

## 利用开源大数据的社会工作实时评估方法\*

张鑫, 欧文孝, 王寅森, 潘岩, 殷风景  
(国防科技大学 系统工程学院, 湖南 长沙 410073)

**摘要:** 鉴于目前社会工作评估评价方面的学术研究还较薄弱, 存在无针对性指标体系, 现有相关评比周期长、成本高等突出问题, 探索构建了利用互联网开源大数据的实时定量评估模型, 建立了针对性指标体系, 提出评分快速计算方法, 并设计实现原型软件。针对选定的社会工作进行案例实证, 通过采集某省14个市州的互联网真实数据来评估其工作力度, 结果与多位受访者的主观感受及相关官方评估基本相符, 验证了所提模型的有效性和合理性。

**关键词:** 定量评估; 开源大数据; 指标评分; 快速计算

**中图分类号:** C93   **文献标志码:** A   **文章编号:** 1001-2486(2022)04-213-10

## Real-time evaluation method of the progress of social work using open-source big data

ZHANG Xin, OU Wenxiao, WANG Yinsen, PAN Yan, YIN Fengjing

(College of System Engineering, National University of Defense Technology, Changsha 410073, China)

**Abstract:** Few quantitative models can be found amenable to the quantitative evaluation of social work, which motivates the presented work. Moreover, a few related existing works suffer from no suitable indicator system and high expenses of time and money. To overcome these limitations, a real-time quantitative evaluation model using the Internet open-source big data was developed, a target indicator system was established, the rapid computation methods of each criterion were proposed, and a prototypical software system was designed and implemented. Case studies on selected social work were conducted, and the work performance of the 14 cities belonging to a certain province in China was evaluated. Results achieved were largely consistent with the subjective feelings of many respondents and relevant official assessment, which validated the efficiency and rationality of the proposed model.

**Keywords:** quantitative evaluation; open-source big data; criterion evaluation; rapid computation

当前,很多社会工作的评估评价还相对薄弱。以拥军优属工作为例,其评价主要存在三方面问题:一是尚无独立的针对性评估体系。已有的考评标准,如《全国双拥模范城(县)考评标准》(下文简称“考评标准”),同时涵盖(政府和群众)拥军优属和(军队)拥政爱民两方面,其中很多指标二者兼顾,难以剥离,只能算是“相关”评估,针对性不够强。二是现有相关评估的间隔长、成本高。一般两年或四年才开展一次,无法随时评、时时评,不利于依据评估或抽检结果及时改进工作,也容易引发“重评时、轻平时”的问题。而且,参评城市需要按标准整理提报材料,评审方还需要组织专班审核材料、打分评比以及开展平时抽检,双方都需要消耗大量人力物力,工作成本较高。三是相关评估的部分评价指标有待优化,具体表现

在:①有些指标只考察有无,未衡量多寡,区分度不够,容易导致“干多干少一个样”;②很多指标需依据政府部门掌握的内部资料,导致详细评分不便公开,不利于显示评估结果的公信力;③部分指标比较综合,涵盖教育、文化、法律、医疗、住房等多领域,不便参评城市之间细化比较,各自发现优势和弱势领域,以便下步针对性地固强补弱。

在现今互联网时代,社会工作相关的通知公告、新闻报道、体验评价等大多能见诸网上,既给相关群体获取信息带来便利,也给利用开源大数据评估这些工作的推进力度提供了可能。鉴于此,研究人员从2014年左右开始尝试利用大数据完成不同评估任务,包括:风险评估<sup>[1-5]</sup>、信用评价<sup>[6-8]</sup>、传播或宣传效果评估<sup>[9-11]</sup>、政策评

\* 收稿日期:2021-09-09

基金项目:国家部委基金资助项目(41412060502)

作者简介:张鑫(1978—),男,四川南江人,研究员,博士,硕士生导师,E-mail:shinezhang\_nudt@qq.com

估<sup>[12-15]</sup>、保健水平评估<sup>[16]</sup>等。通常,评估对象不同,构建的指标体系就不同,所需的数据也不同。比如,评估消费者个人信用指数时,Jiang等<sup>[6]</sup>使用了消费者在信息咨询平台上的问询记录、跨平台失信信息、线上购物信息等;评估公共政策时,则可使用政策文件数据、互联网文本数据(包括新闻、社交帖文)、电子商务数据等<sup>[15]</sup>。但是,与传统利用领域(小)数据进行评估不同,大数据中噪声多、价值密度低,如何快速获取相关数据,并从中提取真正有用信息,是基于大数据的评估方法所面临的共同挑战,高效的数据采集获取和分析挖掘必不可少。值得一提的是,受困于这一挑战,有相当一部分现有研究(如<sup>[8-9, 13-15]</sup>)还处于理论分析和方法构想层面,尚未实际获取大数据、针对所关注评估对象具体开展评估实证。另外,尽管现有研究已经探索了利用大数据来评估很多不同任务对象,但还未发现利用它们来评估拥军优属等社会工作的相关成果被公开报道。

为此,本文以拥军优属作为案例,提出一种利用开源大数据的社会工作实时评估方法,建立综合网络信息特点与指标可测性(即能否利用开源数据测量计算)的分层指标评估体系;设计开发了原型软件,针对给定的参评对象,通过持续爬取其网上信息,以增量方式实现指标得分与综合评分的快速计算。

## 1 评估模型构建与计算

定量评估是指依据统计数据,建立数学模型,并用模型计算评估对象各项指标得分来实现评估分析的方法。其中,“数学模型”即定量评估模型,其构建往往需要:首先确定左右评价的影响因素,其次建立指标体系(包括确立指标、定义指标测算方法、明确指标间的关系和权重等),最后定义相应的综合评分方法或公式。其中,指标筛选可用定性方法(如德尔菲法<sup>[17]</sup>),也可用定量方法(如主成分分析<sup>[18]</sup>或其改进方法<sup>[19]</sup>);指标测算方法和指标关系分析因任务而异,通常需要人工参与;指标赋权则可使用启发式方法(如层次分析法或其改进<sup>[20]</sup>)或机器学习方法(如逻辑回归、支撑向量回归或其他神经网络<sup>[21]</sup>),前者往往离不开人工参与,而后者则需要训练样本,即事先获取的、通常由人工标注的评估样本(这些样本既有指标得分,又有综合评估结果)。此外,虽然机器学习中有增量学习<sup>[22-23]</sup>(incremental learning)的相关研究,但主要关注如何针对新增

样本,或类别、属性快速更新模型参数,而聚焦于新的数据不断获取之后,如何以增量方式快速计算指标得分与总评分的研究成果还比较少见。

以现实生活中的拥军优属工作评估为例,从现有相关考评标准中遴选拥军优属相关指标作为基础,针对网上开源信息的特点,重点从指标可测性和区分性两方面对它们加以调整,借此构建基于开源大数据的拥军优属工作评估指标体系。

### 1.1 评估指标筛选与体系构建

最新版考评标准一共设置了10大考评项目(共78个评分项),分别是:①组织领导坚强有力;②宣传教育广泛深入;③拥军工作扎实有效;④拥政爱民成果显著;⑤政策法规落到实处;⑥双拥活动坚持经常;⑦军民共建富有成效;⑧军政军民关系融洽;⑨群众满意度测评;⑩加分项目。其中,与拥军优属相关的评分项主要分布在①、②、③、⑤、⑥、⑨考评项目中,可归为四大方面:组织领导、宣传教育、政策制定与落实、相关活动开展。下面围绕它们来讨论本文评估指标体系建立。

考评标准中“组织领导”方面的评分项主要考察双拥办等机构设置情况和相关经费是否列入预算。对此,通过互联网检索结果可发现,几乎所有省(自治区)辖市州和(直辖市)市辖区都设置了相关机构,并定期召开党委议军、双拥促进等会议;且从所公开的预算信息中也都能找到双拥和优抚专项预算。所以,考虑到区分性,本文不设立“组织领导”方面的拥军优属评估指标。

“宣传教育”相关评分项主要考察各地线上、线下宣传开展情况。网络时代,互联网已成为舆论宣传主阵地,无论从形式多样性、内容生动性、受众覆盖面,还是从成本控制、节能环保等角度,线上宣传都具有线下不可比拟的优势。因此,本文主要通过爬取各地相关单位的互联网信息,评估其宣传教育开展情况。

“政策制定与落实”方面,考虑到政府部门会在网上颁布相关政策制度,且一经公布通常都会得到较好执行。拥军优属政策往往涉及就业、教育、住房等不同领域,政策覆盖的领域越全,说明当地对优抚群体的照顾越周到,单个领域的政策越多,则往往反映当地在该领域的优抚力度越大。因此,本文从网上采集各地相关政府部门颁布的政策规范类文件,并通过考察它们的领域覆盖面和各领域的政策数量来评估当地拥军优属政策制

定与落实情况。

“相关活动开展”与“宣传教育”类似,各地通常都通过互联网发布新闻动态,其中常包含拥军优属活动(如悬挂光荣牌、走访慰问、组织定向招聘等)开展情况。鉴于此,通过抓取各地相关单位的网上新闻,然后进行事件分析和计量以评估当地相关活动开展情况。

基于以上分析,本文围绕宣传教育、政策制定与落实、相关活动开展三方面,构建如图1所示的评估指标体系。其中,鉴于宣传教育和活动开展情况主要通过相关单位的网上新闻动态发布,所以将二者合并,设立一级指标“新闻动态”,其下不设立二级指标(或者说二级指标是它自身)。政策制定与落实方面,鉴于教育、医疗、住房、就业堪称最重要的民生领域,所以划分5个预定义领域,即上述4个和“其他”,后者涵盖不能划归前4个领域的所有政策文件。相应地,设置1个一级指标(“政策制度”)和5个二级指标(即“教育”“医疗”“住房”“就业”“其他”)。此外,通过与军人军属访谈了解到,各种优先优惠措施(包括车站、机场、医院等公共场所的优先通道设置,酒店、餐饮等民间优惠,以及景区的票价减免等),往往直接左右他们对于一个地方拥军优属力度的观感评价。而且,这些优惠措施也是一种落实于行动的、广大民众更常接触的社会性拥军优属宣传教育。考虑到其重要性,本文单独设立1个一级指标“优先优惠”,包含“优先通道”“民间优惠”“景区优免”3个二级指标。

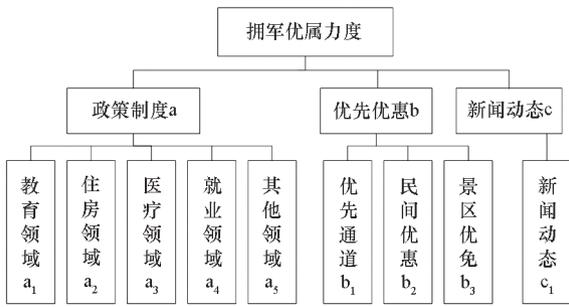


图1 模型评价指标树

Fig. 1 Hierarchy of the evaluation indices

### 1.2 指标赋权

由于上述指标体系中各类二级指标数量不多且相互关联性不明显,所以采用层次分析法(analytic hierarchy process, AHP)进行指标赋权。鉴于考评标准中考评大项和评分项(见1.1节)的赋分体现了双拥领域权威专家对这些项目重要性的集体评判,本文以它们作为确定指标权重的

重要参考,考虑如下三方面:①政策制度是政府颁布的拥军优属法律文件,最能直接反映其拥军优属工作情况,也最具权威性,所以该一级指标权重最高。本文认为教育、住房、医疗、就业和其他这5个民生领域对于军人军属而言同等重要,所以设置相应各二级指标重要性相同。②优先优惠体现了企事业单位拥军优属落实情况和响应力度,往往让军人军属直接有感,不过拥军优属工作并不仅限于民间优先优惠,因此该一级指标应有较大权重,但应略低于政策制度。此外,一个地区的旅游景点数量受地理位置、资源环境等非人力因素限制,为降低其对评估结果公平性的可能影响,设定二级指标中景区优惠的重要性要略低于其他两个指标。③新闻报道活跃度既可反映当地政府对拥军工作的重视程度、宣传力度,也可反映其拥军工作落实情况。但上新闻的往往是当地一些相对重要的工作,而细微工作,包括很多基层活动(如某军供站为退休干部发放防疫物资)时常并未涵盖其中,因此,“新闻动态”的权重应比前两个一级指标更低。

基于上述分析,本文分别构建了一级指标和“优先优惠”下属的二级指标判断矩阵(判断矩阵中的数值表示行列对应指标的相对重要性关系)<sup>[4]</sup>,计算各指标权重并分别进行一致性检验,结果分别如表1和表2所示。

表1 一级指标判断矩阵

Tab. 1 Comparison matrix of the 1st level criteria

一级指标	政策制度	优先优惠	新闻动态	权重
政策制度	1	2	3	0.539 0
优先优惠	1/2	1	2	0.297 2
新闻动态	1/3	1/2	1	0.163 8

表2 二级指标判断矩阵

Tab. 2 Comparison matrix of the 2nd level criteria

二级指标	优先通道	民间优惠	景区优免	权重
优先通道	1	1	2	0.400 0
民间优惠	1	2	2	0.400 0
景区优免	1/2	1/2	1	0.200 0

经计算表1~2中两个判断矩阵的一致性比例(consistency ratio, CR)值分别为0.007 939和0,均小于0.1,一致性检验通过。进一步将两层评价指标判断矩阵计算的权重结果相乘,可算出所有二级指标的全局权重<sup>[4]</sup>,如表3所示。

表 3 二级指标全局权重计算表

Tab. 3 Weights of the second level evaluation indices

二级指标	政策制度	优先优惠	新闻动态	全局权重
	0.539 0	0.297 2	0.163 8	
教育	0.2			0.107 80
住房	0.2			0.107 80
医疗	0.2			0.107 80
就业	0.2			0.107 80
其他	0.2			0.107 80
优先通道		0.4		0.118 88
民间优惠		0.4		0.118 88
景区优先		0.2		0.059 44
新闻动态			1	0.163 80

### 1.3 指标评分方法

考评时间点  $t$  和考评周期  $H$  不同, 参评城市  $l$  的评估得分往往也不同。因此, 如果对照图 1 中的顺序, 将 9 个二级指标的原始得分分别记作  $N_1, \dots, N_9$ , 则它们都应是  $t, H$  和  $l$  的函数。

#### 1.3.1 政策制度类指标

指标得分  $N_1 \sim N_5$  的测算公式定义为:

$$N_i(l, t, H) = PF(l, t, H, i), i = 1, \dots, 5 \quad (1)$$

其中,  $PF(l, t, H, i)$  表示城市  $l$  在相应时段内颁布的属于第  $i$  个民生领域的拥军优属政策数量。这些政策制度通常发布在政府部门的“通知公告”和“规范性文件”栏目中, 可分为两大类: 一类是长期政策, 在新版本出来前长期有效, 且新版本通常是对旧版的部分修订而非颠覆性改变; 另一类是短期政策, 一般只在明确规定的时段内有效。通过数据观测发现, 短期政策的显著特点是其标题中包含“年度”或“年”, 且前面有数字或“本”字。据此, 本文开发网页爬虫, 从城市  $l$  的政府机构网站的“通知公告”和“规范性文件”栏目持续采集相关政策文件, 去重后借助文本分类算法将每个文件划分到 5 个预定义领域中。同时, 通过在标题中检测关键词组合“数字/‘本’ + ‘年/年度’”来判别文件是否为短期政策, 其中单引号引起的是关键词。若为短期政策, 则进一步提取其网页发布时间作为发布时间, 并统计发布时间在从  $t$  开始倒推的  $H$  时间内属于第  $i$  个民生领域的短期政策数量。对于长期政策(即未被判定为短期政策的那些), 则取最早版本时间作为发布时间, 并统计截至时间  $t$  范围内属于第  $i$  个民生领域的所有政策数量, 理由是: 长期政策在理论上到任

意时间  $t$  都有效, 且同一政策的新旧版本已被去重。最后, 两部分求和可得到  $PF(l, t, H, i)$ , 即

$$PF(l, t, H, i) = PF_{perm}(l, t, i) + PF_{temp}(l, t, H, i) \quad (2)$$

其中,  $PF_{perm}(l, t, i)$  是长期政策数量 ( $i = 1, \dots, 5$ ), 与参数  $H$  无关;  $PF_{temp}(l, t, H, i)$  是短期政策数量, 与  $H$  相关。实验中发现, 短期政策多见于就业、住房和教育三个领域。

#### 1.3.2 优先优惠类指标

由于优先优惠一经设立, 就鲜有取消的情况(除非相应的公共设施关停), 所以测算这 3 个指标得分时, 不需要考虑时间段  $H$ , 也即  $N_6 \sim N_8$  分别对应于指定城市  $l$  到  $t$  时刻截止, 其优先通道、民间优惠、景区优免的设置或实施情况。其中, 优先通道和景区优免指标评分的思路是, 辖区内公共场所或景区中设置优先通道或票价减免的占比越高, 得分应越高, 形式化定义为:

$$N_i(l, t, H) = N_i(l, t) = P_i(l, t) / Q_i(l, t) \quad (3)$$

式中,  $i = 6, 8$ , 第一个“=”是为了强调指标评分与参数  $H$  无关, 右侧分式中分母  $Q_6(l, t)$  表示城市  $l$  辖区内截至  $t$  的公共场所总数、 $Q_8(l, t)$  表示景区总数, 而分子  $P_6(l, t)$  或  $P_8(l, t)$  则分别表示设置了优先通道的公共场所数或有票价减免的景区数。为计算  $N_6$ , 本文一方面采集微信小程序“军人优先查询助手”上的优先信息, 以城市  $l$  的名称作为关键词筛选其中相关部分, 计算  $P_6(l, t)$ ; 另一方面以“ $l$  + 机场/高铁站/地铁站/车站/医院”作为关键词检索百度, 对检索结果去重后计量得到  $Q_6(l, t)$ 。对于景区优惠评分  $N_8$ , 本文从旅游网站“途牛旅游网”(http://www.tuniu.com) 上检索和爬取城市  $l$  辖区内的景点信息, 统计  $Q_8(l, t)$ , 并通过匹配关键词“军人/军属/军官/士兵”筛选和统计其中有票价减免的景区数量  $P_8(l, t)$ 。

民间优惠评分  $N_7$  略有不同: 一方面很难找到相对统一的获取各地酒店饭店总量的途径; 另一方面很多城市, 特别是大城市的酒店饭店数量众多, 且不时有新开和关闭的, 总量波动很大。考虑到酒店数量与当地人口数量(包括固定与流动人口)有较强相关性, 本文定义如下公式测算民间优惠力度:

$$N_7(l, t, H) = N_7(l, t) = P_7(l, t) / M(l, t) \quad (4)$$

同样地,  $N_7$  与时段参数  $H$  无关。式中, 分子表示城市  $l$  截至时间  $t$  范围内设有军人军属优惠的饭店酒店总量, 本文通过从微信小程序“军人优先查询助手”中采集相关信息并匹配城市  $l$  后

统计获得;而分母  $M(l,t)$  表示同期城市  $l$  的人口总量,该数据通过网络来源获取。

### 1.3.3 新闻动态类指标

由于新闻文章及所报道的言行活动具有很强的时效性,所以  $N_9$  同时依赖于  $t$ 、 $H$  和  $l$ 。此外,在我国,现代化程度越高、网媒越发达的城市,通常人口总量也越大。而一些三四线城市,虽然其拥军优属工作干得不少,但可能因为网媒发达程度不及,导致网上新闻动态不多、宣传力度不够。为降低网媒发展不平衡可能给本指标测算带来的不利影响,本文定义  $N_9$  的测算方法为城市  $l$  在从  $t$  开始倒推的  $H$  时间内的新闻动态总量再除以其相应时间的人口总量,即:

$$N_9(l,t,H) = R(l,t,H)/M(l,t) \quad (5)$$

其中,  $R(l,t,H)$  表示相应时段的新闻事件计量,  $M(l,t)$  的含义和计算方法与式(4)相同。本文采用如下步骤测算  $R(l,t,H)$ :①采集城市  $l$  及其下辖区县的政府和退役军人事务管理机构网站上“政务动态”或“工作动态”栏目下的文章;②利用“部队/驻军/官兵/拥军/军属/烈士/双拥”等关键词筛选其中与拥军优属有关部分;③对所选文本去重;④通过无监督聚类从去重后的文本中发现事件或者说话题(event/topic),并以每个事件聚簇中最早的文章发布时间作为事发时间;⑤统计在从  $t$  开始倒推的  $H$  时间内发生的事件数量,即  $R(l,t,H)$ 。

### 1.4 模型实时计算

式(1)~(5)定义了各二级指标的原始分测算方法,显然它们的取值范围很不一样。为避免量级不同而影响各自在全局评分中的占比,需要对原始分进行归一化,折算为标准分。思路是,对每个  $N_i(l,t,H)$ ,取参评城市中的最高分和最低分,并按如下方式折算为百分制下的标准分,即:

$$S_i(l,t,H) = \frac{N_i(l,t,H) - N_i^{(\min)}}{N_i^{(\max)} - N_i^{(\min)}} \times 100 \quad (6)$$

其中,  $N_i^{(\max)} = \max_l \{N_i(l,t,H)\}$ 、 $N_i^{(\min)} = \min_l \{N_i(l,t,H)\}$  分别是所有参评城市第  $i$  个指标原始分的最大值和最小值。若将 1.2 节中各二级指标的全局权重分别记作  $W_1, \dots, W_9$ ,则城市  $l$  在从  $t$  开始倒推  $H$  的考评期内的总评分为:

$$S(l,t,H) = \sum_{i=1}^9 W_i \cdot S_i(l,t,H) \quad (7)$$

但是,直接按上述流程进行计算会导致参数  $t$  或  $H$  稍有变化,就需要从网上信息采集开始重来一遍,将非常耗时。而事实上,改变这两个参数又非常必要。因此,本文持续爬取各类网上数据,并

定期(如每天或每月)计算和保存式(1)~(5)中  $PF$ 、 $P$ 、 $Q$ 、 $R$  等统计量,然后利用这些定期统计结果,实现任意时间点和考评周期内评估分数的快速计算。记固定统计间隔为  $T$ ,设  $H = KT$ ,则对于任意  $t$  和  $H$ ,根据式(2)可得:

$$\begin{aligned} PF(l,t,H,i) &= PF_{\text{perm}}(l,t,i) + PF_{\text{temp}}(l,t,H,i) \\ &= PF_{\text{perm}}(l,t,i) + \sum_{k=0}^{K-1} PF_{\text{temp}}(l,t - kT, T, i), i = 1, \dots, 5 \end{aligned} \quad (8)$$

式中,  $PF_{\text{perm}}(l,t,i)$  表示城市  $l$  截至时间  $t$  范围内颁布的属于第  $i$  个民生领域的长期政策的数量,  $PF_{\text{temp}}(l,t - kT, T, i)$  表示每个统计间隔内短期政策的数量。由于长期政策一直有效,所以其计算方法相对特别:

$$PF_{\text{perm}}(l,t,i) = \sum_{k=0}^{\infty} PF_{\text{perm}}(l,t - kT, T, i) \quad (9)$$

其中,  $i = 1, \dots, 5$ ,  $PF_{\text{perm}}(l,t - kT, T, i)$  是从  $t$  开始倒推的第  $k$  个统计间隔内发布的属于第  $i$  个民生领域的长期政策数量;求和上界为正无穷,表示从城市  $l$  可获得的长期政策文件中发布时间最早的开始统计(显然不早于新中国成立的 1949 年)。还需注意的是,若一个长期政策有多个版本,则只有最早版本应计入相应统计间隔的  $PF_{\text{perm}}(l,t - kT, T, i)$  中,余下版本不再重复计入。这样,只要计算并缓存每个固定统计间隔  $T$  中的长、短期政策数量,即可通过查表后求和快速计算任意  $t$  和  $H$  下的指标得分。

余下 4 个指标(原始分)的快速计算方法与此类似,不再赘述。但要注意,优先优惠下的 3 个二级指标的计量方法应参照上面长期政策数量的计算过程,而新闻动态的计量方法则应参照短期政策数量的计算过程。

最后简析一下计算复杂度。实验中发现,能够获取的最早数据的发布时间为 1999 年,所以式(9)中的求和上界实际与考评周期内包含的统计间隔数  $K$  数量级相近。因此,设有  $m$  个城市参评,则按上述快速方法对它们进行综合评估的计算复杂度仅为  $O(Km)$ 。 $K$  和  $m$  的取值通常均在万以下,对于现代计算机而言,很容易做到实时计算,即用户提出评估需求后,在秒以内返回所有参评城市的评估结果。

## 2 算例与分析

为验证上述模型方法的可行性与有效性,接下来首先选择若干城市作为评估对象,接着针对性设计和开发原型系统,实现所需数据的持续采

集、处理与分析,然后利用这些数据实施评估,并对结果进行分析和讨论。

### 2.1 评估对象选取与原型实现

本文选取 H 省所辖 14 个市州作为评估对象,主要考虑两方面:一是同一省份的市州之间通常具有较好的可比性,二是课题组及周边对 H 省熟悉的人相对较多,有利于判断模型评估结果的合理性。接下来将以这些市州的首字母指代它们,当首字母相同时,则在后面加上数字加

以区分。这样,H 省 14 个市州分别为 C1、Z1、X1、H1、S、Y1、C2、Z2、Y2、C3、Y3、H2、L 市和 X2 州。

原型软件的模块构成及工作原理如图 2 所示(其中粗箭头线表示控制流、单线箭头表示数据流),除“评估计算”外各模块的技术途径见表 4。其中,数据“预处理”包括两个子模块,“过滤”主要:①对政策文件,通过在标题中检测“转发”“‘转’+组织名+‘通知/文件’”等关键词组合,

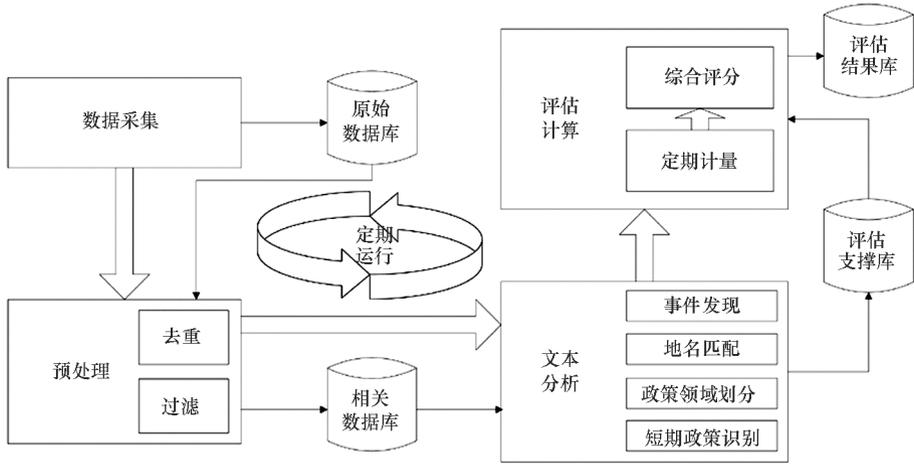


图 2 原型工作流程示意

Fig. 2 Workflow of the prototype

表 4 原型模块技术途径一览表

Tab. 4 Techniques used by the components in the prototype

一级指标	数据需求	数据采集		预处理	文本分析
		数据来源	爬虫技术		
政策制度	政策通知文件	政府部门网站或网页的“通知公告”和“规范性文件”栏目	网页爬虫	①过滤 ②去重	①短期政策识别:基于规则 ②领域划分:有监督分类
	优先通道信息	①公共设施网站 ②军人优先查询助手	①百度元搜索 + 站内元搜索 + 网页爬虫 ②网络抓包与解析	去重	地名匹配:基于扩展地名词典
优先优惠	酒店饭店优惠信息	军人优先查询助手	网络抓包与解析		①地名匹配:基于扩展地名词典 ②关键词匹配:“军人/军属 + 优惠”
	景区优惠信息	途牛旅游网	站内元搜索 + 网页爬虫		①地名匹配 ②关键词匹配:“军人/军属 + 优惠”
新闻动态	新闻文章	政府和退役军人事务管理机构网站上“政务动态”或“工作动态”栏目	网页爬虫	①过滤 ②去重	事件发现:无监督聚类

滤除非本级文件;②对新闻文章,利用“部队/驻军/官兵/拥军/军属/烈士/双拥”等关键词,筛选并保留与拥军优属有关的那些。“去重”旨在去除从多个不同来源采集的相同文章,本文以 SimHash 作为文本表示,通过在线 Single-Pass 无监督聚类实现去重<sup>[24]</sup>。“文本分析”模块:①基于关键词匹配实现短期政策文件识别,见 1.3.1 小节;②基于 TextCNN 模型<sup>[25]</sup>、通过有监督文本分类实现政策制度的领域划分;③通过构建 H 省的扩展地名词典(在其中纳入每个市州的名称、简称和别名)并利用它识别采自“军人优先查询助手”小程序和“途牛旅游网”优先优惠信息的城市归属;④以标题和正文中词条的 word2vec 预训练词向量<sup>[26]</sup>的加权平均作为文本表示,采用在线 Single-Pass 算法实现新闻动态中的事件发现<sup>[27]</sup>。需要强调的是,图 2 中“数据采集”“预处理”“文本分析”等 3 个模块,以及“评估计算”中的“定期计量”子模块,以定时运行方式(如每 12 小时一次),不断采集参评市州的最新数据,经预处理和文本分析后,定期(根据式(8)中参数  $T$  的设置,每天或每月一次)统计计算各二级指标的原始分。当需要评估时,再驱动“评估计算”中的“综合评分”子模块计算所有参评城市在指定时间点和考评周期上的评估得分。

原型自 2020 年 6 月开发完成后即联网部署运行,截至本文完成时,共采集到 H 省相关信息 97.23 万条,这些信息中发布时间最早的为 1999 年 11 月 17 日,经过滤和去重后得到 7 262 条,远少于原始数据,这种量级差异正好符合大数据价值密度低的特点。

## 2.2 评估结果及分析

### 2.2.1 不同考评周期的评估结果

表 5 给出了 H 省 14 个市州以 2020 年 12 月作为考评时间点,考评周期分别为 2 年和 4 年的考评结果。下面从两个不同角度分析评判表 5 中结果的合理性:

首先是人的主观评判。本文从对 H 省拥军优属情况相对熟悉的身边人员(至少体验过或了解 H 省部分市州优抚待遇的才作为备选)中随机找出 7 名受访者,对表 5 中的两个评估结果“总体是否合理?”和“前 6 名是否合理?”进行无记名投票,候选答案设置“合理”“基本合理”“不合理”“不知道”四个选项,并要求四选一。投票结果为,3 人认为评估结果总体上“合理”、4 人认为“基本合理”;5 人认为前 6 名“合理”、2 人认为“基本合理”。由此反映,模型评估结果能够较好

符合人的主观判断。

表 5 H 省各市州截至 2020 年 12 月的总评分

Tab.5 Evaluation results of the cities in H province up to November, 2020

市/州名	4 年总评分	4 年总评排序	2 年总评分	2 年总评排序
C1	82.38	1	80.93	1
C2	49.71	2	40.96	2
Z1	43.69	3	39.63	3
S	40.62	4	37.21	4
X1	30.61	5	27.65	5
H1	28.11	6	23.57	6
Y1	20.64	7	13.12	8
Y2	19.69	8	13.85	7
C3	16.72	9	4.89	14
Z2	14.28	10	7.14	11
H2	5.30	11	8.81	9
L	4.28	12	7.71	10
Y3	2.97	13	6.28	12
X2	2.77	14	6.13	13

其次是与相关评比的对照。2020 年 10 月公布了 H 省全国双拥模范城评比结果。该结果系 2020 年评估的过去 4 年 H 省各参评城市的双拥工作情况。共 8 个城市入选,排序为 C1、C2、H2、X1、S、Z1、H1、L,它们可被视为过去 4 年 H 省双拥工作的前 8 名。与之对应的是表 5 第 2、第 3 列给出的模型评估结果,其中 8 个全国双拥模范城的模型评估排序分别是 1、2、11、5、4、3、6、12(见表 5 第 3 列)。对比可见,二者之间具有较好的一致性:模型评出的前 6 名均入选,且前两名顺序完全一致。但二者也有明显差异,特别是,模型排名 11 和 12 的 H2、L 两市,却进入了全国双拥模范城行列。经调研分析认为,导致这种差异的主要原因可能是:模型所评的是“拥军优属”,只是“双拥”的一个方面,以 H2 为例,该地区驻军部队数量众多,不排除其在“拥政爱民”方面有明显比较优势,而使其双拥排名跻身前列。

由表 5 中结果还可看到,考评周期不同时,相同城市的总评排序可能变得不同。比如表 5 中标红的 C3 市,4 年总评排序第 9,而近 2 年总评排序为 14,表明相对于本省其他市州,该市近两年的拥军优属工作进步速度偏慢。

### 2.2.2 不同时间点上的评估结果

固定考评周期  $H$ , 通过改变考评时间点  $t$ , 可以跟踪评估各参评城市拥军优属工作的发展变化情况。图 3 给出了 14 个市州 2016—2020 年的跟踪评估结果, 考评周期均为 1 年, 考评时间点分别选在每年 12 月。图中结果直观反映了过去 5 年各市州推进拥军优属工作的成效。以 C1 和 S 市为例, C1 的拥军优属工作可谓“节节高升”, S 却先升后降、有些后劲不足。但要注意, 这里给出的是两个城市的总评分, 是相对分而非绝对分(见式(6)和式(7)), 故图 3 反映的并不一定是 S 市的绝对分走低, 而是表明新时期各地都在大力推进拥军优属, 该市的进步没有别的市州快。

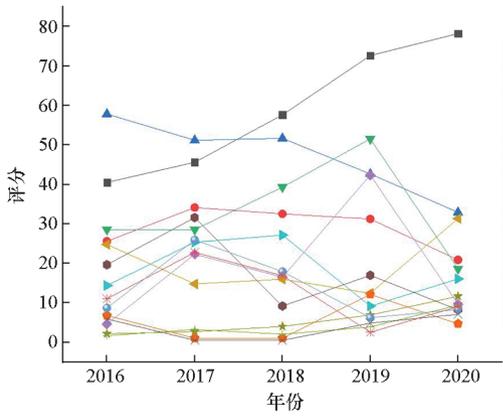
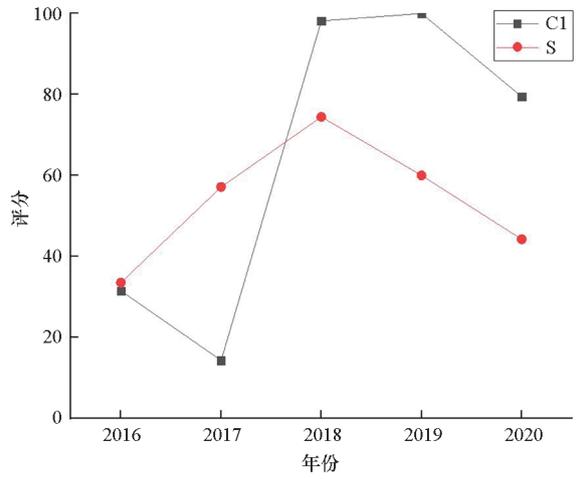


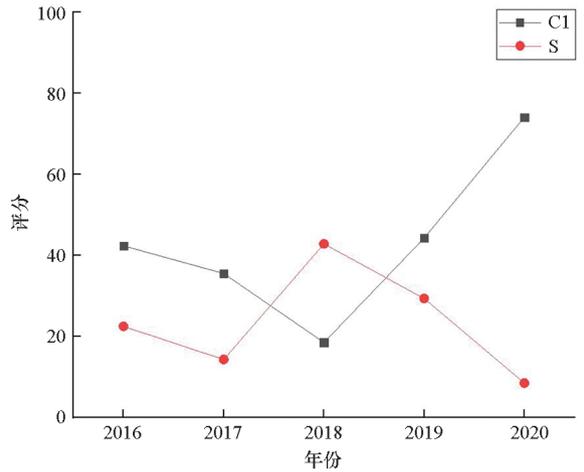
图 3 对 H 省 14 个市州近 5 年的跟踪评估结果  
Fig. 3 Dynamics of the evaluation results of the cities in H province in the recent 5 years

为进一步探究其中原因, 图 4 给出了两个城市 2016—2020 年间 6 个时效性较强的二级指标的标准分(即政策制度类的  $S_1 \sim S_5$  和新闻动态类的  $S_6$ ) 变化和对比情况。可见, C1 市几乎全方位领先; 而 S 市在住房、就业和新闻动态等方面的标准分(相对分)呈下降趋势, 导致整体排位下滑。此结果还反映出, 如果 S 市想要改进工作、提升排名, 下步应重点在住房、医疗、就业等领域的政策制度制定与落实, 以及新闻宣传等方面加大力气、多花功夫。

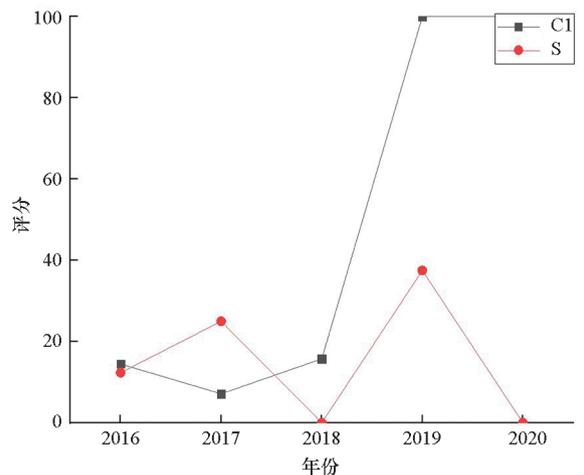
由此可见, 本文模型产生的评估结果可为参评城市改进工作、争先创优指明方向, 提供决策参考。除图 4 给出的结果外, 还可任意设置模型参数  $t$  和  $H$ , 实现随时评、时时评, 密切跟踪参评城市的工作推进发展情况。



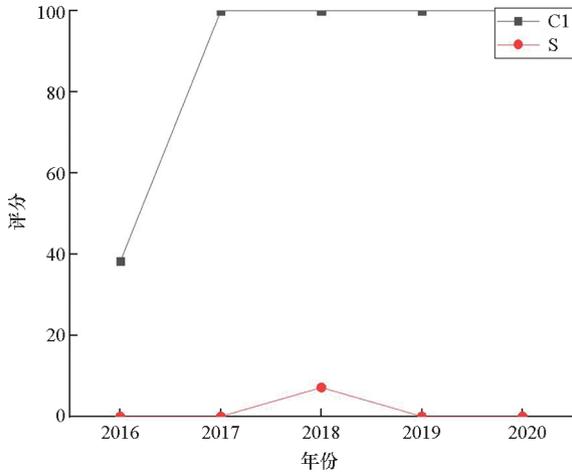
(a) 教育  
(a) Education



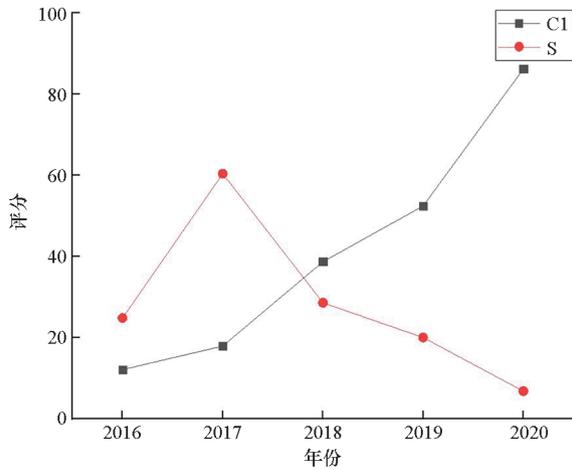
(b) 就业  
(b) Employment



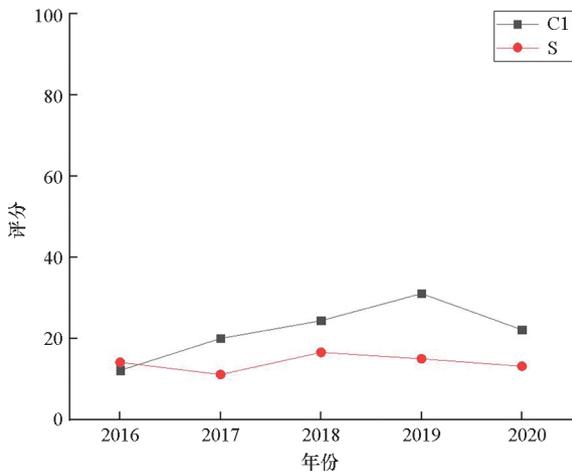
(c) 住房  
(c) Housing



(d) 医疗  
(d) Healthcare



(e) 新闻动态  
(e) Propaganda



(f) 其他  
(f) Others

图4 两个参评城市的6个时效性较强的二级指标得分随时间变化情况

Fig.4 Scores of two cities under consideration with respect to 6 heavily time-dependent second-level indices

### 3 结论

长期以来,针对拥军优属等社会工作的定量评估评价研究还显薄弱,存在无针对性指标体系,现有相关评比周期长、成本高等突出问题。为此,本文探索构建了利用互联网开源大数据实时评估各地社会工作推进情况的定量模型,建立了针对性指标体系,提出了评分快速计算方法,并研制原型软件,以拥军优属为案例,利用网上真实数据进行了模型实证。其中,针对某省各市州的评估排序结果得到多名受访者较一致的认可,且与2020年度该省“全国双拥模范城(县)”命名结果也基本吻合,验证了方法的有效性与合理性。原型已获软件著作权(登记号 2021SR0480424),所提模型也已申报发明专利(公开号 202011483170.1)。本文工作属于利用大数据支撑社会管理的范畴,综合运用管理科学和信息科学的理论技术,特别是层次分析法、文本分析处理、机器学习等方法技术,进行了社会工作评估方面的有益探索。

本文工作也还有很多需要改进和拓展之处,至少包括如下三方面:第一,受众反响是工作评价最重要的依据之一,而当前模型在这方面还有欠缺。下步,拟引入各类社交平台上的网民评论数据,借助情感分析技术进行观点分析,进而评估受众的态度反响,并纳入评估指标体系中。第二,体量大、噪声高、价值密度低是大数据的基本特点,因此本文借助算法自动进行数据过滤、去重和文本分析等,但这些算法都有其精度水平限制(本文实验中文本去重精度约 0.9、相关性过滤精度约 0.92、聚类精度约 0.8、分类精度约 0.87)。尽管实验反映,针对 H 省各市州的最终评估结果比较合理,貌似未明显受到数据分析处理误差的影响,但随着参评范围扩大、数据量显著增加,不排除一些算法的精度水平会明显下降,届时将给评估结果带来怎样的影响值得未来深入探讨。第三,目前原型只关注 H 省,如果未来扩展至全国所有省、自治区、直辖市,包括所辖市州和县区,那么数据量会增大数十倍、甚至数百倍。届时,本文标题中的“大数据”一词将变得更为贴切,但数据体量陡增带来的计算存储压力也必须考虑和应对。鉴于各参评城市的数据及其分析计算,至少在原始分计算结束前相对独立,而后则需要汇总统计,很符合 Map-Reduce 模型,故拟引入该模型进行评估计算的并行化,并将原型移植到 Spark 等支持流式大数据的支撑平台上。

## 参考文献 (References)

- [1] NUGENT T, LEIDNER J L. Risk mining: company-risk identification from unstructured sources [C]//Proceedings of IEEE 16th International Conference on Data Mining Workshops, 2016: 1308 - 1311.
- [2] DHURANDHAR A, GRAVES B, RAVI R, et al. Big data system for analyzing risky procurement entities [C]//Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015: 1741 - 1750.
- [3] ZHAO L T, GUO S Q, WANG Y. Oil market risk factor identification based on text mining technology [J]. Energy Procedia, 2019, 158: 3589 - 3595.
- [4] 陈剑, 肖勇波, 朱斌. 大数据视角下的采购风险评估: 基于某服务采购企业的案例分析 [J]. 系统工程理论与实践, 2021, 41(3): 596 - 612.  
CHEN J, XIAO Y B, ZHU B. Procurement risk evaluation from a big-data perspective: a case study of a procurement service company [J]. Systems Engineering-Theory & Practice, 2021, 41(3): 596 - 612. (in Chinese)
- [5] NYMAN R, KAPADIA S, TUCKETT D. News and narratives in financial systems: exploiting big data for systemic risk assessment [J]. Journal of Economic Dynamics and Control, 2021, 127: 104119.
- [6] JIANG J L, LIAO L, LU X, et al. Deciphering big data in consumer credit evaluation [J]. Journal of Empirical Finance, 2021, 62: 28 - 45.
- [7] WEN C H, YANG J H, GAN L, et al. Big data driven Internet of things for credit evaluation and early warning in finance [J]. Future Generation Computer Systems, 2021, 124: 295 - 307.
- [8] 陈春照, 谢睿, 查婧怡, 等. 基于大数据的中小微企业信贷风险评估和信贷策略研究 [J]. 哈尔滨师范大学自然科学学报, 2021, 37(4): 20 - 30.  
CHEN C Z, XIE R, ZHA J Y, et al. Research on credit risk assessment and credit strategy of small, medium and micro enterprises based on big data [J]. Natural Science Journal of Harbin Normal University, 2021, 37(4): 20 - 30. (in Chinese)
- [9] 商晨迪, 张业安. 大数据时代体育赛事媒体传播效果评估: 实践需求与理论模式 [J]. 体育科研, 2021, 42(3): 19 - 25, 36.  
SHANG C D, ZHANG Y A. Evaluation of sports events media communication effects in the era of big data: practical needs and theoretical models [J]. Sport Science Research, 2021, 42(3): 19 - 25, 36. (in Chinese)
- [10] 刘方坤. 关于建立媒体宣传效果评估体系的思考 [J]. 现代视听, 2020(6): 71 - 74.  
LIU F K. Thoughts on how to build an evaluation system for media propagation effect [J]. Modern Audio-Video Arts, 2020(6): 71 - 74. (in Chinese)
- [11] 姜万川. 媒体宣传效果评估方式简析: 以重庆广播电视集团(总台)综合评价工作为例 [J]. 新闻传播, 2019(18): 35 - 36.  
JIANG W C. A brief analysis of evaluating media propagation effects: taking as an example the Chongqing broadcast & TV station [J]. Journalism Communication, 2019(18): 35 - 36. (in Chinese)
- [12] MCNEELY C L, HAHM J O. The big (data) Bang: policy, prospects, and challenges [J]. Review of Policy Research, 2014, 31(4): 304 - 310.
- [13] 魏航, 王建冬, 童楠楠. 基于大数据的公共政策评估研究: 回顾与建议 [J]. 电子政务, 2016(1): 11 - 17.
- WEI H, WANG J D, TONG N N. Public policy evaluation using big data: review and suggestion [J]. E-Government, 2016(1): 11 - 17. (in Chinese)
- [14] 陈家刚. 大数据时代的公共政策评估研究: 挑战、反思与应对策略 [J]. 河南社会科学, 2019, 27(8): 46 - 51.  
CHEN J G. On public policy evaluation in big data era: challenge, reflection and countermeasure [J]. Henan Social Sciences, 2019, 27(8): 46 - 51. (in Chinese)
- [15] 窦悦, 童楠楠, 易成岐, 等. 大数据视角下我国社会政策评估体系架构与创新路径研究 [J]. 情报理论与实践, 2021, 44(7): 50 - 57.  
DOU Y, TONG N N, YI C Q, et al. Research on the architecture and innovation path of China's social policy evaluation from the perspective of big data [J]. Information Studies: Theory & Application, 2021, 44(7): 50 - 57. (in Chinese)
- [16] HENDRIKX R J P, DREWES H W T, SPREEUWENBERG M, et al. Measuring regional quality of health care using unsolicited online data: text analysis study [J]. JMIR Medical Informatics, 2019, 7(4): e13053.
- [17] 徐蔼婷. 德尔菲法的应用及其难点 [J]. 中国统计, 2006(9): 57 - 59.  
XU A T. The application and the difficulties of Delphi method [J]. China Statistics, 2006(9): 57 - 59. (in Chinese)
- [18] 杜子芳. 多元统计分析 [M]. 北京: 清华大学出版社, 2016.  
DU Z F. Multivariate statistical analysis [M]. Beijing: Tsinghua University Press, 2016. (in Chinese)
- [19] 李莉娟, 张娜. 关于主成分分析的几种常用改进方法 [J]. 统计与决策, 2007(21): 167 - 168.  
LI L J, ZHANG N. Several commonly used improved methods of principal component analysis [J]. Statistics and Decision, 2007(21): 167 - 168. (in Chinese)
- [20] 郭金玉, 张忠彬, 孙庆云. 层次分析法的研究与应用 [J]. 中国安全科学学报, 2008, 18(5): 148 - 153.  
GUO J Y, ZHANG Z B, SUN Q Y. Study and applications of analytic hierarchy process [J]. China Safety Science Journal, 2008, 18(5): 148 - 153. (in Chinese)
- [21] 周志华. 机器学习 [M]. 北京: 清华大学出版社, 2016.  
ZHOU Z H. Machine learning [M]. Beijing: Tsinghua University Press, 2016. (in Chinese)
- [22] GEPPERATH A, HAMMER B. Incremental learning algorithms and applications [C]//Proceedings of European Symposium on Artificial Neural Networks, 2016.
- [23] CASTRO F M, MARÍN-JIMÉNEZ M J, Guil N, et al. End-to-end incremental learning [C]//Proceedings of European Conference on Computer Vision, 2018: 241 - 257.
- [24] 徐萌. 面向对象的 Twitter 事件抽取技术研究 [D]. 长沙: 国防科技大学, 2017.  
XU M. Object-oriented event extraction from Twitter [D]. Changsha: National University of Defense Technology, 2017. (in Chinese)
- [25] KIM Y. Convolutional neural networks for sentence classification [C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 2014: 1746 - 1751.
- [26] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality [C]//Proceedings of the 26th International Conference on Neural Information Processing Systems, 2013: 3111 - 3119.
- [27] ZHANG Z, CHEN L, YIN F J, et al. Improving online clustering of Chinese technology web news with bag-of-near-synonyms [J]. IEEE Access, 2020, 8: 94245 - 94257.