

基于弹幕情感分析的视频片段推荐模型

邓 扬, 张晨曦, 李江峰*

(同济大学 软件学院, 上海 201804)

(* 通信作者电子邮箱 lijf@tongji.edu.cn)

摘 要: 针对传统的视频情感分析方法计算效率较低且结果不易解释等问题, 提出一种基于弹幕文本的视频片段情感识别算法, 并以此作为视频片段的推荐依据。首先对基于情感分析的视频片段推荐问题提出形式化描述。其次, 通过构建基于隐含狄利克雷分布(LDA)的弹幕词语分类, 评估弹幕词语在视频片段中的多维情感向量, 同时, 根据视频片段之间的情感依赖关系推荐视频的情感片段。所提方法的推荐准确度比基于词频-逆文档频率(TF-IDF)的推荐算法提高了 28.9%, 相对于传统 LDA 模型提高了 43.8%。实验结果表明所提模型可有效应用于信息繁杂的不规则文本情感分析。

关键词: 视频片段推荐; 弹幕情感; 主题模型; 情感分析; 情感向量

中图分类号: TP391.1 **文献标志码:** A

Video shot recommendation model based on emotion analysis using time-sync comments

DENG Yang, ZHANG Chenxi, LI Jiangfeng*

(School of Software Engineering, Tongji University, Shanghai 201804, China)

Abstract: To solve the problem that traditional video emotional analysis methods can not work effectively and the results are not easy to explain, a video shot emotional analysis approach based on time-sync comments was proposed, as a basis for the recommendation of video shots. First, a formal description of video shots recommendation based on emotion analysis was studied. Then, after analyzing the classification of time sync comments based on Latent Dirichlet Allocation (LDA) topic model, the emotional vector of the words in time-sync comments were evaluated. Meanwhile, the emotion relationships among the video shots were analyzed for video shots recommendation. The recommendation precision of the proposed method was 28.9% higher than that of the method based on Term Frequency-Inverse Document Frequency (TF-IDF), and 43.8% higher than that of traditional LDA model. The experimental results show that the proposed model is effective in analyzing the complex emotion of different kinds of text information.

Key words: video shot recommendation; time-sync comments emotion; topic modeling; emotion analysis; emotional vector

0 引言

随着互联网的迅速普及, 多媒体处理技术、网络传输、视频数据处理等相关技术快速发展, 来自不同领域的视频数据正在以惊人的速度增长。例如 YouTube, 其每个月的视频播放总次数达到了 46.7 亿次, 另外一个专业视频网站 Hulu 也有超过 20 万个高质量的视频, 其月浏览次数也维持在 9 亿。相对于观看一部完整的视频, 有些时候用户更愿意观看其中一些视频片段。例如一段幽默电影片段的集锦, 或者振奋人心的足球进球集锦; 用户甚至希望可以根据之前看过的视频, 搜索与之情感相似的其他视频片段。视频里所蕴含的情感通常是多种情感交织的混合情感, 所以基于情感的视频片段推荐是一个具有挑战性的研究问题。

近几年随着新媒体技术的发展, 一种新类型的视频——弹幕视频在世界范围内变得愈来愈流行, 即用户可以在观看网络视频中发表自己的实时评论, 并与其他用户产生共鸣或

者互动。在用户所发的弹幕中, 往往包含着与视频当前片段的情感相吻合的文字, 这对展开进一步的研究起到很好的帮助作用。作为一种新型的消息传递方式, 弹幕书写与传统短文本存在很大区别, 主要表现在两个方面: 1) 文本更简短随意。多数评论者选择用短语、短句和符号来表示内心活动。2) 内容口语化、符号化。用户往往通过当下流行的热门语来表达自己的情感。

目前, 有两种基于视频弹幕数据的视频分析技术可以对该问题起到一定借鉴作用。其中, 基于弹幕数据的视频片段自动标签技术^[1-2]提供了一种提取视频片段中弹幕的关键字给这段片段打标签的技术; 基于弹幕的视频精彩镜头的检测方法^[3]主要根据弹幕内容以及数量的变化曲线分析视频中的精彩瞬间。这两种基于内容的分析方法往往难以刻画视频片段的情感趋势, 原因有如下两点: 其一, 情感鲜明的视频片段不能等同于精彩片段, 精彩片段往往出现在弹幕评论密集的区域, 而这种判断模式并不适用于视频情感检测; 其次,

收稿日期: 2016-10-25; 修回日期: 2016-12-21。

作者简介: 邓扬(1991—), 男, 四川成都人, 硕士研究生, 主要研究方向: 信息检索、机器学习; 张晨曦(1960—), 男, 福建龙岩人, 教授, 博士生导师, 博士, 主要研究方向: 分布式计算、嵌入式系统; 李江峰(1983—), 男, 湖北荆州人, 讲师, 博士, CCF 会员, 主要研究方向: 分布式计算、社会网络计算。

通过弹幕数据提取的文字标签通常来源于弹幕中的关键字和高频词,而这些词的所蕴含的情感可能是不易确定的,所以通过标签信息来指代视频片段的情感并不准确。

据调查,目前国内外基于弹幕数据开展的研究工作并不多^[1-6]。针对上述提出的问题,本文提出了一种基于情感分析的视频片段推荐模型,利用弹幕数据分析各个视频片段中用户的情感倾向和交互情况,计算片段的情感以及片段之间的情感相似度,并以此为片段推荐的依据。为检测视频片段的情感,本文设计了一种基于文本主题模型的视频弹幕情感识别方法。首先依据主题模型对弹幕中的词进行主题分类;其次,基于每个词的主题分布和情感词典提出动态评估词语的情感向量的度量算法;最后,通过计算每个视频片段中的每个词的情感向量得到每个片段的情感向量。此外,针对如何评估两个视频片段的情感相似度,本文采用一种基于“全局+局部”的上下文关联情感相似度计算方法,并通过情感相似度分数来推荐视频片段。

1 相关工作

1.1 针对中文的情感分析研究

中文情感识别技术主要分为两种:监督式学习方法和组合方法。监督式学习方法中,文献[7]提出了基于条件随机场(Conditional Random Field, CRF)的冗余情感特征的概念,并对情感分类。文献[8]基于支持向量机(Support Vector Machine, SVM)对长文本中的主观性文字和客观性文字进行分类。组合方法中,文献[9]中研究了四种不同分类方法在中文情感分类中的应用,同时考虑到不同领域需要选择不同分类才能获得更好的分类效果,采用了一种基于Stacking的组合分类方法。目前针对中文的情感分析相对于英文的情感分析,无论从资源还是方法上来说都要相对初步一些,主要存在以下难点:1)中文需要分词;2)中文情感词典需要构建;3)中文理解存在一些难点,而目前尚无较好的解决方法,如“反讽”“褒义贬用”“贬义褒用”;4)对标注数据的规模过于依赖。

1.2 情感分析在网络评论领域的应用

目前针对网络评论的情感分析应用研究多集中在网络舆论和商品评论两个领域。例如,利用情感分析技术对微博舆论中的热点话题、公众观点等进行提取,或者面向商品的用户在线评论进行褒贬识别。文献[10]提出一种监督式机器学习方法来利用局部文本特征对新闻自动分类。文献[11]研究表明,采用一种结合支持向量机的机器学习算法、信息增益的特征选取算法和词频-逆文档频率(Term Frequency-Inverse Document Frequency, TF-IDF)特征项权重计算的组合方法对微博短文本的情感分类效果最好。作为一种新类型的网络评论方式,弹幕写作简单随意,并无严格的语法要求,这导致传统的情感分析方法不能直接适用于弹幕情感研究。

1.3 隐含狄利克雷模型

隐含狄利克雷分布(Latent Dirichlet Allocation, LDA)模型是由Blei等^[12-13]在2003年提出的一种主题模型,由文档、主题、词三层结构组成。2004年Griffiths等在Blei的基础模型上对 β 参数施加了Dirichlet先验分布,使得LDA模型成为一个完整的概率生成模型。目前LDA模型已经广泛地应用到文本挖掘甚至音频分析领域^[14]。对于语料库中的每一篇文章, LDA模型的主题生成过程为:1)求主题的多项式分布参

数 θ , $\theta \sim \text{Dir}(\alpha)$, α 为 θ 的超参数;2)求单词的多项式分布参数 η , $\eta \sim \text{Dir}(\beta)$, β 为 η 的超参数;3)根据 θ 生成主题 z , $z \sim \text{Multinomial}(\theta)$;4)采样过程,根据参数 z 和 η ,从 $p(w|z, \eta)$ 中选择一个词 w ;5)重复步骤3)~4)直到模型收敛。

1.4 情感语料

目前有一些情感词典被开发出来用于情感分析,比如WordNet-Affect。本文从Ren_CECps中文情感语料库^[15]提取多维情感词典,该语料库已经广泛用于中文情感的研究^[16-17]。该语料库包含了1487篇中文博客,涵盖了878164个情感词,每个情感词在语料中都被标注为(惊讶、悲伤、喜爱、高兴、憎恨、期待、焦虑、生气)8个维度表示的向量,每个维度用0到1表示情感从弱到强。

考虑到弹幕中有大量的网络用语,本文采用文献[18]中所提的针对网络词汇的情感词典(简称Catchwords词典)作为补充。该网络用语情感词典涵盖了733个网络用语,并将其分为7个情感分类(乐、好、怒、愁、惊、恶、惧)。

2 问题描述

2.1 背景

本节介绍本文相关背景,包括弹幕视频的特点和结构,以及相关定义。一个用户观看到视频中某个场景,可能会在这个时间点写一些文字发送到视频中分享他的观看体验,这就是弹幕。当后面的用户再看到该视频中的这个场景,之前用户发送的弹幕就会出现在屏幕上。所有的弹幕均为用户在观看过程中发送的,弹幕反映了用户在观看这段片段的情绪,这种影响具有一定的传播效应,甚至依据当前这一幕的弹幕可以预测视频后面的情感发展。

弹幕可以定义为一个三元组,其包含了用户输入内容、发送时间、用户ID。本文主要对三种形式的弹幕内容进行分析:中文、英文、数字。其中:英文包含了部分网络用语,如“QAQ”(表示哭泣);而考虑数字的原因,有部分常用网络流行词,如“233”(表示大笑)和“666”(表示厉害)等,对情感分析很有帮助。

如图1所示,弹幕视频截图上方的文字为用户发送的弹幕,例如“要出事了”是某用户于该视频的120 min 10 s到120 min 40 s之间发送的一条弹幕。

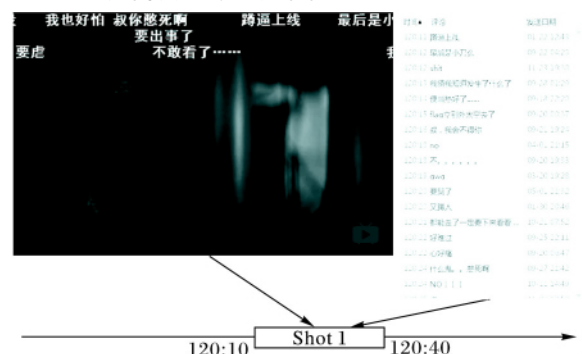


图1 bilibili上的弹幕视频示例

Fig. 1 Video example of time-sync on bilibili

2.2 形式化描述

设视频集合为 $V = \{v_1, v_2, \dots, v_{|V|}\}$, $|V|$ 为集合大小,其中视频 v_i 的时间长度为 T_{v_i} ,切分为 $|v_i|$ 个视频片段 $v_i = \{s_{i,1}, s_{i,2}, \dots, s_{i,|v_i|}\}$, $s_{i,k}$ 代表第 i 个视频中的第 k 个视频片段

($1 \leq k \leq |v_i|$) s_{ik} 的时间长度为定值 T_s 。因此, $|v_i| = \lceil (T_{v_i} - T_s) / t_d + 1 \rceil$ 其中: t_d 代表视频片段之间的滑动距离, 即相邻两段片段的相隔时间距离。视频片段集合为 S , 总共包含 $|S|$ 个视频片段。 W 代表 S 中的词集合, 包含 $|W|$ 个词。弹幕 c 定义为三元组 (w_c, t_c, μ_c) , 其中: w_c 代表弹幕 c 中的词集合, t_c 代表弹幕 c 在视频中的时间戳, μ_c 代表发送弹幕 c 的用户。因此片段 s_k 可以定义为:

$$s_k = \{ \langle d_{w_{k1}} : |w_{k1}| \rangle, \langle d_{w_{k2}} : |w_{k2}| \rangle, \dots, \langle d_{w_{kn_k}} : |w_{kn_k}| \rangle \}$$

其中: n_k 代表 s_k 中不一样的词的个数; w_{ki} 代表片段 s_k 第 i 个词; $|w_{ki}|$ 代表 w_{ki} 在片段 s_k 中出现的次数; $d_{w_{kn_k}}$ 代表 w_{kn_k} 对应的编号。

本文要解决的问题是给定目标片段 s_{target} , 求解视频集合中与 s_{target} 情感相似度高的片段集合:

$$c_{v, s_{\text{target}}} = \max_{c'_{v, s_{\text{target}}} \in v} (c'_{v, s_{\text{target}}} | s_{\text{target}}, p_{\text{target}}, S, W); \forall v \in V$$

其中: $s_{\text{target}} \in v_{\text{target}}$; $c_{v, s_{\text{target}}}$ 代表来自 v 的情感相似度排名最高的 n_{top} 个片段集合; $c'_{v, s_{\text{target}}}$ 代表 $c_{v, s_{\text{target}}}$ 的候选值。最后对 V 中的每个视频 v_i 找到最优 $c_{v_i, s_{\text{target}}}$ 返回推荐集合:

$$C_{s_{\text{target}}} = \{ c_{v_1, s_{\text{target}}}, c_{v_2, s_{\text{target}}}, \dots, c_{v_{|V|}, s_{\text{target}}} \}$$

3 视频片段推荐模型

如图2 视频片段推荐的工作流程包含三个子模块: 1) 数据预处理; 2) 视频片段情感识别; 3) 情感相似度分析。

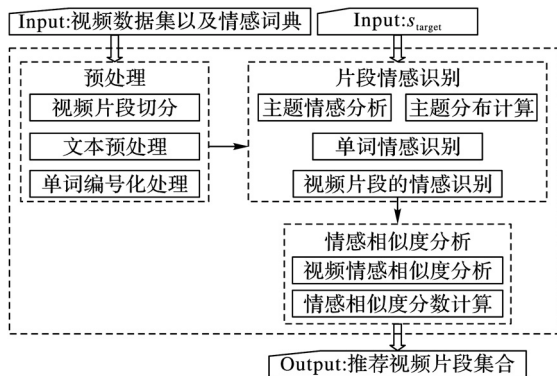


图2 视频片段推荐流程

Fig. 2 Process of video shot recommendation

3.1 弹幕数据预处理

数据预处理阶段主要包含了视频片段切分、文本预处理、词语编号化处理三个步骤。首先根据 T_s 和 t_d 对所有视频的弹幕数据进行切分, 并对片段进行初步的筛选; 其次, 从 Ren_CECps 中文情感语料库提取情感词典。本文通过对每个情感词的向量总和求平均得到情感词的平均情感向量, 公式如下:

$$e_w = \frac{1}{T_w} \sum_{i=1}^{T_w} e_{w_i}$$

其中: T_w 代表情感词 w 在 Ren_CECps 中文情感语料库中出现的总次数; e_{w_i} 代表情感词 w 第 i 次在语料库中出现时被标注的情感向量。后文中用“Ren_CECps 词典”代表本文从该语料库中所提取的情感词典。为缩短模型训练的时间, 本文对所有词编号化处理, 并剔除难以识别的网络符号以及出现概率极低的词。

3.2 视频片段的情感识别

弹幕的词分布可视为一个概率模型, 在同一个视频片段

中的词相互之间具有情感依赖, 因此, 本文提出一种基于多主体的视频片段情感识别 (Multi-Topic Emotion Recognition, MTER) 算法。MTER 利用每个词的隐含情感主题分布来评估词的情感值, 并用情感向量表示。考虑到大量的词的情感没有 Ren_CECps 词典中有标注, 且中文词语在不同语境下可能会有不同的情感含义, 本文采用词的最大权重的主题来计算出该词在当前片段中的临时情感。

根据前面的定义, s_k 可以被视为一系列词的集合, 根据 Ren_CECps 词典, 每个词 w_{ki} 的情感可以表示为一个维度为 8 的情感向量: $e_{w_{ki}} = (e_{w_{ki}}^1, e_{w_{ki}}^2, \dots, e_{w_{ki}}^8)$, 每个情感词在 Ren_CECps 词典由 8 个情感维度表示。在提出本文方法之前, 一个基于统计的情感计算方法, 即对 s_k 中能在 Ren_CECps 词典中检索到的词的情感向量求和, 从而得到 s_k 的情感向量:

$$E_{s_k} = \left(\sum_{w \in s_k \cap R} |w| e_w^1, \sum_{w \in s_k \cap R} |w| e_w^2, \dots, \sum_{w \in s_k \cap R} |w| e_w^8 \right)$$

其中: R 表示 Ren_CECps 词典; $\sum_{w \in s_k \cap R} |w| e_w^j$ 是第 j 个维度的情感权重 ($1 \leq j \leq 8$)。当大多数词能够在情感词典里检索到时, 这种统计方法的效果很好。然而基于本文的数据统计, 弹幕中有只有 46% 的词可以在 Ren_CECps 词典中检索到, 以及 12% 的词可以在 Catchwords 词典中检索到, 所以仅基于词典的统计方法显然会降低情感分析的准确度。针对上述问题, 本文引入 LDA 主题模型作为统计方法的补充。

3.2.1 基于主题模型计算视频片段情感向量

本文提出的模型中, 视频片段视作一组弹幕词语构成的集合, 不考虑词与词的先后关系。将 LDA 模型的思想引入视频片段的情感识别模型, 假设模型训练初始设置 K 个隐含主题 $\{t_1, t_2, \dots, t_K\}$, 所以每个词在视频片段中都能对应一个主题, 每个视频片段能计算得出一个主题分布。在所提模型中, LDA 的训练过程如下: 1) 为片段中每个弹幕词语初始化一个主题; 2) 根据其他弹幕词语的主题为每一个词重新生成一个新主题; 3) 基于 Gibbs Sampling 对每个词的主题重新采样, 并指定迭代次数 η 。LDA 过程的结果如表 3 所示。

表1 主题分布示例

Tab. 1 Example of topic allocation

主题	词的主题分布
主题 1	0.004* 大笑, 0.004* 哈哈, \dots 0.001* 高能
主题 2	0.005* 什么, 0.005* cry, \dots 0.001* 不要
...	...
主题 n	0.005* 帅气, 0.003* 很赞, \dots 0.001* 漂亮

如表 1 所示, 字体加粗的词语代表可以在 Ren_CECps 词典中检索到, 且各主题中每个词按照权重降序排列。其中 w_{ti} 代表 Topic t 的第 i 个词, α_{ti} 代表 w_{ti} 在 t 中的权重。随着 i 的增加 α_{ti} 逐渐趋于 0, 这表示在一个主题中, 权重越小的词与该主题的相关性越低。因此, 为了保证方法的有效性, 本文只考虑每个主题的 top P 个词。每个主题的情感向量的计算公式如下:

$$E_t = \left[\frac{1}{M} \sum_{i=1}^P (1 + \alpha_{ti}) e_{w_{ti}}^1, \frac{1}{M} \sum_{i=1}^P (1 + \alpha_{ti}) e_{w_{ti}}^2, \dots, \frac{1}{M} \sum_{i=1}^P (1 + \alpha_{ti}) e_{w_{ti}}^8 \right]$$

其中: $\sum_{i=1}^P (1 + \alpha_{ti}) e_{w_{ti}}^j$ 代表第 j 个情感因子的情感权重值; M 代

表 8 个情感因子中情感强度最大的值,其目的是对每个情感维度的强度作归一化处理。这里如果 w_{ti} 不能在 Ren_CECps 词典中检索到,则 $e_{w_{ti}} = (0^1, 0^2, \dots, 0^8)$ 。接下来根据 s_k 的主题分布中,权重最大的主题 t_{km} ,可以对 s_k 中不能被词典检索到的词的情感进行评估,所以 s_k 中的词 w_{ki} 的情感计算公式如下:

$$e_{ki} = \begin{cases} e_{w_{ki}}, & w_{ki} \in R \\ \omega_{km} \times (1 + \alpha_{t_{km}^i}) \times E_{t_{km}}, & \text{其他} \end{cases}$$

其中: ω_{km} 代表 t_{km} 的权重; $E_{t_{km}}$ 代表 t_{km} 的情感向量。因此 s_k 的情感向量计算公式如下:

$$E_{s_k} = \left(\sum_{i=1}^n |w_{ki}| e_{ki}^1, \sum_{i=1}^n |w_{ki}| e_{ki}^2, \dots, \sum_{i=1}^n |w_{ki}| e_{ki}^8 \right)$$

3.2.2 基于上下文信息调整视频片段情感向量

视频片段的情感具有传递性,归属于同一个视频中的视频片段具有一定的情感相关性和相似性,尤其是时间相隔较近的片段。本文对视频片段情感向量的调整公式如下:

$$E'_{s_{i,k}} = \mu E'_{s_{i,k-1}} + (1 - \mu) E_{s_{i,k}}$$

其中: $\mu (0 < \mu < 1)$ 是影响因子, μ 越大代表 $s_{i,k}$ 之前的片段对 $s_{i,k}$ 的情感影响也大。可得出所有在 $s_{i,k}$ 之前的片段都会影响 $s_{i,k}$ 的情感向量,且 $s_{i-1,k}$ 的影响最大。

综上,基于 MTER 的视频片段情感识别算法描述如下。

算法 1 视频片段情感识别算法。

输入: 视频片段集合 S ; LDA 过程的隐含主题数量 K ; LDA 过程的主题采样次数 η 。

输出: $E_{s_k}, \forall s_k \in S$ 。

数据预处理: 对弹幕文本集合进行 LDA 主题采样。初始化每个视频片段 s_k 的情感向量 $E_{s_k} = (0^1, 0^2, \dots, 0^8)$ 。

for each topic t do:

 Calculate E_t

end for

for each shot s_k do

 for each word w_{ki} in s_k do

 if $w_{ki} \in R$ then

 Update E_{s_k} with $e_{w_{ki}}$ //把 word w 的情感向量累加到 E_s

 else

 Calculate e_{ki} for w_{ki}

 Update E_{s_k} with e_{ki}

 end if

 end for

end for

for each video v_i do

 for each shot $s_{i,k}$ do

 Adjust $E_{s_{i,k}}$

 end for

end for

3.3 情感相似度分析

计算得出每个视频片段的情感向量之后,本文提出一种片段间的情感相似度的计算方法。鉴于片段所属的视频的全局情感会很大程度上影响到片段本身的情感,例如,来自两部喜剧的片段的相似度有很大概率高于分别来自喜剧和悲剧的两个片段的情感相似度。因此,本文结合全局情感(video-level)和局部情感(shot-level)来计算两个片段之间的情感相似度。计算方法如下:

一部完整的视频包含了一系列视频片段,而视频对应的弹幕数据中包含的网络情感词汇数量远大于一个视频片段。

因此,计算一部完整视频的情感可直接对视频弹幕中的网络情感词的数量进行统计。本文定义视频的情感为 $E_v = (n_1, n_2, \dots, n_7)$ (Catchwords 词典共有 7 个情感分类),其中 $n_i (1 \leq i \leq 7)$ 代表视频 v 在 Catchwords 词典中的第 i 个分类的情感词的数量,本文根据杰卡德距离(Jaccard Distance)来计算两个视频之间的整体情感相似度:

$$E_v\text{-sim}(E_{v_j}, E_{v_k}) = \frac{\sum_{i=1}^7 \min(n_{ji}, n_{ki})}{\sum_{i=1}^7 \max(n_{ji}, n_{ki})}$$

根据本文对视频片段的情感定义 $E_i = (e^1, e^2, \dots, e^8)$,可根据余弦相似度(cosine similarity)对两个视频片段的情感向量求相似度:

$$E_s\text{-sim}(E_{s_j}, E_{s_k}) = \frac{\sum_{i=1}^8 e_j^i \times e_k^i}{\sqrt{\sum_{i=1}^8 (e_j^i)^2} \times \sqrt{\sum_{i=1}^8 (e_k^i)^2}}$$

因此,任意两个视频片段的情感相似度的计算公式定义为:

$$E_s\text{-sim}(E_{s_{i,k}}, E_{s_{j,l}}) = E_v\text{-sim}(E_{v_i}, E_{v_j}) \times E_s\text{-sim}(E_{s_k}, E_{s_l})$$

最后根据视频片段 $s_{i,k}$ 和 s_{target} 之间的情感相似度对 v_i 中的视频片段排名,取前 n_{top} 的片段作为 $c_{v_i \neq target}$,最终得到 $C_{s_{target}}$ 。

综上,视频片段推荐算法描述如下。

算法 2 视频片段推荐算法。

输入: 视频片段集合 S ; 目标片段 s_{target} 。

输出: 视频片段集合 $C_{s_{target}}$ 。

for each shot s do

 Calculate $E_s\text{-sim}(E_{s_{target}}, E_s)$

end for

for each video v do

 for each shot s do

 if s is top n_{top} of $E_s\text{-sim}$ in v

 Add s to $c_{v_i \neq target}$

 end if

 end for

 Add $c_{v_i \neq target}$ to $C_{s_{target}}$

end for

return $C_{s_{target}}$

4 实验比较与分析

4.1 数据说明

本文的数据由两部分组成: 1) 国内视频网站“优酷”的一段混剪视频,这段混剪视频由十部不同电影中的一些激励人心的片段组成; 2) 视频片段集,从国内弹幕视频网站“bilibili”上下载上述混剪视频相关的十部电影的弹幕数据。这些弹幕数据根据本文的视频切分规则切分为一系列弹幕数据块,每一块数据对应一个视频片段。收集所有覆盖混剪视频相关内容的励志视频片段作为目标片段集合,最后收集得到 17 个目标片段。

本文选用的中文分词工具是 Jieba,通过对弹幕数据进行分词、清洗、去噪之后,最后剩余 352 892 个词。通过设置 $T_s = 120$ s, $t_d = 100$ s 来把视频切割成一系列片段,并筛选弹幕数量至少含有 8 条的视频片段,且每一部电影至少包含 50

个视频片段。此外,为了避免视频中的片头片尾曲对情感分析造成干扰,本文剔除每部电影的前两个片段和最后两个片段。数据集概况如表2所示。

表2 实验数据概况
Tab. 2 Overview of dataset

符号	数据说明	值
$n_{\text{objective}}$	目标片段集合的大小	17
$ V $	电影总数	10
$ S $	视频片段总数	669
$ W $	词数量	352 892
$ w_c $	平均每条弹幕包含的词数量	6.2
$ v_c $	平均每部电影包含的弹幕数量	5 317

4.2 实验设置

本文设计了几组对比实验来验证模型的准确性。TF-IDF是文档关键词抽取的典型算法。本文中基于TF-IDF的对比实验被分为“TF-IDF-R”和“TF-IDF”。对于“TF-IDF-R”,本文根据每个片段中的词频向量和Ren_CECps词典来计算视频片段的情感向量;而在“TF-IDF”中仅通过片段中的TF-IDF词频向量来描述视频片段的情感特征。本文提出的情感相似度计算方法均适用于这两种方法。“TF-IDF-R”和“TF-IDF”都在实验效果中表现稳定。

本文还设置了原始LDA模型作为MTER的对比实验,与TF-IDF不同,LDA和MTER都是基于迭代计算的方法,原始LDA模型中的参数和MTER中的LDA过程的实验参数保持一致,主要参数设置如下:主题采样迭代次数 $\eta=400$;隐含主题个数 $K=20$;LDA中的主题分布的先验值 $\alpha=0.1$;对于MTER,情感影响因子 $\mu=0.15$;主题中的有效单词个数 $P=100$ 。

实验过程包括以下几个步骤:1)对电影切分后的所有视频片段(包含上文所述的 $n_{\text{objective}}$ 个励志片段)分别计算情感向量(见算法1);2)依次取目标片段集合中的视频片段作为 s_{target} ,计算 S 中其他片段与 s_{target} 的情感相似度分数,并在每个电影 v_i 的视频片段中取分数最高的 n_{top} 个视频片段加入推荐列表(见算法2);3)检测每个 s_{target} 的推荐准确度。

本文分别求出每个目标片段的 F_1 值,并求出平均值 $\text{mean}F_1$,公式定义如下:

$$\text{Precision} = n_{\text{correct}} / |C_{s_{\text{target}}}|$$

$$\text{Recall} = n_{\text{correct}} / n_{\text{objective}}$$

$$\text{mean}F_1 = \frac{1}{n_{\text{objective}}} \sum_{i=1}^{n_{\text{objective}}} F_1^{(i)}$$

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

其中: n_{correct} 代表上述混剪视频中的励志片段被推荐给 s_{target} 的数量, $F_1^{(i)}$ 代表第 i 个 s_{target} 的 F_1 值。

4.3 参数学习

主要对MTER中的四个主要参数进行学习:1)LDA过程的隐含主题个数 K ;2)主题采样迭代次数 η ;3)情感影响因子 μ ;4)主题中的有效单词个数 P 。参数学习结果如图3所示。

当主题数量为20左右, $\text{mean}F_1$ 达到最优值。当 $K > 20$ 时, $\text{mean}F_1$ 开始迅速下降。由第3章所述, K 越大导致 ω_{km} 越小,主题模型对情感向量的影响降低。而当 $K < 10$ 时,对于风格不同的视频片段 t_m 为同一主题的概率会增大,这样同样不

能发挥主题模型在情感识别中的优势。

P 取100左右, $\text{mean}F_1$ 达到最优值。其原因在于随着词在 t_{km} 排名越靠后, $\alpha_{t_{km}}$ 趋于0,词的影响力也越小。此时采用更多数量的词参与计算,会使得主题情感的准确性降低。

主题采样迭代次数 η 为400时,模型效果达到最佳效果。当 η 超过400,LDA过程将会出现过拟合的现象。

当 $\mu=0.15$ 时,模型达到最优;当 $\mu > 0.15$ 时, $\text{mean}F_1$ 下降的原因 μ 增大导致了片段的情感区分度降低,导致模型误差。

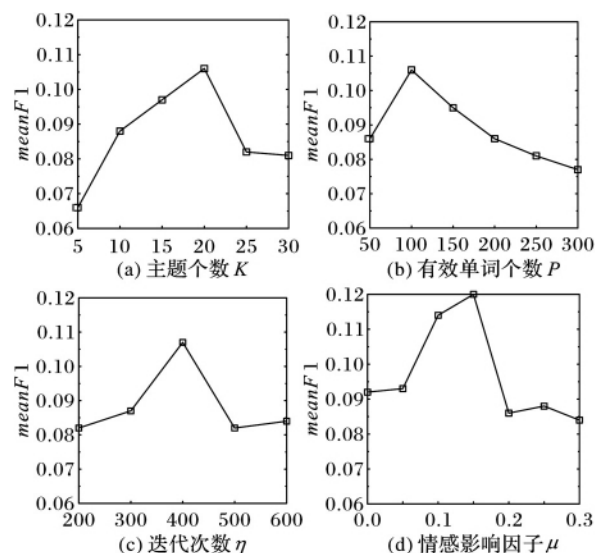


图3 MTER 参数学习结果

Fig. 3 Parameter learning results of MTER

4.4 实验结果

实验反映了不同 n_{top} 如何影响推荐结果。当 $n_{\text{top}}=15$,推荐总数远大于目标片段的数量,导致四种方法的准确率都达到最低值;当 n_{top} 从15降低到10,MTER的召回率降低了19%,而TF-IDF-R、LDA、TF-IDF分别降低27%、24%、32%;当 n_{top} 降低到5,每个电影中只有5个分数最高的视频片段加入推荐列表,此时的推荐结果正确性非常依赖目标片段的弹幕数据,导致四种方法的差距非常小,即便如此MTER依然优于其他方法。基于MTER的推荐方法和其他三种对比实验的结果如图4~5所示。

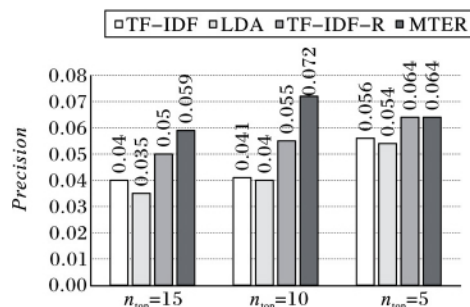


图4 4种算法准确率结果对比

Fig. 4 Precision comparison of four algorithms

将 $\text{mean}F_1$ 的实验结果列入表3。可以看出不借助情感词典的方法TF-IDF和LDA的实验结果并不理想,其最好结果是TF-IDF在 $n_{\text{top}}=5$ 取得,对应的 $\text{mean}F_1=0.084$ 。借助情感词典的方法的 $\text{mean}F_1$ 最低为0.089。实验中最优的

$meanF_1$ 是 MTER 在 $n_{top} = 10$ 取得。当 n_{top} 取得 10 和 15 时, MTER 比 TF-IDF-R 分别提高了 30% 和 19%。然而在不借助词典的情况下, TF-IDF 在 n_{top} 取三个不同值的情况下均优于 LDA。其主要原因为 LDA 根据片段的主题分布来计算片段之间的相似度, 然而部分词尤其是网络用语适用于多种类型的场景, 这种在各个片段中出现概率都很大的词可能会在不同的主题中都有很大的权重, 这样就降低了主题模型在情感识别中的效果。

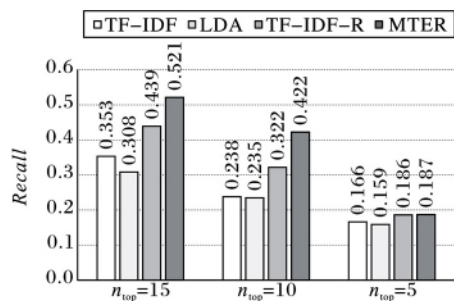


图 5 4 种算法的召回率结果对比

Fig. 5 Recall comparison of four algorithms

表 3 4 种算法 $meanF_1$ 的结果对比

Tab. 3 $meanF_1$ comparison of four algorithms

方法	$meanF_1$		
	$n_{top} = 15$	$n_{top} = 10$	$n_{top} = 5$
TF-IDF	0.072	0.070	0.084
LDA	0.063	0.068	0.081
TF-IDF-R	0.089	0.094	0.095
MTER	0.106	0.123	0.096

通过以上实验分析可以发现, 基于 MTER 的视频片段推荐方法明显优于基于普通 LDA 和基于 TF-IDF 的推荐方法, 其主要原因是基于 MTER 的推荐算法考虑到了当遇到没有情感标注的词, 即一个完全陌生的词, 如何通过主题模型计算出该词的情感值。相对于其他三种方法, MTER 考虑到当一个陌生的词在不同场景中可能会有情感差异, 并结合该词当前的主题分布实时计算出该词在当前视频片段中的情感向量。而基于 TF-IDF 的实验方法只考虑了各片段中词之间的数量关系和已有的情感词向量来评估片段的情感, 当面对情感词典数据缺乏的情况下无法对未知词的情感进行探索, 因此基于 MTER 的推荐方法优于其他对比方法。

结合上述结果分析, 不借助情感词典的方法 TF-IDF 和 LDA 在视频片段推荐中表现较差, 尤其是 $n_{top} = 15, 10$ 时; 而当 $n_{top} = 5$ 时, 四种方法的效果差距不大。MTER 在 n_{top} 为三种不同值的情况下均优于其他三种方法, 且在 $n_{top} = 10$ 时表现最佳。

5 结语

本文重点对视频片段的情感问题进行研究, 提出了基于弹幕情感分析的视频片段推荐机制。作为一种新类型的文本消息, 弹幕反映了视频的情感。本文提出了一种新颖的分析机制, 借助主题模型来挖掘弹幕文本的情感因子。此外, 还提出了基于全局特征和局部特征的视频片段情感相似度分析方法。实验结果表明了所提模型在分析不规则文本的复杂情感特征方面的能力, 以及弹幕在视频情感分析领域的适用性。

参考文献 (References)

- [1] WU B, ZHONG E, TAN B, et al. Crowdsourced time-sync video tagging using temporal and personalized topic modeling[C]// Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2014: 721 - 730.
- [2] LV G, XU T, CHEN E, et al. Reading the videos: temporal labeling for crowdsourced time-sync videos based on semantic embedding[C]// Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. Menlo Park, CA: AAAI Press, 2016: 3000 - 3006.
- [3] XIAN Y, LI J, ZHANG C, et al. Video highlight shot extraction with time-sync comment[C]// HOTPOST 2015: Proceedings of the 7th International Workshop on Hot Topics in Planet-scale Mobile Computing and Online Social Networking. New York: ACM, 2015: 31 - 36.
- [4] HAMASAKI M, TAKEDA H, HOPE T, et al. Network analysis of an emergent massively collaborative creation community[C]// Proceedings of the Third International ICWSM Conference. Menlo Park, CA: AAAI Press, 2009: 222 - 225.
- [5] WU Z, ITO E. Correlation analysis between user's emotional comments and popularity measures[C]// Proceedings of the 2014 IIAI 3rd International Conference on Advanced Applied Informatics. Piscataway, NJ: IEEE, 2014: 280 - 283.
- [6] YOSHII K, GOTO M. MusicCommentator: generating comments synchronized with musical audio signals by a joint probabilistic model of acoustic and textual features[EB/OL]. [2016-03-10]. <https://staff.aist.go.jp/m.goto/PAPER/ICEC2009yoshii.pdf>.
- [7] ZHAO J, LIU K, WANG G. Adding redundant features for CRFs-based sentence sentiment classification[C]// EMNLP 2008: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA, USA: Association for Computational Linguistics, 2008: 117 - 126.
- [8] ZHOU L, XIA Y, LI B, et al. WIA-opinmine system in NTCIR-8 MOAT evaluation[EB/OL]. [2016-03-10]. <http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings8/NTCIR/15-NTCIR8-MOAT-ZhouL.pdf>.
- [9] 李寿山, 黄居仁. 基于 Stacking 组合分类方法的中文情感分类研究[J]. 中文信息学报, 2010, 24(5): 56 - 61. (LI S S, HUANG C R. Chinese sentiment classification based on stacking combination method[J]. Journal of Chinese Information Processing, 2010, 24(5): 56 - 61.)
- [10] YU H, HATZIVASSILOGLU V. Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences[C]// EMNLP 2003: Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA, USA: Association for Computational Linguistics, 2003: 129 - 136.
- [11] 刘志明, 刘鲁. 基于机器学习的中文微博情感分类实证研究[J]. 计算机工程与应用, 2012, 48(1): 1 - 4. (LIU Z M, LIU L. Empirical study of sentiment classification for Chinese microblog based on machine learning[J]. Computer Engineering and Applications, 2012, 48(1): 1 - 4.)
- [12] BLEI D M, NG A Y, JORDAN M I. Latent Dirichlet allocation[J]. Journal of Machine Learning Research, 2003, 3: 993 - 1022.

(下转第 1134 页)

资源泄漏问题,但是面对使用 Button 进行相关资源操作的测试程序,它的检测准确率还有待提高。对 RelFix 的测试结果表明,RelFix 能够比较理想地实现资源泄漏的修复,并且修复过后的 Android 应用程序的原有功能没有受到影响。但是,它对资源泄漏检测的工具具有一定依赖性,资源泄漏检测工具的有效性也会对它的修复结果产生影响。

此外,ResLeakBench 不仅能够对本文使用的测试对象 Relda2 和 RelFix 进行测评,还可以对其他关于 Android 应用软件资源泄漏地检测和修复的工具进行测评。对于资源相关工具的使用者来说,ResLeakBench 能够为提供选择策略。用户可以利用本文构造出的基准测试集进行对多个工具进行验证,通过对实验结果的比较,得出比较符合他们要求的工具。对于资源工具的开发来说,他们可以利用 ResLeakBench 对工具进行改进。

参考文献 (References)

- [1] 杨娇. 基于 Android 平台手机通讯录的设计与实现[D]. 西安: 西安电子科技大学, 2013. (YANG J. Design and implementation for a mobile phone contacts on Android platform [D]. Xi'an: Xidian University, 2013.)
- [2] IDC. Smartphone OS market share, 2016 Q1 [EB/OL]. [2016-06-01]. <http://www.idc.com/promo/smartphone-market-share/os;jsessionid=B13EFFFFA3D9B1C2CB357D32DBF4204BA>.
- [3] WU T, LIU J, XU Z, et al. light-weight, inter-procedural and callback-aware resource leak detection for Android apps [J]. IEEE Transactions on Software Engineering, 2016, 42(11): 1054 - 1076.
- [4] LIU J, WU T, YAN J, et al. Fixing resource leaks in Android apps with light-weight static analysis and low-overhead instrumentation [C]// Proceedings of the 2016 IEEE 27th International Symposium on Software Reliability Engineering. Piscataway, NJ: IEEE, 2016: 342 - 352.
- [5] Flyingstarwb. 目前高性能计算基准测试程序分类[EB/OL]. [2016-04-01]. <http://blog.csdn.net/flyingstarwb/article/details/4225913/>.
- [6] 王良. Benchmark 性能测试综述[J]. 计算机工程与应用, 2006, 42(15): 45 - 48. (WANG L. Summa of benchmark performance test[J]. Computer Engineering and Applications, 2006, 42(15): 45 - 48.)
- [7] GUO C, ZHANG J, YAN J, et al. Characterizing and detecting resource leaks in Android applications [C]// Proceedings of the 2013 IEEE/ACM 28th International Conference on Automated Software Engineering. Piscataway, NJ: IEEE, 2013: 389 - 398.
- [8] ARZT S, RASTHOFER S, FRITZ C, et al. FlowDroid: precise context, flow, field, object-sensitive and lifecycle-aware taint analysis for Android apps [J]. ACM SIGPLAN Notices, 2014, 49(6): 259 - 269.
- [9] Google. App components [EB/OL]. [2016-04-01]. <https://developer.android.com/guide/components/index.html>.
- [10] 王铭. 回调函数在软件设计中的应用[J]. 河南教育学院学报(自然科学版), 2003, 12(3): 44 - 46. (WANG G. The application of callback function in software design[J]. Journal of Henan Education Institute (Natural Science Edition), 2003, 12(3): 44 - 46.)
- [11] Google. Activity lifecycle [EB/OL]. [2016-04-01]. <https://developer.android.com/training/basics/activity-lifecycle/index.html>.
- [12] Google. Dalvik bytecode [EB/OL]. [2016-04-01]. <https://source.android.com/devices/tech/dalvik/dalvik-bytecode.html>.
- [13] TUMBLESON C, WISNIEWSKI R. A tool for reverse engineering Android APK files [EB/OL]. [2016-04-01]. <https://ibotpeaches.github.io/Apktool/>.
- [14] 姚尚朗, 靳岩. Android 开发入门与实战[M]. 北京: 人民邮电出版社, 2009: 57 - 136. (YAO S L, JIN Y. Introduction and Practice of Android Development [M]. Beijing: Posts & Telecom Press, 2009: 57 - 136.)
- [15] ZHANG J. Resourcetable [EB/OL]. [2016-04-01]. <http://lcs.ios.ac.cn/~zj/ResourceTable.html>.

This work is partially supported by National Natural Science Foundation of China (91418206).

LIU Jierui, born in 1990, M. S. candidate. His research interests include program analysis and testing.

YAN Jun, born in 1980, Ph. D., associate research fellow. His research interests include program analysis and testing.

YANG Hongli, born in 1970, Ph. D. associate research fellow. Her research interests include program analysis and verification.

(上接第 1070 页)

- [13] HEINRICH G. Parameter estimation for text analysis [EB/OL]. [2016-03-10]. <http://www.arbylon.net/publications/text-est2.pdf>.
- [14] WU B, ZHONG E, HORNER A, et al. Music emotion recognition by multi-label multi-layer multi-instance multi-view learning [C]// Proceedings of the 22nd ACM International Conference on Multimedia. New York: ACM, 2014: 117 - 126.
- [15] REN J. Document for Ren-CECps 1.0 [EB/OL]. [2016-03-10]. <http://al-www.is.tokushima-u.ac.jp/member/ren/RenCECps1.0/Ren-CECps1.0.html>.
- [16] QUAN C, REN F. Sentence emotion analysis and recognition based on emotion words using Ren-CECps [J]. International Journal of Advanced Intelligence, 2010, 2(1): 105 - 117.
- [17] REN F, QUAN C. Linguistic-based emotion analysis and recognition for measuring consumer satisfaction: an application of affective computing [J]. Information Technology and Management, 2012, 13(4): 321 - 332.
- [18] 郑飏飏, 徐健, 肖卓. 情感分析及可视化方法在网络视频弹幕数据分析中的应用[J]. 现代图书情报技术, 2015, 31(11): 82 - 90. (ZHENG Y Y, XU J, XIAO Z. Utilization of sentiment analysis and visualization in online video bullet-screen comments [J]. New Technology of Library and Information Service, 2015, 31(11): 82 - 90.)

DENG Yang, born in 1991, M. S. candidate. His research interests include information retrieval, machine learning.

ZHANG Chenxi, born in 1960, Ph. D., professor. His research interests include distributed computing, embedded system.

LI Jiangfeng, born in 1983, Ph. D., lecturer. His research interests include distributed computing, social network computing.