间段评论往变小方向拐。通过分别统计一定时间段内出现评论变大、变小、持平的占比,来表示评论变化拐度情况。在表1列举的11种评论拐度图形中,任意一种图形出现一次则计数为1,所有评论拐度图形出现的次数之和为拐度总次数,用 $G_{total}$ 表示。

表1 评论变化拐度情况列表

Tab. 1 Comments on the change of the inflection situation list

序号	评论拐度图形	评论变化率比较	评论变化
G1		$\Delta C_{pre,j+1} > \Delta C_{pre,j} > 0$	
G2		$\Delta C_{pre,j+1} \geq 0, \Delta C_{pre,j} < 0$	增大
G3		$\Delta C_{pre,j} > \Delta C_{pre,j+1} \geq 0$	
G4		$\Delta C_{pre,j} = \Delta C_{pre,j+1} = 0$	持平
G5		$\Delta C_{pre,j+1} < \Delta C_{pre,j} \leq 0$	
G6		$\Delta C_{pre,j} < \Delta C_{pre,j+1} < 0$	减小
G7		$\Delta C_{pre,j+1} < 0 < \Delta C_{pre,j}$	

变大的占比(g\_up)为

$$\sum (G1 + G2 + G3) / G_{total}$$

持平的占比(g\_bal)为

$$\sum G4 / G_{total}$$

变小的占比(g\_down)为

$$\sum (G5 + G6 + G7) / G_{total}$$

(5)文章每种情绪投票占比

网民阅读完文章后,可以通过投票方式表达对此文章的情绪态度。总投票数和每类情绪的投票数可以通过爬虫代码从网页直接获取。使用每种情绪的投票占比来表示网民对这篇文章的实际情绪,其由每类情绪的投票数除以总投票数得到。情绪共有6种:分别为愤怒(angry)、震惊(shocked)、搞笑(funny)、新奇(novel)、感动(moved)、难过(sad)各投票数。

综上,本文预选的所有预警指标特征见表2。

用于文章预警的特征包含连续特征和离散特征,表2中离散特征名称中标注\*,其余为连续特

征。由于这些特征的量纲不同,其度量范围和数量差别较大。因此,为便于综合评价,减少数量级差异,本文使用min-max方法对特征进行归一化处理,使其值在[0,1]之间。

表2 文章预选指标特征列表

Tab. 2 Feature list of article pre-selected indicators

序号	特征	特征名称
f1	topic_importance	主题重要性 I
f2	mood_emotion	* 文章情绪 $E_{article}$
f3	participants_num	参与人数 $P$
f4	comments_num	评论总数 $C$
f5	comment_emotion	* 评论情感倾向 CE
f6	g_total	拐度总次数 $G_{total}$
f7	g_up	评论增大的占比 $G_{up}$
f8	g_bal	评论持平的占比 $G_{bal}$
f9	g_down	评论减小的占比 $G_{down}$
f10	angry	愤怒情绪投票占比 $p_{E\_agr}$
f11	moved	感动情绪投票占比 $p_{E\_mor}$
f12	shocked	震惊情绪投票占比 $p_{E\_shk}$
f13	funny	搞笑情绪投票占比 $p_{E\_fun}$
f14	novel	新奇情绪投票占比 $p_{E\_nov}$
f15	sad	难过情绪投票占比 $p_{E\_sad}$

## 2.2 指标特征分析

通过分析文章的相关信息,预选15个特征用于文章预警研究,本节使用皮尔逊相关性判断特征与预警相关性,通过相关性分析可以得到本文预选的预警特征指标是否与预警相关。

表3 文章预选特征指标与预警的相关性

Tab. 3 The correlation between pre-selected feature indicators and early warning

特征序号	相关性	特征序号	相关性
f1	0.003 12	f9	0.154 32 **
f2	0.762 47 **	f10	0.110 42
f3	0.337 21 **	f11	-0.294 91 **
f4	0.244 78 **	f12	0.554 67 **
f5	-0.671 89 **	f13	-0.330 19 **
f6	0.132 45 **	f14	0.303 18 **
f7	0.113 47 **	f15	0.102 28
f8	-0.193 11 **		

注:小数点后保留5位,\*为显著性 $p < 0.05$ ,\*\*为显著性 $p < 0.01$

由表3可知,本文预选的特征指标除主题重要性f1、读者愤怒情绪投票f10、难过情绪投票f15相关性最低且不显著以外,其它特征都表现较强的显

著性。主题重要性  $f_1$  相关性最低且不显著,其原因可能是因为该方法的量化是通过主题词在全文中出现的概率来反应主题重要性,不同文章通过该方法计算得到的概率值相差不大,导致每篇文章的主题重要性区分度不高。由于该特征与预警相关性小于 0.1,可筛除该指标。因此,通过上述特征分析,本文预选的特征基本与预警显著相关。

### 3 基于 XGBoost 的预警模型建立

本文目标旨在舆情监测过程中收集到一篇文章相关数据后,能够对其判断是否要对文章舆情预警,以及确定预警级别。参考《国家突发公共事件总体应急预案》<sup>[8]</sup>,其按突发公共事件产生的危害程度、可控性和影响程度、发展态势、紧急程度等因素,本文将文章预警等级分为 5 个等级,分别为特别严重(1 级)、严重(2 级)、较重(3 级)、一般(4 级)、不预警(0 级)。同时舆情监测预警系统中需要对监测的预警对象的预警级别及时准确反馈,以达到预警目的。因此,在模型的选择上需要使用一个速度快、准确率高的模型。

由 chen 等人<sup>[9]</sup>提出的 XGBoost (Extreme Gradient Boosting),因其速度快、准确率高等优点而受到广泛应用。XGBoost 方法是 GBDT (Gradient Boosting Decision Tree) 梯度提升树的改进版本,其与 GBDT 不同在于基学习器除决策树外还支持线性学习器,并加入正则项使得偏差与方差均衡。传统的 GBDT 在优化过程中只用到一阶泰勒展开,而 XGBoost 用到泰勒一阶展开和二阶展开。

梯度提升树是由 Boosting 方法结合 Gradient 梯度得到的。Boosting 是集成学习中的一种,其核心思想是通过迭代过程中前一轮的误差率动态更新训练集权重,每一轮都是一个弱学习器,由多个弱学习器集成强学习器实现回归和分类。而 GBDT 方法的每一轮迭代目标是找一棵决策树模型的弱学习器使得本轮的损失函数最小。GBDT 模型是由  $k$  个基学习器组成加法运算,可表示为

$$F(x_i) = \sum_{k=1}^K f_k(x_i), f_k \in F \quad (1)$$

其中,  $F$  表示所有基学习器组成的函数空间。对于  $n$  个样本,其损失函数为

$$L = \sum_{i=1}^n L(y_i, \hat{y}_i) \quad (2)$$

损失函数表示模型的偏差,最小化损失函数就是最小化模型的偏差。为了使模型的偏差和方差达

到较好的平衡,加入正则项来抑制模型的复杂度,因此模型的目标函数表示为

$$Obj = \sum_{i=1}^n L(y_i, \hat{y}_i) + \sum_{i=1}^n \Omega(f_k), f_k \in F \quad (3)$$

式中  $\Omega$  表示基学习器的复杂度,即模型的结构风险。

由于本文以决策树模型作为基学习器,因此可以使用树的深度、叶节点的个数等反应模型的复杂度。GBDT 前向分布算法的思想是从前往后建立基学习器,以此来逐渐优化,逼近目标函数  $Obj$  的过程。该过程以一个常数项开始,每次添加一个新的函数,其过程如下:

$$\hat{y}_i^{(0)} = 0 \quad (4)$$

$$\hat{y}_i^{(1)} = f_1(x_i) = \hat{y}_i^{(0)} + f_1(x_i) \quad (5)$$

$$\hat{y}_i^{(2)} = f_1(x_i) + f_2(x_i) = \hat{y}_i^{(1)} + f_2(x_i) \quad (6)$$

.....

$$\hat{y}_i^{(t)} = \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \quad (7)$$

上述过程中(0)~(t)表示从第 0 轮到第 t 轮,每一轮添加一个基学习器,主要在于 GBDT 的目标函数,即每个基学习器的加入都是以优化目标函数为目的。公式(7)表示第 t 轮的预测值计算,  $f_t(x_i)$  为要加入的基学习器,则此时的目标函数为

$$Obj^{(t)} = \sum_{i=1}^n L(y_i, \hat{y}_i^{(t)}) + \sum_{i=1}^n \Omega(f_i) = \sum_{i=1}^n L(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_i) \quad (8)$$

使用泰勒公式展开公式(8)中的目标函数。令

$g_i$  为目标函数中  $\hat{y}_i^{(t-1)}$  的一阶偏导,令  $h_i$  为  $\hat{y}_i^{(t-1)}$  二阶偏导,  $\hat{y}_i^{(t-1)}$  是 t 前一轮的训练结果,  $y_i$  是其对应的真实值,可作为常数处理,则由公式(8)转化为

$$Obj^{(t)} \approx \sum_{i=1}^n \left[ g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_i) \quad (9)$$

当对目标函数公式(8)只求一阶偏导时,模型为 GBDT,求一阶、二阶偏导时,模型为 XGBoost,即公式(9)为 XGBoost 的目标函数,损失函数不同对应着不同的  $g_i$  和  $h_i$ 。在 XGBoost 中复杂度  $\Omega(f_i)$  的公式为

$$\Omega(f_i) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (10)$$

其中,  $\lambda$  为学习率;  $T$  表示叶子节点个数;  $w_j$  为叶子节点的权重。  $\Omega(f_i)$  作为结构风险,将叶子节点的个数加入惩罚项,以限制模型的复杂度,并使用

L2 正则避免过拟合。

模型构建如图 1 所示,模型的输入为上节中通过特征分析后得到的文章预警特征指标,输出为文章预警级别。通过文章数据集提取文章数据,并量化各项文章预警特征,输入到 XGBoost 模型中,训练得到基于 XGBoost 的网络舆情文章预警模型。该模型的内部是由多棵 CART 树构成,一篇文章的预警等级由其所在的多棵树中的叶子节点权重共同确定,除第一棵树以外,每一棵树都是训练上一棵树的损失值,使得整个样本的损失尽可能减少,得到最优的分类结果。

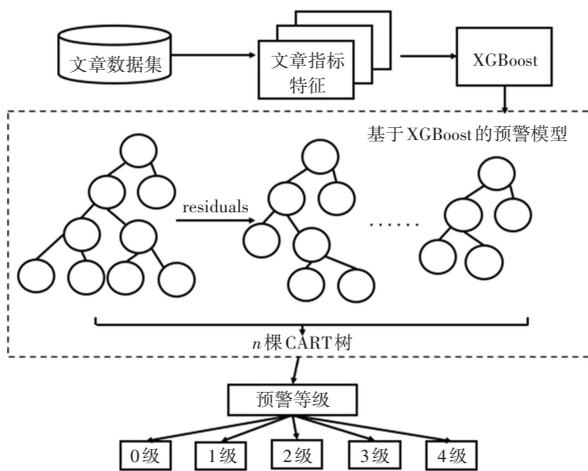


图 1 基于 XGBoost 的文章预警建模过程

Fig. 1 The modeling process of article early warning based on XGBoost

## 4 实验

### 4.1 实验数据

实验数据是使用爬虫代码,爬取新浪社会新闻网站情绪排行榜上的新闻数据而得,其中包括新闻标题、新闻内容、每篇文章对应的情绪投票结果及每篇新闻的评论数据。这些情绪分别为感动、震惊、搞笑、难过、新奇、愤怒 6 类情绪。选择 6 名网络舆情研究的学者,分别对收集到的 3 310 篇文章标记预警等级,通过 Fleiss Kappa 一致性检验后,得到  $Kappa = 0.758$  ( $Kappa$  系数划分为大于 0.75 为优秀, 0.40~0.75 为正常至良好, 低于 0.40 为差), 因此使用标记预警等级后的数据集作为实验的真实值。每类预警级别(从 1 级到 4 级预警严重程度逐级递减, 0 级表示不预警)的样本数见表 4。

表 4 每类预警级别样本数

Tab. 4 Number of samples for each warning level

预警等级	1 级	2 级	3 级	4 级	0 级
样本数	341	348	311	306	2 004

## 4.2 实验结果与分析

### 实验一 预警特征筛选结果及分析

通过 XGBoost 模型训练得到的指标特征重要性得分如图 2 所示。横轴表示特征重要性得分值 ( $F\_score$ ), 纵轴对应表 2 中的每个特征指标。

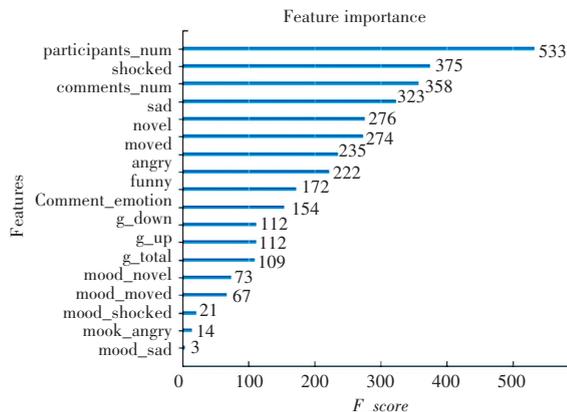


图 2 指标特征重要性

Fig. 2 The importance of features

可以看出,文章的参与人数 ( $participants\_num$ ) 具有最高的重要性得分,说明一篇文章的参与人数越多,越有可能需要预警。其次是网民对文章报道事件的震惊情绪 ( $shocked$ ),说明当看完文章的网民产生震惊情绪时越有可能需要预警。特征重要性得分排名前三的特征是衡量预警等级的最重要特征。而文章情绪特征重要性得分较低,说明其对预警等级的重要性较小。其中,文章情绪为难过 ( $mood\_sad$ ) 时最低,说明文章情绪表现为难过时,对预警等级的分类影响较弱。

### 实验二 基于 XGBoost 的文章预警模型效果及分析

表 5 展示了采用 3 种方法建立网络舆情文章预警模型的性能指标对比结果。其中 SVM 方法对预警等级的划分效果最差,而基于决策树的预警模型比基于 SVM 的预警模型效果要好。性能最好的是基于 XGBoost 建立的预警模型,其能够达到 77% 的平均准确率,且其  $F1 - measure$  的值能够达到 56.2%。说明本文提出使用 XGBoost 建立的预警模型能够很好的实现对网络舆情文章进行预警等级的分类。

表 5 三种模型性能对比

Tab. 5 Performance comparison of three models

模型	ACC	Recall	Precision	F1 - measure
决策树 <sup>[11]</sup>	0.69	0.476	0.475	0.475
SVM <sup>[10]</sup>	0.59	0.200	0.118	0.149
XGBoost(本文方法)	0.77	0.559	0.567	0.562

为了查看每个预警等级被分类准确的性能效

果, 本文选择准确率较高的基于决策树和基于 XGBoost 建立的预警模型的 ROC 曲线进行对比展示, 如图 3、图 4 所示。图中不同颜色代表不同的预警等级(0 级~4 级)的 ROC 曲线, 深蓝色的表示整体平均的 ROC 曲线。area 对应每个预警等级 ROC 曲线下的面积, 即 AUC 值, 该值越高, 表示预警模型越能判别出该预警等级。

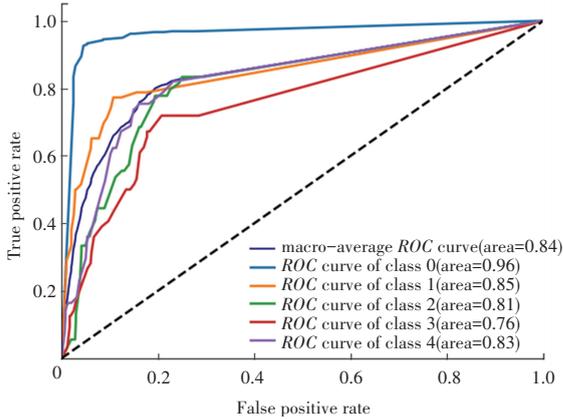


图 3 基于决策树的预警模型 ROC 曲线

Fig. 3 ROC curve of early warning model based on decision tree

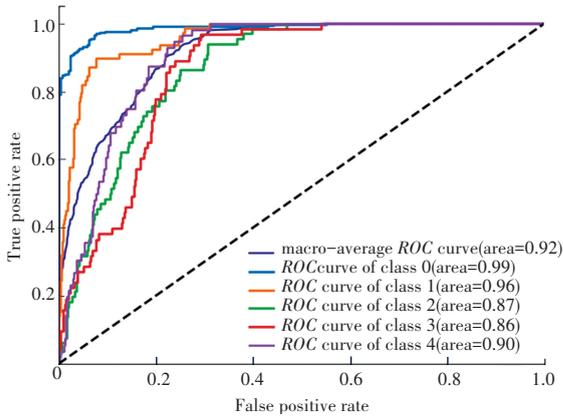


图 4 基于 XGBoost 的预警模型 ROC 曲线

Fig. 4 ROC curve of early warning model based on XGBoost

从图 3 和图 4 中可知, 基于 XGBoost 的预警模型在每个预警等级的判定命中率上均高于基于决策树的预警模型。其中, 宏平均 AUC 相差 8%, 除等级 0(无警)的 AUC 相差 3% 以外, 其他预警等级的 AUC 相差 6%~10% 之间, 说明基于 XGBoost 的模型更有助于预警等级的判定。

综上, 本文提出以数据驱动的方式客观验证提出的特征指标的有效性, 使用 XGBoost 建立网络舆情文章预警模型对文章预警等级分类在宏平均评估指标 ACC、F1、及 AUC 均取得了较好的效果, 显著优于使用 SVM 建立预警模型以及非集成的决策树模型, 且对文章预警等级为非常严重的判断明显优于

其他等级, 有效防止高严重级别的等级被误判为低严重级别。

## 5 结束语

针对现有研究中舆情预警多集中在以主观经验设计的指标体系为核心的静态预警模型研究, 指标体系的建立主观性太强, 依赖于专家经验, 缺乏基于数据驱动的实证研究。本文通过分析引起文章舆情危机的因素, 为新闻文章的预警等级的综合判定设计一套全面可量化的预警特征集, 以真实数据验证所提指标体系对舆情文章预警的重要性, 建立了全面可量化的文章舆情预警指标, 使得预警指标体系更具客观性。不仅解决了现有指标体系预警方法的主观性, 并提出使用集成方法 XGBoost 建立文章预警模型, 实现对文章预警级别的判定。通过对比实验的结果表明, 本文提出的方法明显优于现有的基于决策树的预警模型和基于 SVM 的预警模型, 能够有效实现对新闻文章预警等级的判断。

## 参考文献

- [1] 曾润喜. 网络舆情指标体系、方法与模型比较研究 [J]. 情报杂志, 2014, 33(4): 96-101.
- [2] 王晰巍, 邢云菲, 韦雅楠, 等. 大数据驱动下的社交网络舆情用户情感主题分类模型构建研究——以“移民”主题为例 [J]. 信息资源管理学报, 2020, 10(1): 29-38, 48.
- [3] Run Xi. Construction of early warning mechanism of network public opinion in colleges and universities under big data environment [J]. Academic Journal of Computing & Information Science, 2022, 5(2).
- [4] 王宝智, 黄萍萍, 叶坤葵. 网络舆情监测及预警指标体系构建与研究 [J]. 开封文化艺术职业学院学报, 2020, 40(6): 221-222.
- [5] 董光文, 袁健. 基于时间序列的方面级网络舆情情感演化模型 [J]. 智能计算机与应用, 2022, 12(12): 62-69.
- [6] 谭坤彦, 杨孔雨. 基于神经网络的舆情情感分析研究热点与趋势——基于 CiteSpace 的可视化分析 [J]. 智能计算机与应用, 2022, 12(8): 33-42.
- [7] LEI J, RAO Y, LI Q. Towards building a social emotion detection system for online news [J]. Future Generation Computer Systems, 2014, 37(7): 438-448.
- [8] 国务院. 国家突发公共事件总体应急预案 [J]. 安全与健康, 2006(7): 57-60.
- [9] CHEN T, GUESTRIN C. Xgboost: A scalable tree boosting system [C]//Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016: 785-794.
- [10] 张庆青. 基于支持向量机的网络舆情预警策略研究 [J]. 网络安全技术与应用, 2016(6): 63-64.
- [11] 徐建国, 刘梦凡, 刘慧慧. 突发事件网络舆情风险预警模型研究 [J]. 软件导刊, 2020, 19(7): 70-75.