# Public opinion monitoring through collective semantic analysis of tweets

**Article** *in* Social Network Analysis and Mining · July 2022

DOI: 10.1007/s13278-022-00922-8



**3 authors**, including:

Dionisis Karamouzas
Centre for Research and Technology Hellas



Ioannis Mademlis
Aristotle University of Thessaloniki



ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/362246843

Public opinion monitoring through collective semantic analysis of tweets

通过推文的集体语义分析进行舆论监测ArticleinSocial Network Analysis and Mining · July 2022

DOI: 10.1007/s13278-022-00922-8



3 authors, including:

Dionisis Karamouzas
Centre for Research and Technology Hellas



Ioannis Mademlis
Aristotle University of Thessaloniki

# Public opinion monitoring through collective semantic analysis of tweets

Dionysios Karamouzas[1†], Ioannis Mademlis[2*†] and Ioannis Pitas[2]

[1]Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece.
[2]Department of Informatics, Aristotle University of Thessaloniki, Thessaloniki, Greece.

*Corresponding author(s). E-mail(s): imademlis@csd.auth.gr;
Contributing authors: dionkara@ece.auth.gr; pitas@csd.auth.gr;
†The first two authors contributed equally and are joint first authors.

**Abstract**

The high popularity of Twitter renders it an excellent tool for political research, while opinion mining through semantic analysis of individual tweets has proven valuable. However, exploiting relevant scientific advances for collective analysis of Twitter messages in order to quantify general public opinion has not been explored. This paper presents such a novel, automated public opinion monitoring mechanism, consisting of a semantic descriptor that relies on Natural Language Processing (NLP) algorithms. A four-dimensional descriptor is first extracted for each tweet independently, quantifying text polarity, offensiveness, bias and figurativeness. Subsequently, it is summarized across multiple tweets, according to a desired aggregation strategy and aggregation target. This can then be exploited in various ways, such as training machine learning models for forecasting day-by-day public opinion predictions. The proposed mechanism is applied to the 2016/2020 US Presidential Elections tweet datasets and the resulting succinct public opinion descriptions are explored as a case study.

**Keywords:** public opinion, Twitter analysis, social media analysis, sentiment analysis, opinion mining, Deep Neural Networks





# 通过推文的集体语义分析进行舆论监测

Dionysios Karamouzas[1†], Ioannis Mademlis[2*†] 和Ioannis Pitas

[1]电气与计算机工程系，亚里士多德
萨洛尼卡大学，萨洛尼卡，希腊。
2信息学系，雅典大学塞萨洛尼基分校，
塞萨洛尼基，希腊。

*通讯作者。邮箱：imademlis@csd.auth.gr;
合作作者：dionkara@ece.auth.gr;pitas@csd.auth.gr;
†前两位作者贡献相同，并列第一作者。
作者。

**Abstract**

Twitter的极高人气使其成为政治研究的优秀工具，而通过单个推文的语义分析进行意见挖掘已被证明是有价值的。然而，利用相关科学进展对Twitter消息进行集体分析以量化公众舆论尚未被探索。本文提出了一种新颖的自动化舆论监测机制，该机制包含一个基于自然语言处理（NLP）算法的语义描述符。首先，为每条推文独立提取一个四维描述符，量化文本极性、攻击性、偏见和比喻性。随后，根据所需的聚合策略和聚合目标，对多条推文进行汇总。这可以以多种方式利用，例如训练机器学习模型以进行每日公众舆论预测。所提出的机制应用于2016/2020年美国总统选举推文数据集，并作为案例研究探索了由此产生的简洁舆论描述。

**关键词：** 舆论，Twitter分析，社交媒体分析，情感分析，意见挖掘，深度神经网络

# 1 Introduction

Social media have gradually risen to be central elements of modern life in the Western world. The easy access to on-line platforms and the benefits of constant social interaction keep the number of users steadily growing. They allow people to directly express their opinions and feelings using a variety of media, like text, images, videos, etc.

Consequently, these platforms can be used to monitor public opinion related to a subject of particular interest. Public opinion "represents the views, desires, and wants of the majority of a population concerning a certain issue, whether political, commercial, social, or other" [1]. Twitter seems to be the medium of choice for stating opinions regarding sociopolitical matters like COVID-19, elections, racism, etc. The massive number of people utilizing Twitter for staying up-to-date and expressing their views has provided politicians with the opportunity to put their message across quickly and cheaply, without going through the traditional media briefings and news conferences [2].

The potential of Twitter regarding political events was first highlighted during the US presidential elections of 2008, where Barack Obama used the platform efficiently for his campaign [3]. After that successful Twitter campaign, all major candidates and political parties quickly established a social media presence. Moreover, the popularity of Twitter provides a unique opportunity for e-government initiatives, especially with regard to simplified communication between government institutions and citizens [4]. This may allow for greater transparency and increased citizen confidence in local institutions.

Given this very high potential of Twitter, opinion mining of tweets can provide us with valuable information. Automated semantic text analysis tools, relying on modern Artificial Intelligence (AI)-based Natural Language Processing (NLP) algorithms, can identify the mood of a tweet (e.g., polarity/sentiment [5]) with remarkable precision. Deep Neural Networks (DNNs) have greatly advanced the relevant state-of-the-art, especially in *sentiment analysis*: the task of assigning a class label to a corpus of written text, where each class expresses a possible sentiment of the author concerning the content of the text. Sentiment may simply be *polarity* (ranging from very negative to very positive attitude), or multi-dimensional (identifying the presence or absence of different emotions). Additional semantic text properties that are correlated with opinion, besides sentiment, can also be identified using almost identical algorithms (for instance, *bias* or *sarcasm*).

Thus, for instance, modern AI makes it possible to get a feeling about which candidate is more likely to win the next elections by analyzing the sentiment of related tweets [6]. Moreover, opinion mining can be used to determine the public's views regarding a crucial matter, e.g., a referendum [7], and help the government take the right decisions.

However, collective and multidimensional semantic analysis of tweets, based on state-of-the-art DNNs, in order to quantify and monitor general public opinion has not been significantly explored, despite the obvious potential. Most



# 1 引言

社交媒体逐渐成为西方现代生活中的核心要素。在线平台的便捷访问和持续社交互动的好处使用户数量稳步增长。它们允许人们使用各种媒体（如文本、图像、视频等）直接表达他们的观点和感受。

因此，这些平台可以用来监测与特定主题相关的舆论。舆论“代表了一个国家的大多数人在某个问题上（无论政治、商业、社会或其他）的观点、愿望和需求”[1]。推特似乎已成为表达关于COVID-19、选举、种族等社会政治问题的观点的首选媒介。大量使用推特获取最新信息并表达观点的人们，为政治家提供了快速且廉价地传递信息的机会，而无需通过传统的媒体简报和新闻发布会 [2]。

Twitter在政治事件方面的潜力最早在2008年美国总统选举期间被凸显，当时巴拉克·奥巴马有效地利用该平台进行他的竞选活动 [3]。在那次成功的Twitter竞选活动之后，所有主要候选人和政党迅速建立了社交媒体存在感。此外，Twitter的流行为电子政务计划提供了一个独特的机会，特别是在简化政府机构与公民之间的沟通方面 [4]。这可能允许更大的透明度和当地机构公民信心的增加。

鉴于Twitter的这种巨大潜力，推文的意见挖掘可以为我们提供有价值的信息。基于现代人工智能（AI）的自然语言处理（NLP）算法的自动化语义文本分析工具，可以以惊人的精度识别推文的情绪（例如，极性/情绪 [5]）。深度神经网络（DNN）极大地推动了相关最先进技术，特别是在<样式 id='2'>情绪分析</样式>方面：将一个类标签分配给书面文本语料库的任务，其中每个类表示作者对文本内容的可能情绪。情绪可能简单地是<样式 id='4'>极性</样式>(从非常负面到非常积极的态度)，或者是多维度的（识别不同情绪的存在或不存在）。除了情绪之外，与意见相关的其他语义文本属性也可以使用几乎相同的算法来识别（例如，<样式 id='6'>偏见</样式>或<样式 id='8'>讽刺</样式>）。

因此，例如，通过分析相关推文的情感，现代人工智能可以让人感受到哪个候选人更有可能赢得下一届选举 [6]。此外，意见挖掘可用于确定公众对某个关键事项的看法，例如一项全民公投 [7],，并帮助政府做出正确的决策。

然而，尽管具有明显的潜力，基于最先进的深度神经网络进行推文的集体和多维度语义分析，以量化并监测公众舆论，尚未得到显著探索。大多数

related methods only extract limited amounts of semantic content (typically polarity), instead of multiple semantic dimensions, while the tweets are processed individually; the outcomes are simply summarized for manual human overview. Automated collective analysis of social media-extracted content from multiple semantic aspects, in order to identify tendencies in the overall public opinion as a whole, is rather scarce.

This paper attempts to investigate the relevant unexplored possibilities of multidimensional collective tweet analysis through state-of-the-art DNNs. Thus, a novel, automated public opinion monitoring mechanism is proposed, consisting of a composite, quantitative, semantic descriptor that relies on DNN-enabled Natural Language Processing (NLP) algorithms. A four-dimensional vector, i.e., an instance of the proposed descriptor, is first extracted for each tweet independently, quantifying *text polarity, offensiveness, bias and figurativeness*. Subsequently, the computed descriptors are summarized across multiple tweets, according to a desired aggregation strategy (e.g., arithmetic mean) and aggregation target (e.g., a specific time period). This can be exploited in various ways; for example, aggregating the tweets of each days separately allows us to construct a multivariate timeseries which can be used to train a forecasting AI algorithm, for day-by-day public opinion predictions [8]. In order to evaluate the usefulness of the proposed mechanism, it has been applied to the large-scale 2016 US Presidential Elections tweet dataset. The resulting succinct public opinion descriptions are explored as a case study.

The remainder of this paper is organized in the following manner. Section 2 discusses related previous literature, focusing not on NLP or timeseries forecasting methods (which are exploited by us in a black-box manner) but on various existing mechanisms for extracting and monitoring public opinion by applying NLP and/or timeseries forecasting on social media posts. Section 3 presents the proposed semantic descriptor of public opinion. Section 4 discusses experimental evaluation on the 2016/2020 US Presidential Elections tweet datasets and the succinct public opinion descriptions that were derived using the proposed mechanism. Subsequently, Section 5 discusses key-findings on the employed datasets which were extracted using the proposed mechanism, as well as the latter's main novel contributions. Finally, Section 6 draws conclusions from the preceding presentation.

## 2 Related Work

This Section presents a brief overview of existing AI-based (NLP and/or DNN) approaches to semantic analysis of social media text for: a) quantitative description of public opinion, and b) timeseries forecasting.

### 2.1 Public opinion description

There is growing scientific interest on analyzing social media posts since the early 2010s, with Twitter dominating relevant research. For instance, [9] evaluated current public opinion regarding political advertising on Facebook, by



相关方法仅提取有限的语义内容（通常为极性），而不是多个语义维度，而推文是单独处理的；结果仅简单地总结供人工查看。从多个语义方面自动分析社交媒体提取的内容，以识别整体舆论的整体趋势，相对较少。

本文试图通过最先进的DNN研究多维集体推文分析的相关未开发可能性。因此，提出了一种新颖的自动化舆论监测机制，该机制由一个依赖于DNN支持的自然语言处理（NLP）算法的复合、定量、语义描述符组成。首先，为每条推文独立提取一个四维向量，即所提出描述符的实例，量化文本极性、攻击性、偏见和比喻性。随后，根据所需的聚合策略（例如，算术平均值）和聚合目标（例如，特定时间段），将计算出的描述符跨多条推文进行总结。这可以以各种方式利用；例如，分别聚合每天推文，使我们能够构建一个多变量时间序列，该时间序列可用于训练预测人工智能算法，进行每日舆论预测 [8]。为了评估所提出机制的有效性，它已应用于大规模2016年美国总统选举推文数据集。生成的简洁舆论描述作为案例研究进行了探索。

本文的其余部分按以下方式组织。第2 节讨论了相关的先前文献，重点关注不是 NLP 或时间序列预测方法（我们以黑盒方式利用这些方法），而是应用 NLP 和/或时间序列预测在各种现有机制上提取和监测社交媒体帖子上的公众舆论。第3节介绍了所提出的公众舆论语义描述符。第 4 节讨论了在 2016/2020 年美国总统选举推文数据集上进行的实验评估，以及使用所提出机制推导出的简洁公众舆论描述。随后，第5 节讨论了使用所提出机制从所采用数据集中提取的关键发现，以及该机制的主要创新贡献。最后，第 6 节总结了前面的介绍。

## 2 相关工作

本节简要概述了现有的基于 AI（NLP 和/或 DNN）的方法，用于对社交媒体文本进行语义分析，以实现：a) 公众舆论的定量描述，以及 b) 时间序列预测。

### 2.1 意见描述

自2010年代初以来，分析社交媒体帖子引起了越来越多的科学兴趣，而Twitter主导了相关研究。例如，[9]评估了当前关于Facebook政治广告的公众意见，通过

automatically extracting topics of discussion from relevant tweets published in October 2019. Manual inspection of these topics was conducted, leading to the conclusion that user perception of Facebook advertising is gradually decreasing, as a result of privacy concerns related to trust in the platform. Similarly, having the goal of analyzing tweets from the candidates' perspective, [10] and [11] processed the content of Donald Trump's and Hilary Clinton's tweets during the US 2016 presidential elections. The goal was to qualitatively identify which issue each candidate emphasized and what communication strategies they used.

However, this paper does not concern topic modeling and manual inspection of identified topics. Instead, it relates to methods that exploit semantic content attributes of tweets to quantify public opinion in an automatic manner. These methods can be broadly categorized into: a) non-semantic ones, which do not perform AI-based semantic analysis on tweets, and b) semantic ones, which typically perform a type of AI-enabled sentiment analysis/opinion mining on tweets.

### 2.1.1 Non-semantic methods

Non-semantic methods only consider keyword frequencies and/or tweet volume, resulting in rather inaccurate and/or purely qualitative insights. For instance, [12] performed a statistical analysis on tweets from the Spanish 2019 presidential campaign, which were selected based on keywords and quantified through their volume over time. The goal was to reveal political discourse of the parties engaged and highlight the main messages conveyed and their resulted impact in the share of candidates' voice. Machine learning classifiers were only used to detect spammers and the conclusions were purely qualitative. Similarly, [13] investigated social media activity during a political event, by analyzing the US 2016 GOP debate and observing the volumes of special keywords in Twitter posts.

Evidently, the mechanism proposed in this paper is entirely different in nature from these non-semantic approaches: it assigns each tweet a 4D semantic numerical vector acquired with DNN/NLP-based text classifiers, thus going a step further from conventional statistical approaches, and then aggregates these outputs for all tweets in order to construct an overall, quantitative public opinion descriptor.

### 2.1.2 Semantic methods without aggregation

Semantic methods are more advanced and provide more accurate results. The majority of such methods operate on Twitter, but do not treat the relevant tweets collectively and mostly just consider the volume of tweets per sentiment class. For instance, [1] presented a framework for monitoring the evolution (16 months) of public opinion related to the topic of climate change. The proposed mechanism is able to on-the-fly identify and monitor sentiment in a desired set of individual tweets, possibly as they are being published, but only rudimentary analysis is performed to these outcomes as a collection. Thus, public





自动从2019年10月发布的相关推文中提取讨论主题。对这些主题进行了人工检查，得出结论：由于对平台信任相关的隐私问题，用户对Facebook广告的感知逐渐下降。同样，为了从候选人的角度分析推文，[10]和[11] 处理了美国2016年总统选举期间唐纳德·特朗普和希拉里·克林顿的推文内容。目标是定性识别每位候选人强调的问题以及他们使用的沟通策略。

然而，本文不涉及主题建模和已识别主题的人工检查。相反，它涉及利用推文的语义内容属性以自动方式量化舆论的方法。这些方法可以大致分为：a) 非语义方法，这些方法不对推文进行基于人工智能的语义分析，以及 b) 语义方法，这些方法通常对推文执行某种基于人工智能的情感分析/观点挖掘。

### 2.1.1 非语义方法

非语义方法仅考虑关键词频率和/或推文数量，导致结果较为不准确和/或纯粹定性的洞察。例如，[12] 对西班牙2019年总统选举的推文进行了统计分析，这些推文是根据关键词选择并随着时间的推移通过其数量进行量化的。目标是揭示参与各方的政治话语，突出传达的主要信息及其对候选人声音份额的影响。机器学习分类器仅用于检测垃圾邮件发送者，结论纯粹是定性的。类似地，[13] 通过分析美国2016年共和党辩论和观察推文中特殊关键词的量，研究了政治事件期间社交媒体活动。

显然，本文提出的机制在本质上与这些非语义方法完全不同：它为每条推文分配一个由DNN/NLP-based文本分类器获得的4D语义数值向量，从而超越了传统的统计方法，然后对所有推文的输出进行聚合，以构建一个整体的、定量的公众意见描述符。

### 2.1.2 无聚合的语义方法

语义方法更先进，并提供更准确的结果。此类方法的大多数都在Twitter上运行，但不会将相关的推文集体处理，并且主要只是考虑每个情感类别的推文数量。例如，[1] 提出了一种监控与气候变化主题相关的公众舆论演变（16个月）的框架。所提出的机制能够实时识别和监控所需一组单个推文中的情感，可能是在它们发布时，但仅对这些结果进行初步分析，作为集合。因此，公众

opinion as a whole is scarcely considered. Having a different goal in mind, [14] conducted sentiment analysis of tweets for predicting election outcome. A simple CNN model was employed for that task, while the total volume of tweets (non-semantic attribute) was compared against sentiment polarity (volume of positive tweets) to find out which is a better predictor of election results. The semantic descriptor was found to be more accurate. Advancing on this line of research, [15] aimed to predict not only the winner but also the voting share of each candidate in the 2019 Spanish Presidential elections, by considering the volume of positive tweets per candidate. Still, the semantic attributes themselves were not aggregated over the set of all tweets.

In a similar manner, [16] performed opinion mining on Italian tweets about vaccination for a 12-month period and simply counted the number of tweets per polarity class ("against", "in favor" or "neutral") for each month. It was found that vaccine-related events influenced the distribution of polarity classes, but no aggregation of the semantic content was performed. Under an almost identical general idea, [17] aimed to determine the critical time window of public opinion concerning an event, by applying multi-emotional sentiment classification to microblog posts in Sina Weibo (published within a short time period of approximately 10 days after certain events). The volume of tweets per class (among the employed 7 emotion classes) was simply examined to find out that monitoring the negative emotions trend is crucial for predicting the influence of events.

### 2.1.3 Semantic methods with aggregation

In contrast to these approaches, a set of more advanced semantic methods do perform aggregation of the semantic content and thus treat the tweets collectively, by constructing a semantic public opinion descriptor in the form of a low-dimensional timeseries. This is exactly the method family to which the mechanism proposed in this paper belongs to. For instance, [18] explored the change of public sentiment in China after "Wenzhou Train Collision", by performing sentiment analysis on posts from the Sina Weibo microblogging platform, by aggregating eight identified emotions per tweet over time in order to produce an 8D daily vector, being monitored as 8 separate timeseries for a 10 day interval. However, sentiment analysis accuracy at the time was not high and the results were not particularly useful. Similarly, in [19], tweets about the 2017 Anambra State gubernatorial election in Nigeria are semantically analyzed and the outcomes are aggregated for every two-hour interval posts. The produced time-series cover an 18-hour time-frame on the election day.

Operating also in this direction, [20] employed tweet semantic analysis and aggregation to construct an average daily sentiment timeseries for each party, covering a 21-day pre-election period during the US presidential elections. Finally, in [21], similar ideas were applied to the prediction of cryptocurrency price returns through collective semantic analysis of tweets. Relevant timeseries were constructed through day-by-day aggregation of individual tweet semantic outcomes augmented with financial data, covering a period of 2 months, and a





舆论作为一个整体很少被考虑。出于不同的目标，[14]对推文进行了情感分析以预测选举结果。为此任务采用了简单的CNN模型，同时将推文总量（非语义属性）与情感极性（正面推文的数量）进行比较，以找出哪个是选举结果的更好预测指标。语义描述符被证明更准确。沿着这条研究路线，[15] 旨在预测2019年西班牙总统选举的获胜者以及每位候选人的投票份额，方法是考虑每位候选人的正面推文数量。然而，语义属性本身并没有在所有推文的集合上聚合。

以类似的方式，[16] 对意大利关于疫苗接种的推文进行了12个月的舆论挖掘，并简单地计算了每个月每个情感类别（“反对”、“支持”或“中立”）的推文数量。发现与疫苗相关的事件影响了情感类别的分布，但没有对语义内容进行聚合。在几乎相同的一般思想下，[17] 旨在确定公众对某事件的关注关键时间窗口，方法是应用多情感情感分类到新浪微博（在特定事件发生后约10天内发布的微博）。简单地检查了每个类别（在所使用的7个情感类别中）的推文数量，发现监测负面情绪趋势对于预测事件的影响至关重要。

### 2.1.3 带有聚合的语义方法

与这些方法不同，一组更高级的语义方法确实执行语义内容的聚合，从而将推文集体处理，通过构建低维时间序列形式的语义公众意见描述符。这正是本文提出的机制所属的方法系列。例如，[18] 探讨了“温州动车追尾”后中国公众情绪的变化，通过对新浪微博平台上的帖子进行情感分析，通过在时间上聚合每条推文中的八个已识别情绪，以产生一个8D的每日向量，该向量被监控为10天间隔的8个单独时间序列。然而，当时的情感分析准确率并不高，结果也并不特别有用。类似地，在 [19]，中，尼日利亚2017年安南布拉州州长选举的推文被语义分析，并且结果被聚合到每个两小时间隔的帖子中。产生的时间序列覆盖了选举日上的18小时时间范围。

也在这个方向上操作，[20] 采用推文语义分析和聚合来构建每个政党的平均每日情绪时间序列，涵盖美国总统选举期间21天的预选期。最后，在 [21]，中，类似的方法被应用于通过推文的集体语义分析来预测加密货币价格回报。通过每日聚合单个推文的语义结果并辅以金融数据构建相关时间序列，涵盖2个月的时间，并

learning model was trained for timeseries forecasting. Twitter-derived public sentiment was found to indeed have predictive power, but it was not enough on its own for accurate forecasting. Overall, methods of this type are most similar to ours, but the scale of experimental evaluation (e.g., temporal duration of constructed timeseries) is significantly limited compared to this paper. Other important differences are discussed in the following Subsection.

### 2.1.4 Semantic analysis dimensions and algorithms

The vast majority of the semantic methods presented above only utilize tweet sentiment, thus they can be considered as exploiting one-dimensional text semantics. This is most obvious in cases where the semantic analysis outputs a polarity (e.g., binary or ternary classification into positive/negative tweets, or into positive/neutral/negative ones). This limited approach is the most dominant one [14] [15] [20] [16] [21]. However, one-dimensional sentiment analysis can be considered to be the case even when multi-emotional classifiers are being employed instead of simple polarity. E.g., in [18] (expect, joy, love, surprise, anxiety, sorrow, angry and hate), [17] (happiness, like, sadness, disgust, astonishment, anger, and fear) and [1] (joy, inspiration, anger, discrimination, support). Although it is a more nuanced approach, these emotions still fall under the general umbrella of sentiment, thus these methods keep ignoring other semantic text attributes. The only case where multidimensional semantics are considered is [19], where 2 different opinion dimensions (polarity and bias) are both taken into account. In contrast, this paper proposes *a 4-dimensional mechanism jointly considering polarity, bias, figurativeness and offensiveness, which are all different text attributes, and experimentally verifies their usefulness.*

Another relevant aspect is how semantic tweet analysis is performed in such published methods. The vast majority among them employ outdated algorithms for text description, relying on lexicons or older representations. E.g., [18] uses the HowNet lexicon [22], [17] uses an emotional dictionary [23] and a negative word dictionary, [20] uses SentiStrength [24], [16] uses a Bag-of-Words approach, etc. Only a few rely on modern DNN-based word representation schemes, such as [17] and [1] which exploit Word2Vec [25]. The situation is even more dire when examining the type of learning models employed for actual semantic analysis. Almost all of the presented methods utilize outdated approaches, such as [17], [16] or [19], which exploit a simple K-Nearest Neighbours (KNN) classifier, a Support Vector Machine (SVM) or Textblob's Naive Bayes Classifier[1], respectively. [15] evaluates a variety of traditional (non-neural) machine learning algorithms, while [21] exploits a sentiment analysis rule set [26]. Recent DNN-based learning models were only exploited in [1] (Bidirectional Long Short-Term Memory network) and [14] (Convolutional Neural Network).

[1] https://textblob.readthedocs.io/en/dev/





为时间序列预测训练了学习模型。研究发现，推文衍生公共情绪确实具有预测能力，但仅凭它本身无法进行准确预测。总体而言，这类方法与我们的方法最相似，但与本文相比，实验评估的规模（例如，构建的时间序列的时间长度）明显有限。其他重要差异将在以下小节中讨论。

### 2.1.4 语义分析维度和算法

上述大多数语义方法仅利用推文情绪，因此它们可以被视为利用一维文本语义。这在语义分析输出极性（例如，二元或三元分类为正面/负面推文，或为正面/中性/负面）的情况下最为明显。这种有限的方法是最主流的 [14] [15] [20] [16] [21]。然而，即使使用多情感分类器而不是简单的极性，一维情绪分析也可以被认为是这种情况。例如，在 [18] (期待、喜悦、爱、惊讶、焦虑、悲伤、愤怒和憎恨)， [17] (快乐、喜欢、悲伤、厌恶、惊讶、愤怒和恐惧)以及 [1] (喜悦、灵感、愤怒、歧视、支持)。尽管这是一种更细致的方法，但这些情绪仍然属于广义上的情绪，因此这些方法仍然忽略了其他语义文本属性。唯一考虑多维语义的情况是 [19], 同时考虑了2个不同的意见维度（极性和偏见）。相比之下，本文提出了一个联合考虑极性、偏见、形象性和攻击性的四维机制，这些都是不同的文本属性，并实验验证了它们的有效性。

另一个相关方面是如何在已发布的这些方法中执行推文的语义分析。其中绝大多数采用过时的文本描述算法，依赖于词汇表或旧有的表示方法。例如，[18] 使用HowNet词汇表， [22], [17] 使用一个情感词典， [23]和一个负面词词典， [20] 使用SentiStrength， [24], [16] 使用词袋模型，等等。只有少数依赖于现代基于深度神经网络（DNN）的词表示方案，例如 [17]和 [1] 利用Word2Vec [25]。当考察实际语义分析所采用的学习模型类型时，情况甚至更加糟糕。几乎所有提出的方法都利用过时的方法，例如 [17], [16] 或 [19], ，它们分别利用简单的K-最近邻（KNN）分类器、支持向量机（SVM）或Textblob的朴素贝叶斯分类器[1]。 [15] 评估了多种传统的（非神经）机器学习算法，而 [21] 利用了一个情感分析规则集 [26]。最近的基于深度神经网络的学习模型仅在[1] (双向长短期记忆网络)和 [14] (卷积神经网络)中被利用。

[1] https://textblob.readthedocs.io/en/dev/

Contrary to all of the above approaches, the mechanism proposed in this paper relies end-to-end on *state-of-the-art DNN solutions, both for word representation and for semantic analysis.*

## 2.2 Timeseries forecasting

Forecasting of timeseries derived by sentiment analysis of tweets has mainly been previously employed for predicting future financial indices. Thus, [27] explored the effect of different major events occurring during 2012–2016 on stock markets. A similar approach was followed in [28]. [29] examined the use of polarity values, extracted from tweets about the United States foreign policy and oil companies, in order to forecast the direction of weekly WTI crude oil prices. Finally, [30] investigated whether a public polarity indicator extracted from daily tweets on stock market or movie box office can indeed improve the forecasting of a related timeseries.

In all of these cases, the only exploited semantics dimension was polarity. It was shown that forecasting accuracy improves by using polarity, but public opinion extracted through tweets was only employed as an auxiliary information source for a financial-domain task, complementing a main source (e.g., stock exchange data in [27]). Moreover, outdated word representation and opinion mining algorithms were employed in all of these papers: [27] utilized SentiWordNet[2] lexicons, [29] exploited SentiStrength and the Stanford NLP Sentiment Analyzer[3], while [30] used a Naive Bayes classifier.

Similarly, older algorithms were mainly used for the timeseries forecasting task itself. [27] exploited linear regression and Support Vector Regression (SVR), [29] utilized SVM, Naïve Bayes and Multi-Layer Perceptron (MLP) learning models, while [30] relied on linear regression, MLP and SVMs to predict the target timeseries' immediate trajectory, considering the polarity or volume of tweets as input features.

In contrast to the above methods for financial forecasting exploiting tweet-derived polarity estimations, *this paper focuses on the proposed semantic public opinion descriptor itself and its potential uses for political analysis (including forecasting of public opinion).* This descriptor compactly captures *multiple opinion/semantic dimensions, instead of simply polarity.* Timeseries forecasting is utilized as an example application, among others, and *state-of-the-art DNN models are exploited during all stages, in contrast to previous methods in the literature.*

# 3 Proposed Mechanism

The proposed novel automated public opinion monitoring mechanism consists of a composite, quantitative, semantic descriptor that relies on NLP/DNN-based classifiers. By utilizing them, a four-dimensional vector, i.e., an instance of the proposed descriptor, is first extracted for each tweet independently, thus

[2] https://github.com/aesuli/SentiWordNet
[3] https://nlp.stanford.edu/sentiment/

Springer Nature 2021 LaTeX template

通过推文的集体语义分析进行舆论监测 7

与所有上述方法相反，本文提出的方法完全依赖于 最先进的*DNN*解决方案，无论是词表示还是语义分析。

## 2.2 时间序列预测

通过推文的情感分析得到的时间序列预测主要以前用于预测未来的金融指数。因此， [27]研究了2012-2016年间发生的不同重大事件对股票市场的影响。类似的方法在 [28]中采用了。 [29] 研究了从关于美国外交政策和石油公司的推文中提取的极性值，以预测每周WTI原油价格的方向。最后， [30] 研究了从每日关于股票市场或电影票房的推文中提取的公共极性指标是否确实可以提高相关时间序列的预测。

在所有这些情况下，唯一利用的语义维度是极性。结果表明，使用极性可以提高预测精度，但通过推文提取的舆论仅作为金融领域任务的辅助信息来源使用，补充主要信息源（例如， [27]中的证券交易所数据）。此外，所有这些论文都采用了过时的词表示和意见挖掘算法: [27]使用了SentiWordNet2 词汇表， [29] 利用了SentiStrength和斯坦福NLP情感分析器3，而 [30]使用了朴素贝叶斯分类器。

类似地，旧算法主要用于时间序列预测任务本身。 [27] 利用了线性回归和支持向量回归（SVR）， [29] 使用了SVM、朴素贝叶斯和多层感知器（MLP）学习模型，而 [30] 则依赖线性回归、MLP和SVM来预测目标时间序列的即时轨迹，并将推文的情感或数量作为输入特征。

与上述利用推文衍生情感估计进行金融预测的方法不同，本文关注所提出的语义公众意见描述符本身及其在政治分析（包括公众意见预测）中的潜在应用。该描述符紧凑地捕获了多个意见/语义维度，而不仅仅是情感。时间序列预测被用作一个示例应用，等等，在所有阶段都利用了最先进的*DNN*模型，这与文献中的先前方法不同。

# 3 Proposed Mechanism

所提出的创新自动化公众意见监测机制由一个复合的、定量的、语义描述符组成，该描述符依赖于基于NLP/DNN的分类器。通过利用它们，首先为每条推文独立提取一个四维向量，即所提出描述符的一个实例，从而

2 https://github.com/aesuli/SentiWordNet 3 https://nlp.stanford.edu/sentiment/

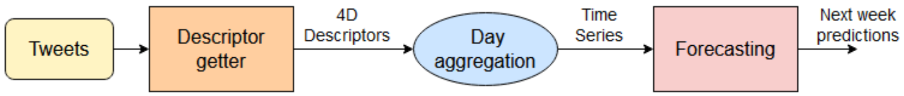


**Fig. 1** Quantitative public opinion forecasting using the proposed mechanism/semantic descriptor.

quantifying *text polarity, offensiveness, bias and figurativeness*. Subsequently, the computed descriptors are summarized across multiple tweets, according to a desired aggregation strategy (e.g., arithmetic mean) and aggregation target (e.g., a specific time period). The summarized descriptors can be exploited in various ways: for instance, by aggregating them on a day-by-day basis allows us to construct a multivariate timeseries which can be used to train a forecasting DNN for predicting future summarized descriptors. As an example, the pipeline of such a public opinion forecasting application/case study employing our proposed mechanism is depicted in Figure 1.

Below, the steps of computing the proposed public opinion descriptor are analyzed in detailed, along with the algorithmic machinery for implementing the process.

## 3.1 Step 1: Selecting the desired pool of tweets

The Twitter API allows easy and automated extraction of tweets based on manually set criteria about their topic and date. For instance, the presence of specific keywords and/or hashtags, the tweet timestamp, the fact of having been published within a desired range of dates, etc. Monitoring public opinion concerning an issue (e.g., attitude towards the incumbent party during an extended pre-election period) evidently requires smart adjustment of these criteria, so that an actually relevant corpus of user messages can be obtained. However, in general, this is extremely straightforward and simple, therefore we will not elaborate further.

## 3.2 Step 2: Individual descriptor extraction per tweet

The second step of the proposed mechanism is to semantically describe each tweet from the selected message pool as a 4-dimensional (4D) real-valued opinion vector. This description vector is separately extracted for each Twitter message. A set of 4 pretrained DNN models are employed to this end. Based on the state-of-the-art in NLP, two different neural architectures were employed. The hybrid CNN-LSTM from [31] was separately trained three time ex nihilo, using three different public annotated datasets for recognizing *offensiveness*, *bias* and *figurativeness* (sarcasm, irony and/or metaphor) in tweets. Additionally, a state-of-the-art pretrained, publicly available neural model[4] was employed for *polarity* recognition. In all four cases, the desired

[4] https://github.com/DheerajKumar97/US-2020-Election-Campaign-Youtube-Comments-Sentiment-Analysis



8 *Public opinion monitoring through collective semantic analysis of tweets*

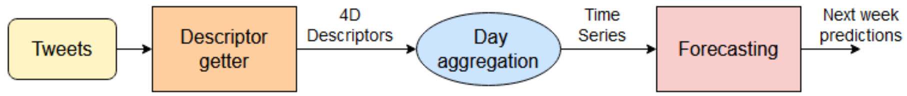


**图1** 使用所提出的机制/语义描述符进行定量公众舆论预测。

量化 文本极性、攻击性、偏见和比喻性。随后，根据所需的聚合策略（例如，算术平均值）和聚合目标（例如，特定时间段），将计算出的描述符跨多个推文进行汇总。汇总的描述符可以以多种方式利用：例如，通过按日汇总，我们可以构建一个多变量时间序列，该时间序列可用于训练预测未来汇总描述符的预测DNN。作为示例，采用我们提出的机制的公众舆论预测应用程序/案例研究的流程图如图1。

下面，详细分析了计算所提出的公众舆论描述符的步骤，以及实现该过程的算法机制。

## 3.1 第1步：选择所需的推文池

Twitter API允许根据手动设置的关于其主题和日期的标准轻松且自动地提取推文。例如，特定关键词和/或标签的存在、推文时间戳、在所需日期范围内发布的事实等。监控关于某个问题（例如，在延长选举前的时期内对现任党的态度）的公众舆论显然需要对这些标准进行智能调整，以便获得实际相关的用户消息语料库。然而，通常情况下，这非常直接和简单，因此我们不会进一步阐述。

## 3.2 第 2 步：针对每条推文提取个体描述符

所提出机制的第二步是从选定的消息池中，将每条推文语义描述为一个四维（4D）实值意见向量。该描述向量针对每条 Twitter 消息分别提取。为此，采用了四套预训练的 DNN 模型。基于 NLP 的最新技术，采用了两种不同的神经网络架构。来自 [31] 的混合 CNN-LSTM 架构分别使用三种不同的公共标注数据集，从零开始训练三次，用于识别攻击性、偏见 和 形象性 (讽刺、反讽和/或隐喻)在推文中的存在。此外，还采用了最新预训练的、公开可用的神经网络模型[4] 用于 极性 识别。在所有四种情况下，所需的

4 https://github.com/DheerajKumar97/US-2020-Election-Campaign-Youtube-Comments-Sentiment-Analysis

task was posed as binary text classification, with corresponding tweet labels (offensive/non-offensive, biased/non-biased, figurative/literal, negative/positive). Importantly, forcing classification to be as uncomplicated as possible (discriminating between two classes is typically easier than discriminating between multiple classes) renders the employed DNNs more robust, accurate and dependable for the actual problem tackled by the proposed mechanism, i.e., public opinion monitoring. All trained classifiers output a real value within the range $[0, 1]$ for each test tweet, with a value of 0/1 implying perfect and uncontested assignment of one of the two opposite labels (e.g., 0/1 means definitely and fully negative/positive sentiment, respectively, in the case of polarity).

Practical details about training the four DNNs that form the backbone of the proposed mechanism follow below.

#### 3.2.1 Training Datasets

**SemEval-2019 Task 6 sub-task A (S19-T6)** [32]: This dataset contains 14,100 tweets annotated as offensive/non-offensive and was used for training the offensiveness recognition DNN.

**Political Social Media Posts** (PSMP) from Kaggle[5] : This dataset contains 5,000 messages from Twitter and Facebook annotated as neutral/partisan and was used to create a bias recognition DNN. The presence of Facebook messages in the dataset did not pose a problem, as they also lie in the general category of short opinionated texts, similarly to tweets.

**Tweets with Sarcasm and Irony** (TSI) [33]: This dataset contains approximately 76,000 tweets annotated as ironic/sarcastic/figurative/literal. In the context of this paper, the first three classes were grouped in a single class called “figurative”, so that a binary figurativeness recognition DNN could be trained.

**YouTube Comments** (YTC): Moving on to the pretrained polarity recognition DNN, it was originally trained on a dataset with 12,559 YouTube comments. The comments were scrapped from 8 different YouTube videos related to the 2020 US presidential elections. Annotation was performed automatically via TextBlob[6] and so a positive/negative label was assigned to each comment.

#### 3.2.2 Text Preprocessing

Identical preprocessing was applied to the three training datasets, i.e., stopwords, hashtags, mentions and URLs were removed. These entities either provide us no semantic information, or only encode information about the discussed topic. However, the proposed mechanism assumes that the topic has been manually selected by the user (in Step 1); therefore, the presented automated individual tweet descriptor relying on the four pretrained DNNs

[5] https://www.kaggle.com/crowdflower/political-social-media-posts
[6] https://textblob.readthedocs.io/en/dev/

Springer Nature 2021 LaTeX template

*Public opinion monitoring through collective semantic analysis of tweets* 9

该任务被设定为二元文本分类，相应的推文标签（攻击性/非攻击性、有偏见/无偏见、比喻性/字面性、负面/正面）。重要的是，将分类尽可能简化（区分两类通常比区分多类更容易）使得所使用的DNNs对实际问题更具鲁棒性、准确性和可靠性，该问题由所提出的机制处理，即舆论监测。所有训练好的分类器对每个测试推文输出一个值，该值在 $[0, 1]$ 范围内，值为0/1表示对两个相反标签的完美且无争议的分配（例如，0/1分别表示在极性情况下绝对和完全的负面/正面情绪）。

关于训练构成所提出机制核心的四个DNN的实用细节如下。

#### 3.2.1 训练数据集

**SemEval-2019 Task 6 sub-task A (S19-T6)** [32]: 该数据集包含14,100条被标注为攻击性/非攻击性的推文，用于训练攻击性识别DNN。

**政治社交媒体帖子**(PSMP) from Kaggle[5] : 该数据集包含来自推特和脸书的5,000条消息，被标注为中立/党派性，并用于创建一个偏见识别DNN。数据集中包含脸书消息并非问题，因为它们也属于短篇观点性文本的一般类别，类似于推文。

**带有讽刺和反讽的推文**(TSI) [33]: 该数据集包含约76,000条被标注为讽刺/反讽/比喻/字面意义的推文。在本文的背景下，前三个类别被归为一个名为“比喻”的单一类别，以便训练一个二元比喻识别DNN。

**YouTube评论**(YTC): 转向预训练的情感识别DNN，它最初在一个包含12,559条YouTube评论的数据集上进行训练。这些评论是从与2020年美国总统选举相关的8个不同YouTube视频上抓取的。注释是通过TextBlob[6] 自动进行的，因此每条评论都被分配了一个正面/负面标签。

#### 3.2.2 文本预处理

对三个训练数据集应用了相同的预处理，即移除了停用词、标签、提及和URL。这些实体要么不提供语义信息，要么仅编码关于讨论主题的信息。然而，所提出的机制假设主题已由用户手动选择（在步骤1中）；因此，所展示的基于四个预训练DNN的自动个体推文描述符

5 https://www.kaggle.com/crowdflower/political-social-media-posts 6 https://textblob.readthedocs.io/en/dev/

only captures complementary semantic information, such as polarity, bias, etc. Additionally, lemmatization was applied to avoid having multiple words with identical meaning. Finally, all words were converted into lower-case. These are typical text preprocessing options in NLP.

### 3.2.3 Neural Models

The DNN architecture that was separately trained for offensiveness, bias and figurativeness recognition [31] consists in a hybrid, parallel BiLSTM-CNN. The input text representations (computed after preprocessing), fed to this neural architecture during both the training and test stages, are derived by using 200-dimensional embedding vectors from a pretrained GloVe model [34]. The CNN applies convolution of kernel sizes 3, 4 and 5, thereby learning fixed length features of 3-grams, 4-grams, and 5-grams, respectively. The convolution is followed by a ReLU activation function. These convolutional features are then downsampled by using a 1-D max pooling function. The CNN outputs are concatenated, combined with the BiLSTM output and jointly fed into a fully connected output neural layer, activated by a sigmoid function to produce the final semantic score: a real number in the range [0, 1].

The pretrained neural model employed for polarity recognition is based on a BiLSTM architecture. A fully-connected embedding layer is used in the front-end of the network, that has learnt to map each input word to a 200-dimensional vector representation. These embeddings are fed to a BiLSTM layer followed by a max-pooling operation. Then, multiple ReLU-activated fully-connected neural layers with dropout are employed to further reduce the dimensionality of the output. The produced dense feature representation is fed to the final sigmoid-activated fully-connected layer that gives us the final real-valued sentiment score for polarity in the range [0, 1].

Table 1 summarizes the achieved recognition accuracy (%) of each of the four opinion classifiers on the test set of the respective training dataset.

**Table 1** Achieved accuracy of each of the four opinion classifiers on the test set of the respective training dataset.

| *Model* | *Accuracy* |
|---|---|
| Bias | 75.64% |
| Figurativeness | 84.45% |
| Polarity | 88.00% |
| Offensiveness | 84.00% |

During the test stage for all four DNN models, each pretrained DNN is actually employed for computing a different part of the individual 4D real-valued tweet descriptor for each incoming tweet. The output semantic score denotes how offensive/biased/figurative/negative the message is judged to be. An output score of 0 means very high possibility of it being non-offensive, non-biased, literal or negative, respectively. An output score of 1 means very high possibility of it being offensive, biased, figurative or positive, respectively.



仅捕获补充语义信息，如极性、偏见等。此外，应用了词形还原以避免具有相同含义的多个词。最后，所有单词都转换为小写。这些是NLP中典型的文本预处理选项。

### 3.2.3 神经模型

分别用于攻击性、偏见和比喻识别而单独训练的DNN架构 [31]由混合、并行BiLSTM-CNN组成。在训练和测试阶段输入到该神经架构的文本表示（在预处理后计算），是通过使用预训练的GloVe模型的200维嵌入向量获得的 [34]。CNN应用3、4和5的卷积核大小，从而分别学习3-gram、4-gram和5-gram的固定长度特征。卷积后是一个ReLU激活函数。然后，这些卷积特征通过使用1-D最大池化函数进行下采样。CNN输出被连接，与BiLSTM输出联合输入到一个由sigmoid函数激活的全连接输出神经层，以产生最终的语义分数：一个实数，范围在 [0, 1]。

用于情感极性识别的预训练神经网络模型基于BiLSTM架构。网络前端使用一个全连接嵌入层，该层已学习将每个输入词映射到200维的向量表示。这些嵌入被输入到BiLSTM层，然后进行最大池化操作。接着，使用多个带有dropout的ReLU激活的全连接神经网络层来进一步降低输出维度。产生的密集特征表示被输入到最终的sigmoid激活的全连接层，该层给出我们范围内的最终实值情感得分 [0, 1]。

Table 1 总结了每个意见分类器在相应训练数据集的测试集上所达到的识别准确率(%)。

**Table 1** 每个意见分类器在相应训练数据集的测试集上所达到的准确率。

| 模型 | 准确率 |
|---|---|
| Bias | 75.64% |
| 比喻性 | 84.45% |
| 极性 | 88.00% |
| 攻击性 ss | 84.00% |

在所有四个DNN模型的测试阶段，每个预训练的DNN实际上用于计算每个传入推文的不同部分的4D实值推文描述符。输出语义分数表示该消息被判断为具有攻击性/偏见/比喻/负面程度。输出分数为0表示其具有非攻击性、非偏见、字面或负面的可能性非常高。输出分数为1表示其具有攻击性、偏见、比喻或正面的可能性非常高。

A score near 0.5 would imply that the tweet is judged to be neutral in the corresponding attribute, or it simply cannot be classified.

### 3.2.4 Hyperparameters

The optimal hyperparameters used for training the DNN classifiers were obtained by manual tuning and are presented in Table 2. The pretrained polarity classifier used the hyperparameters found in the relevant software repository[7].

N_layers denotes the number of hidden layers in the LSTM and N_hidden is the size (nodes) of these layers. Weight_decay denotes the L2 regularization factor. Lr_decay denotes the learning rate multiplying factor. Wd_multiplier denotes the weight decay multiplying factor. Batch_size denotes the number of tweets processed at each step of the optimization. Dropout, dropout_enc and dropout_op denote the dropouts used after the LSTM, Embedding and Output layer respectively.

**Table 2** Hyperparameters used for training the sentiment classifiers.

| *Hyperparameter* | *Values* |
|---|---|
| n_hidden | 200 |
| n_layers | 2 |
| dropout | 0.5 |
| weight_decay | 1e-7 |
| dropout_enc | 0.2 |
| dropout_op | 0.5 |
| lr_decay | 0.7 |
| wd_multiplier | 6 |
| learning_rate | 5e-3 |
| batch_size | 52 |

## 3.3 Step 3: Aggregation

Having obtained an individual 4D, real-valued semantic descriptor per tweet, the next step is to aggregate the derived vectors into the desired public opinion descriptor. To do this, the user has to select the **aggregation strategy** and the **aggregation target**.

The target refers to the granularity of the aggregation process and directly influences the final number of aggregate public opinion description vectors. Two possible choices are the most straightforward:

- *Complete aggregation*. All individual tweet descriptors are merged into a single aggregate public opinion description vector, representing the entire message pool extracted in Step 1.
- *Temporally segmented aggregation*. The overall range of dates out of which the entire message pool was extracted in Step 1 is partitioned in isochronous,

[7]https://github.com/DheerajKumar97/US-2020-Election-Campaign-Youtube-Comments-Sentiment-Analysis

Springer Nature 2021 LaTeX template



接近0.5的分数意味着推文在相应属性中被判断为中立，或者它根本无法被分类。

### 3.2.4 超参数

用于训练DNN分类器的最佳超参数是通过手动调整获得的，并呈现在表 2中。预训练的情感分类器使用了在相关软件仓库中找到的超参数7。

N layers表示LSTM中的隐藏层数量，N hidden是这些层的大小（节点）。Weight decay表示L2正则化因子。Lr decay表示学习率乘数因子。Wd multiplier表示权重衰减乘数因子。Batch size表示每次优化步骤中处理的推文数量。Dropout、dropout enc和dropout op分别表示LSTM、Embedding和Output层之后的dropout。

**表 2** 用于训练情感分类器的超参数。

| *Hyperparameter* | 值 |
|---|---|
| n 隐藏 | 200 |
| n 层数 | 2 |
| dropout | 0.5 |
| 权重衰减 | 1e-7 |
| dropout 编码 | 0.2 |
| dropout op | 0.5 |
| lr decay | 0.7 |
| wd multiplier | 6 |
| learning rate | 5e-3 |
| batch size | 52 |

## 3.3 Step 3: Aggregation

获取每条推文的个体4D实值语义描述符后，下一步是将派生出的向量聚合为所需的公众意见描述符。为此，用户必须选择 **聚合策略** 和 **聚合目标**。

目标指的是聚合过程的粒度，并直接影响最终聚合的公众意见描述向量的数量。两种可能的选择是最直接的：

- 完成聚合。所有单个推文描述符被合并成一个单一的聚合公众意见描述向量，代表在步骤1中提取的整个信息池。
- 时间分段聚合。日期的整体范围 在步骤1中提取的整个消息池在步骤3中被分区

通过推文的集体语义分析进行舆论监测3.3 第 3 步：聚合

non-overlapping and consecutive time periods. All individual tweet descriptors falling under each period are merged into a single aggregate public opinion description vector. This is separately performed for each period.

With complete aggregation, the outcome is a single 4D vector. With temporally segmented aggregation, the outcome is a 4D timeseries. Examples of temporally segmented aggregation targets would be day-by-day or week-by-week aggregation. Depending on the application, different additional aggregation targets may also be envisioned.

The aggregation strategy refers to how a set of individual 4D tweet descriptors are combined into a single aggregate 4D descriptor. Three possible choices are the most straightforward:

- *Element-wise vector mean.*
- *Element-wise vector median.*
- *Element-wise vector trimmed mean.*

All three of these choices may be implemented simply by performing computations separately along each of the four descriptor dimensions. As before, different additional aggregation strategies may also be envisioned, depending on the application.

## 4 Evaluation

The proposed mechanism was evaluated on the well-known 2016 and 2020 United States Presidential Election Tweets datasets, using a day-by-day temporally segmented aggregation target. All three aggregation strategies described in Section 3 were separately followed and assessed. Details and results follow in the next Subsections.

### 4.1 Datasets

The 2016 US Presidential Election tweet dataset from Kaggle[8] contains 61 million rows. Their overall time range is from 2016-08-30 to 2017-02-28, with 20 days missing, leading to a total of 163 days. From this initial dataset, we retained approximately 32 million tweets after applying a common cleaning process: removal of empty rows, of non-English text, of duplicate tweets and of messages that contained less than 5 words after text preprocessing. This is important for proper semantic analysis, since an adequate number of words per tweet is essential to achieving high opinion mining accuracy. Subsequently, the keywords "Clinton", "Obama" and "Trump" were exploited for partitioning the messages into ones referring to Democrats and ones referring to Republicans.

Overall, this entire manual process was equivalent to performing Step 1 of the proposed mechanism. It was not needed in its entirety for the smaller second dataset that we employed, i.e., the US Election 2020 Tweets from

[8] https://www.kaggle.com/paulrohan2020/2016-usa-presidential-election-tweets61m-rows



非重叠且连续的时间段。每个时间段内的所有单个推文描述符都被合并成一个单一的聚合舆论描述向量。这针对每个时间段分别进行。

完全聚合后，结果是单个4D向量。时间分段聚合后，结果是4D时间序列。时间分段聚合的目标示例可以是按日或按周聚合。根据应用，还可能设想不同的其他聚合目标。

聚合策略指的是如何将一组单个4D推文描述符组合成一个单一的聚合4D描述符。三种可能的选择是最直接的：

- 元素向量均值.
- 逐元素向量中位数.
- 逐元素向量截尾均值.

这三个选项都可以简单地通过在每个四个描述符维度上分别进行计算来实现。与之前一样，也可以根据应用场景设想不同的附加聚合策略。

## 4 评估

所提出的机制在著名的 2016 年和 2020 年美国总统选举推文数据集上进行了评估，使用按日时间分割的聚合目标。第 3 节中描述的三个聚合策略被分别遵循和评估。详细信息和结果将在下一小节中提供。

### 4.1 数据集

Kaggle上的2016年美国总统选举推文数据集[8] 包含6100万行。其总体时间范围是从2016-08-30到2017-02-28，有20天缺失，导致总共有163天。从这个初始数据集中，在应用常见的清理过程后，我们保留了大约3200万条推文：删除空行、非英文文本、重复推文以及文本预处理后包含少于5个词的消息。这对于正确的语义分析很重要，因为每条推文中足够的词数对于实现高意见挖掘精度至关重要。随后，利用关键词“克林顿”、“奥巴马”和“特朗普”将消息分为指代民主党人和共和党的消息。

总的来说，这个整个手动过程相当于执行所提出机制的第1步。对于我们使用的小型第二个数据集（即美国2020年选举推文）来说，并不需要完全执行这一步。

8https://www.kaggle.com/paulrohan2020/2016-usa-presidential-election-tweets61m-rows

Kaggle[9], since its tweets are preseparated in two partisan groups (Democrats and Republicans). It contains 1.72 million rows in total, with an overall time range from 2020-10-15 to 2020-11-08, meaning 25 days in total. From this initial dataset, approximately 720 thousand tweets were kept after applying the cleaning process described above.

Subsequently, the proposed mechanism was separately applied to the two message pools of each dataset. As a result, two day-by-day 4D descriptor timeseries were derived, covering overall the exact same time range: one for the Republicans and one for the Democrats (separately for each dataset). Indicatively, for the 2016 US Presidential Election tweet dataset, generating the descriptors for all relevant tweets required 24 hours. Day-by-day aggregation required 10 minutes for each aggregation strategy. Experiments were performed on a desktop computer with an AMD Ryzen 5@3.2GHz CPU, 16GB of DDR4 RAM and an nVidia GeForce GTX1060 (6GB RAM) GP-GPU.

## 4.2 Analysis 1: Timeseries Forecasting

The first type of evaluation performed on the derived timeseries was to assess their predictability using AI-enabled forecasting. Since the two constructed timeseries compactly capture public opinion about the two respective parties during a heated pre/post-election period, forecasting has obvious political usefulness: it may allow an interested organization to predict near-future changes in its public image, using only Twitter data. Of course, in this context, "near-future" implies a forecasting horizon of a few weeks at the most.

### 4.2.1 Implementation

The two timeseries were day-by-day 4D descriptions of evolving public opinion about the Democrats and the Republicans, respectively. However, since three different aggregation strategies were employed (mean, median, trimmed mean), in fact six 4D timeseries were derived overall, with all of them covering the same period. Since forecasting is typically performed on univariate timeseries, each descriptor channel was then handled separately, leading to a total of 24 different timeseries.

A moderate 7-day forecasting horizon was selected, since this allows for rather reliable predictions while still being practically useful. A recent stacked LSTM architecture was adopted [35], with each LSTM cell being followed by a fully connected neural layer. The overall DNN was trained separately for each timeseries, using Back-Propagation Through Time (BPTT) and a Continuous Coin Betting (COCOB) optimizer [36]. This training approach was selected over alternative optimizers (Adagrad and Adam) because it displayed superior performance in [35]. The fact that COCOB attempts to minimize the loss function by self-tuning its learning rate, accelerates convergence in comparison to other gradient descent-based algorithms with a constant or decaying learning rate, where the convergence becomes slower close to the optimum.

[9] https://www.kaggle.com/datasets/manchunhui/us-election-2020-tweets



Kaggle9, 由于其推文被预先分为两个党派团体（民主党人和共和党人）。它总共包含 1.72 百万行，总体时间范围从 2020-10-15 到 2020-11-08，意味着总共 25 天。从这个初始数据集开始，在应用上述清理过程后，保留了大约 720 千条推文。

随后，所提出的机制被分别应用于每个数据集的两个消息池。结果，生成了两个按天划分的 4D 描述符时间序列，覆盖了完全相同的时间范围：一个用于共和党人，一个用于民主党人（在每个数据集中分别）。指示地，对于 2016 年美国总统选举推文数据集，为所有相关推文生成描述符需要 24 小时。按天聚合需要每聚合策略 10 分钟。实验是在一台配备 AMD Ryzen 5@3.2GHz CPU、16GB DDR4 RAM 和 nVidia GeForce GTX1060（6GB RAM）GP-GPU 的台式计算机上进行的。

## 4.2 分析 1：时间序列预测

对派生出的时间序列进行的首次评估是使用人工智能驱动的预测来评估其可预测性。由于这两个构建的时间序列紧凑地捕捉了在激烈的前/后选举期间关于两个相关政党的公众意见，因此预测具有明显的政治用途：它可能允许一个感兴趣的机构仅使用 Twitter 数据预测其公众形象的未来变化。当然，在这种情况下，“未来”意味着最多几周的预测范围。

### 4.2.1 实施

这两个时间序列分别是关于民主党人和共和党人不断演变的公众意见的每日 4D 描述。然而，由于采用了三种不同的聚合策略（均值、中位数、截尾均值），实际上派生了六个 4D 时间序列，它们都涵盖了相同的时期。由于预测通常在单变量时间序列上执行，因此每个描述通道随后被单独处理，导致总共产生了 24 个不同的时间序列。

选择了一个适中的7天预测范围，因为这样可以进行相当可靠的预测，同时仍然具有实际用途。采用了一种最近的堆叠LSTM架构 [35]，，每个LST M单元后面都跟着一个全连接神经网络层。整体DNN分别针对每个时间序列进行训练，使用反向传播通过时间（BPTT）和连续硬币投注（COCOB）优化器 [36]。选择这种训练方法而不是其他优化器（Adagrad和Adam），是因为它在 [35]中表现出优异的性能。COCOB通过自调学习率来尝试最小化损失函数，与其他具有恒定或衰减学习率的基于梯度下降的算法相比，它加速了收敛，而在接近最优值时，这些算法的收敛速度会变慢。

9 https://www.kaggle.com/datasets/manchunhui/us-election-2020-tweets

The use of BPTT for updating model parameters during training is necessary when employing LSTMs.

Optimal hyperparameters for training the forecasting DNN model were obtained by using Sequential Model-based Algorithm Configuration (SMAC) [37] and are presented in Table 3. minibatch_size denotes the number of timeseries considered for each full backpropagation in the LSTM. Epoch_size denotes how many times the dataset is traversed within each epoch. L2_regularization and Gaussian noise added to the input are used to reduce overfitting. LSTM unit weights were initialized using a random_normal_initializer.

**Table 3** Hyperparameters used for training the forecasting model.

| *Hyperparameter* | *Value* |
|---|---|
| cell_dimension | 20 |
| gaussian_noise_stdev | 1e-4 |
| l2_regularization | 1e-4 |
| max_epoch_size | 1 |
| max_num_epochs | 2 |
| minibatch_size | 4 |
| num_hidden_layers | 1 |
| random_normal_initializer_stdev | 1e-4 |

Out of the two evaluation datasets, only the 2016 one was used for training the forecasting model. A segment was withheld for test purposes from the end of each timeseries, with a length equal to the forecasting horizon; the remaining data constituted the training dataset. Moreover, this pretrained model was also separately tested on the 2020 dataset. The results from testing on both datasets are presented in the sequel.

Deseasonalisation was applied as a common preprocessing step, since DNNs are weak at modelling seasonality [38]. That was achieved by decomposing each timeseries into seasonal, trend, and remainder components, in order to subsequently remove the seasonality component, by employing STL decomposition. If a timeseries exhibited no seasonality, this step simply returned zero seasonality.

Sliding window schemes were adopted for feeding inputs to the DNN and deriving the outputs, with the output window size $n$ set to be equal to the size of the forecasting horizon $H = 7$. The input window size $m$ was empirically set to $9 = n1.25$. Each training timeseries was broken down into blocks of size $m + n$, thus forming the input–output pairs for each LSTM cell instance.

### 4.2.2 Metrics and Results

A set of common timeseries forecasting evaluation quantitative metrics were employed for assessing the predictability of the computed timeseries.

First, the *Symmetric Mean Absolute Percentage Error* (SMAPE) is defined as follows:



在训练期间使用BPTT更新模型参数时，当采用LSTMs时这是必要的。

通过基于序列模型的算法配置（SMAC）获得用于训练预测DNN模型的最佳超参数[37] 并且以表格3的形式呈现。minibatch size表示每次完整反向传播时考虑的时间序列的次数。Epoch size表示每个epoch内遍历数据集的次数。L2正则化和添加到输入的高斯噪声用于减少过拟合。LSTM单元权重使用随机正态初始化器进行初始化。

**表3** 用于训练预测模型的超参数。

| 超参数 | 值 |
|---|---|
| 单元格维度 | 20 |
| 高斯噪声标准差 | 1e-4 |
| L2正则化 | 1e-4 |
| 最大 epoch 大小 | 1 |
| 最大 epoch 数 | 2 |
| 小批量大小 | 4 |
| 隐藏层数量 | 1 |
| 随机正态初始化标准差 | 1e-4 |

在两个评估数据集中，仅使用了2016年的数据来训练预测模型。从每个时间序列的末尾保留了一段用于测试，其长度等于预测范围；剩余数据构成了训练数据集。此外，这个预训练模型也在2020数据集上单独进行了测试。在两个数据集上的测试结果在后续部分呈现。

去季节化作为常见的预处理步骤被应用，因为深度神经网络在建模季节性方面较弱 [38]。这是通过将每个时间序列分解为季节性、趋势和剩余分量来实现的，以便随后通过使用STL分解去除季节性分量。如果一个时间序列没有季节性，这个步骤简单地返回零季节性。

滑动窗口方案被采用用于向深度神经网络输入数据并导出输出，输出窗口大小 $n$ 设置为等于预测范围 $H = 7$。输入窗口大小 $m$ 经经验性设置为 $9 = n1.25$。每个训练时间序列被分解为大小为 $m + n$ 的块，从而形成每个LSTM单元实例的输入-输出对。

### 4.2.2 指标与结果

采用一组常见的时序预测评估定量指标来评估计算时序的可预测性。

首先，对称平均绝对百分比误差 (SMAPE) 定义如下：

**Table 4** Forecasting results on the US 2016 Presidential Election Tweets dataset for the six constructed timeseries. "Dem" denotes the Democrats, "Rep" denotes the Republicans, while "mean", "med" and "trim" imply the three respective aggregation strategies: mean, median and trimmed mean. In each case, the SMAPE/MASE metrics have been independently averaged across the four descriptor channels using both the mean and the median operator. A lower value is better for both metrics, while SMAPE is a percentage.

| *Timeseries* | *Mean SMAPE* | *Median SMAPE* | *Mean MASE* | *Median MASE* |
|---|---|---|---|---|
| Dem-Mean | 0.1096 | 0.0563 | 1.2396 | 1.0799 |
| Dem-Med | 0.1380 | 0.0913 | 1.0620 | 1.1711 |
| Dem-Trim | 0.1798 | 0.0676 | 1.3364 | 1.0885 |
| Rep-Mean | 0.0529 | 0.0280 | 0.7689 | 0.6937 |
| Rep-Med | 0.0492 | 0.0330 | 0.5158 | 0.5303 |
| Rep-Trim | 0.0737 | 0.0314 | 0.6931 | 0.6549 |

**Table 5** Forecasting results on the US 2020 Presidential Election Tweets dataset for the six constructed timeseries. "Dem" denotes the Democrats, "Rep" denotes the Republicans, while "mean", "med" and "trim" imply the three respective aggregation strategies: mean, median and trimmed mean. In each case, the SMAPE/MASE metrics have been independently averaged across the four descriptor channels using both the mean and the median operator. A lower value is better for both metrics, while SMAPE is a percentage.

| *Timeseries* | *Mean SMAPE* | *Median SMAPE* | *Mean MASE* | *Median MASE* |
|---|---|---|---|---|
| Dem-Mean | 0.1793 | 0.1490 | 1.6975 | 1.6943 |
| Dem-Med | 0.3431 | 0.2088 | 1.7526 | 1.6620 |
| Dem-Trim | 0.2847 | 0.1903 | 1.7535 | 1.7434 |
| Rep-Mean | 0.0961 | 0.0813 | 1.5233 | 1.4572 |
| Rep-Med | 0.1867 | 0.1303 | 1.7228 | 1.6949 |
| Rep-Trim | 0.1472 | 0.0999 | 1.5961 | 1.5637 |

$$SMAPE = \frac{100\%}{H} \sum_{k=1}^{H} \frac{|F_k - Y_k|}{(|Y_k| + |F_k|)/2}, \tag{1}$$

where $H$, $F_k$, and $Y_k$ indicate the size of the horizon, the forecast of the DNN and the ground-truth forecast, respectively.

Due to the low interpretability and high skewness of SMAPE [39], the scale-independent *Mean Absolute Scaled Error* (MASE) metric was also employed. For non-seasonal timeseries, it is defined as follows:

$$MASE = \frac{\frac{1}{H}\sum_{k=1}^{H}|F_k - Y_k|}{\frac{1}{T-1}\sum_{k=2}^{T}|Y_k - Y_{k-1}|}, \tag{2}$$

In Eq. (2), the numerator is the same as in SMAPE, but normalised by the average in-sample one-step naive forecast error. A MASE value greater than 1 indicates that the performance of the tested model is worse on average than the naive benchmark, while a value less than 1 denotes the opposite. Therefore, this error metric provides a direct indication of forecasting accuracy relatively to the naive benchmark.

Since these metrics are computed for univariate timeseries forecasting, we employed mean and median aggregation across the four descriptor channels for each of the six timeseries. The results obtained for the 2016 and 2020 datasets

Springer Nature 2021 LaTeX template

*Public opinion monitoring through collective semantic analysis of tweets* 15

**表4** 美国2016年总统选举推文数据集上六个构建的时间序列的预测结果。"Dem"表示民主党，"Rep"表示共和党，而"mean"、"med"和"trim"分别表示三种相应的聚合策略：均值、中位数和截尾均值。在每种情况下，SMAPE/MASE指标均已独立地使用均值和中位数运算符在四个描述通道上进行了平均。对于这两个指标，较低值更好，而SMAPE是一个百分比。

| 时间序列 | 均值*SMAPE* | 中位数*SMAPE* | 均值*MASE* | 中位数*MASE* |
|---|---|---|---|---|
| Dem-均值 | 0.1096 | 0.0563 | 1.2396 | 1.0799 |
| Dem-Med | 0.1380 | 0.0913 | 1.0620 | 1.1711 |
| Dem-Trim | 0.1798 | 0.0676 | 1.3364 | 1.0885 |
| Rep-Mean | 0.0529 | 0.0280 | 0.7689 | 0.6937 |
| Rep-Med | 0.0492 | 0.0330 | 0.5158 | 0.5303 |
| Rep-Trim | 0.0737 | 0.0314 | 0.6931 | 0.6549 |

**Table 5** 美国2020年总统选举推文数据集上六种构建的时间序列的预测结果。"Dem"表示民主党，"Rep"表示共和党，而"mean"、"med"和"trim"分别表示三种相应的聚合策略：均值、中位数和修剪均值。在每种情况下，SMAPE/MASE指标均已独立地使用均值和中位数运算符在四个描述通道上进行平均。对于这两个指标，较低值更好，而SMAPE是一个百分比。

| 时间序列 | 平均 *SMAPE* | 中位数 *SMAPE* | 平均 *MASE* | 中位数 *MASE* |
|---|---|---|---|---|
| Dem-Mean | 0.1793 | 0.1490 | 1.6975 | 1.6943 |
| Dem-Med | 0.3431 | 0.2088 | 1.7526 | 1.6620 |
| Dem-Trim | 0.2847 | 0.1903 | 1.7535 | 1.7434 |
| Rep-Mean | 0.0961 | 0.0813 | 1.5233 | 1.4572 |
| Rep-Med | 0.1867 | 0.1303 | 1.7228 | 1.6949 |
| Rep-Trim | 0.1472 | 0.0999 | 1.5961 | 1.5637 |

$$SMAPE = \frac{100\%}{H} \sum_{k=1}^{H} \frac{|F_k - Y_k|}{(|Y_k| + |F_k|)/2}, \tag{1}$$

其中 $H$、$F_k$ 和 $Y_k$ 分别表示视野范围、DNN 的预测和真实预测的大小。

由于SMAPE的低可解释性和高偏斜性 [39], 尺度无关的平均绝对尺度误差(MASE)指标也被采用。对于非季节性时间序列，它定义如下：

$$MASE = \frac{\frac{1}{H}\sum_{k=1}^{H}|F_k - Y_k|}{\frac{1}{T-1}\sum_{k=2}^{T}|Y_k - Y_{k-1}|}, \tag{2}$$

在公式（2）中，分子与SMAPE相同，但被样本内一步朴素预测误差的平均值标准化。MASE值大于1表示测试模型的平均性能比朴素基准更差，而值小于1则表示相反。因此，这个误差指标提供了相对于朴素基准的直接预测准确度指示。

由于这些指标是针对单变量时间序列预测计算的，我们针对六个时间序列中的每个时间序列，在四个描述符通道上使用均值和中位数聚合。2016年和2020数据集获得的结果

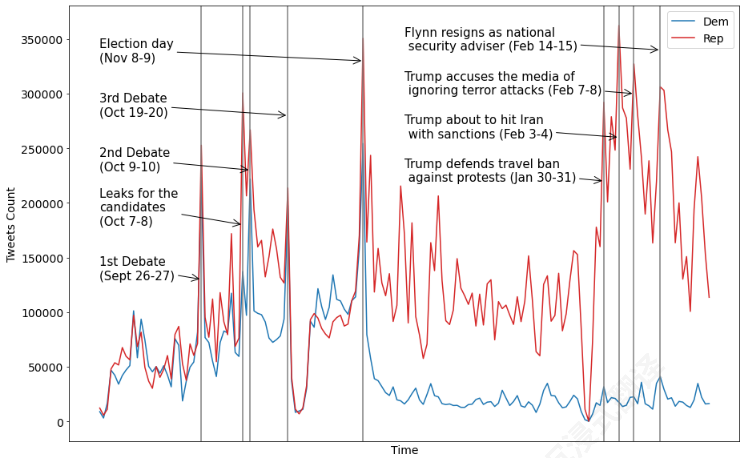


**Fig. 2** Daily number of tweets for Democrats (Dems) and Republicans (Reps) in 2016 dataset. The two dates given per event are the date of that event (first) and the date of the respective reaction in Twitter (second).

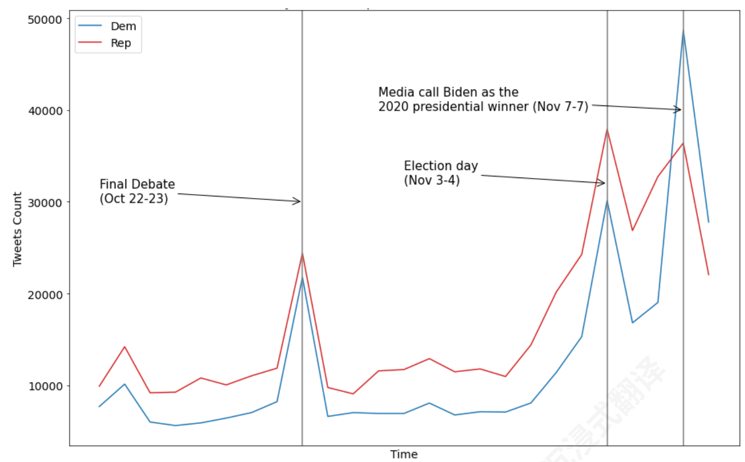


**Fig. 3** Daily number of tweets for Democrats (Dems) and Republicans (Reps) in the 2020 dataset. The two dates given per event are the date of that event (first) and the date of the respective reaction in Twitter (second).

are shown in Tables 4 and 5, respectively, where larger SMAPE or MASE values indicate worse forecasting accuracy.

In general, the timeseries constructed using the proposed mechanism seem to be predictable to an acceptable degree by using the employed DNN model. Moreover, forecasting behaves similarly for both datasets, leading us to draw a common set of conclusions. First, the mean aggregation strategy resulted in the timeseries with the best overall forecasting behaviour. Second, based on both metrics, it is clear that forecasting performs worse for the Democrats than for the Republicans, implying that public opinion concerning them (as



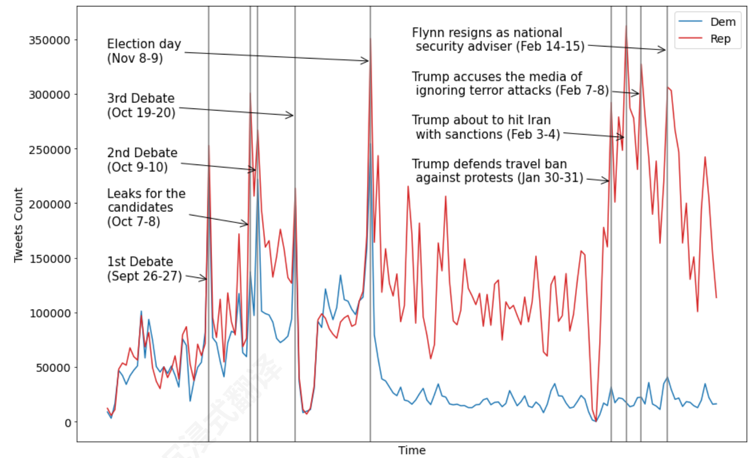


**图2** 2016数据集中民主党（Dems）和共和党（Reps）的每日推文数量。所给的两个日期分别代表事件发生日期（第一个）和Twitter中的相应反应日期（第二个）。

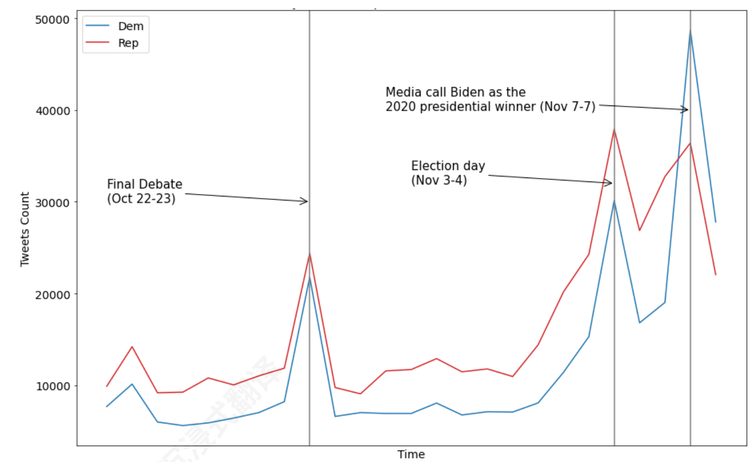


**图3** 2020数据集中民主党（Dems）和共和党（Reps）的每日推文数量。所给的两个日期分别代表事件发生日期（第一个）和Twitter中的相应反应日期（第二个）。

分别显示在表 4 和 5中，其中较大的SMAPE或MASE值表示更差的预测精度。

一般来说，使用所提出的机制构建的时间序列似乎可以通过所使用的DNN模型以可接受的程度进行预测。此外，两个数据集的预测行为相似，导致我们得出了一套共同的结论。首先，均值聚合策略产生了具有最佳整体预测行为的时间序列。其次，根据这两个指标，很明显民主党（Dems）的预测表现不如共和党（Reps），这意味着关于他们的舆论（作为

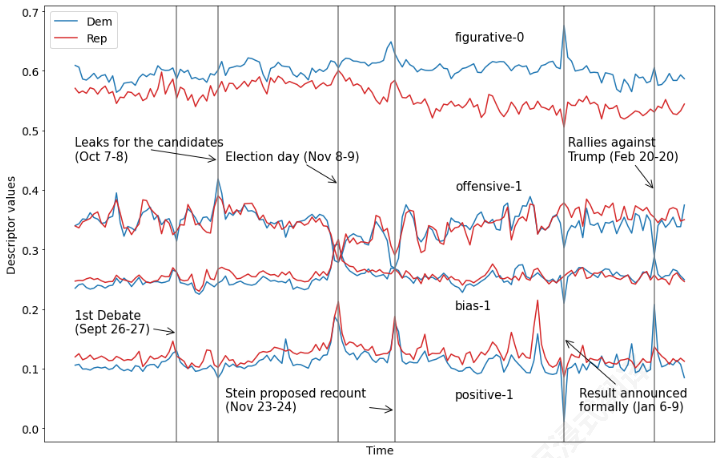


**Fig. 4** Per-channel day-by-day values of the 4D timeseries constructed from the 2016 data using the proposed descriptor and the mean aggregation strategy, separately for Democrats (Dems) and Republicans (Reps). The two dates given per event are the date of that event (first) and the date of the respective reaction in Twitter (second).

expressed in Twitter) was less stable and predictable during the examined period. Finally, we notice a drop in accuracy when testing on the 2020 dataset (small in absolute terms), compared to the 2016 one. This is to be expected, since the forecasting DNN was pretrained only on the training set of the 2016 dataset.

## 4.3 Analysis 2: Visualizations and Qualitative Evaluation

A set of visualizations were computed from the 4D timeseries constructed using the proposed mechanism, in order to facilitate manual inspection of the outcome. Given the conclusions of Subsection 4.2, only the timeseries derived by mean aggregation were exploited here. This Subsection presents this qualitative evaluation process and its results, along with auxiliary information about the original 2016/2020 US Presidential Elections datasets.

First, Figure 2 depicts the number of tweets posted every day in the complete dataset's time range (from 2016-08-30 to 2017-02-28), separately for the Democrats and the Republicans. For the most important events like the three presidential debates and the election day, increased Twitter traffic is observed for both parties. Leaks for both Clinton and Trump that took place in 2016-10-07 seem to have affected more the latter candidate, as the majority of posts expressed an opinion about him. Equal traffic is observed for both parties just before the election day (2016-11-08), since in that stage Twitter plays a significant role in the campaign of both candidates. However, the number of tweets concerning Democrats drops significantly after the election day and their defeat. In contrast, people kept tweeting frequently about the winner and center of attention Donald Trump regarding his actions as a president of the US, like the travel ban, Iran sanctions, etc.



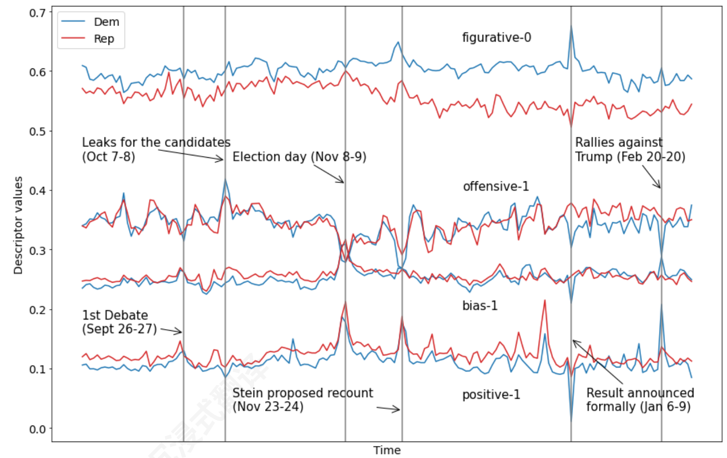


**图4** 使用所提出的描述符和均值聚合策略，分别针对民主党人（Dems）和共和党人（Reps），从2016年数据中构建的4D时间序列的逐通道逐日值。每个事件给出的两个日期是该事件（第一个）和Twitter中的相应反应（第二个）的日期。

在Twitter中表达的）在考察期间不太稳定和可预测。最后，我们注意到在测试2020数据集时（绝对值小），准确率有所下降，与2016年相比。这是可以预料的，因为预测DNN仅在2016数据集的训练集上进行了预训练。

## 4.3 分析2：可视化和定性评估

使用所提出的机制从构建的4D时间序列计算了一组可视化，以促进对结果的手动检查。鉴于4.2节的结论，这里仅利用了均值聚合得到的时间序列。本节介绍此定性评估过程及其结果，以及有关原始2016/2020美国总统选举数据集的辅助信息。

首先，图 2 描绘了完整数据集时间范围内（从2016-08-30到2017-02-28）每天发布的推文数量，分别针对民主党人和共和党人。在最重要的活动，如三场总统辩论和选举日，两党的推文流量都增加了。2016-10-07发生的克林顿和特朗普的泄密事件似乎对后者候选人影响更大，因为大多数帖子表达了对他的意见。选举日前（2016-11-08），两党流量相等，因为在那个阶段推特在两位候选人的竞选活动中发挥了重要作用。然而，选举日后民主党相关的推文数量显著下降，他们也被击败了。相比之下，人们频繁地发帖谈论赢家和焦点人物唐纳德·特朗普，关于他作为美国总统的行动，如旅行禁令、伊朗制裁等。

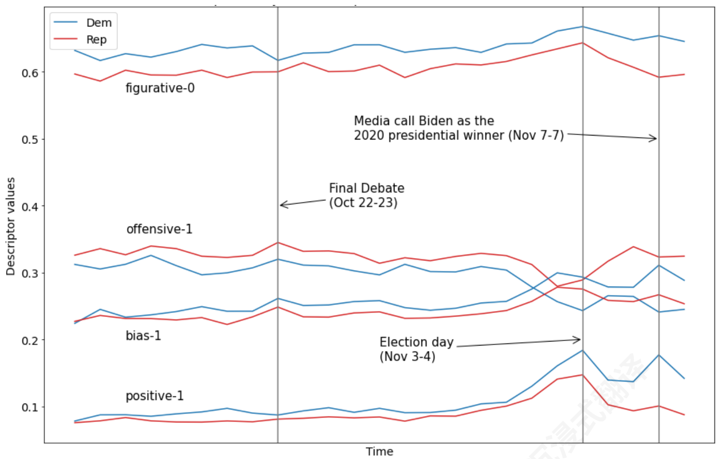


**Fig. 5** Per-channel day-by-day values of the 4D timeseries constructed from the 2020 data using the proposed descriptor and the mean aggregation strategy, separately for Democrats (Dems) and Republicans (Reps). The two dates given per event are the date of that event (first) and the date of the respective reaction in Twitter (second).

Figure 3 depicts the number of tweets posted from 2020-10-15 to 2020-11-08, separately for the Democrats and the Republicans. Again, increased Twitter traffic is observed for both parties during the most important events. However less events are observed in the 2020 plot, which is due to the smaller size of the dataset as a whole, in comparison compared to the 2016 dataset. In general, there were more tweets posted about Trump up until November 7. On that day, the media called Biden as the 2020 presidential winner and Twitter traffic exploded for Biden, surpassing Trump posts by a significant margin. This obviously makes sense as Biden's victory is officially announced for the first time.

Having established original Twitter traffic patterns, concurrent daily values of the 4D timeseries constructed using the proposed descriptor and the mean aggregation strategy are depicted in Figure 4, separately for each party of the 2016 US election (party affiliation is color-coded). In this Figure, as in many following ones, we label each timeseries by one of the two opposite class labels (e.g., "positive" or "figurative") followed by the tweet classifier output value which implies this label (e.g., a tweet fully and undoubtedly classified as figurative/positive, has been assigned a value of 0/1 by the figurativeness/polarity classifier, respectively). It is evident that the timeseries maintain a stable class along all four dimensions and across the entire time range for both parties: their class is negative (in polarity), unbiased, non-offensive and literal. This indicates a general public stance towards the competing politicians, which reflects a judgemental and indignant (negative + literal) but simultaneously educated (unbiased + non-offensive) public.

Figure 5 depicts the mean daily 4D descriptor for each party of the 2020 US election. Again, the timeseries maintain the same stable classes with the



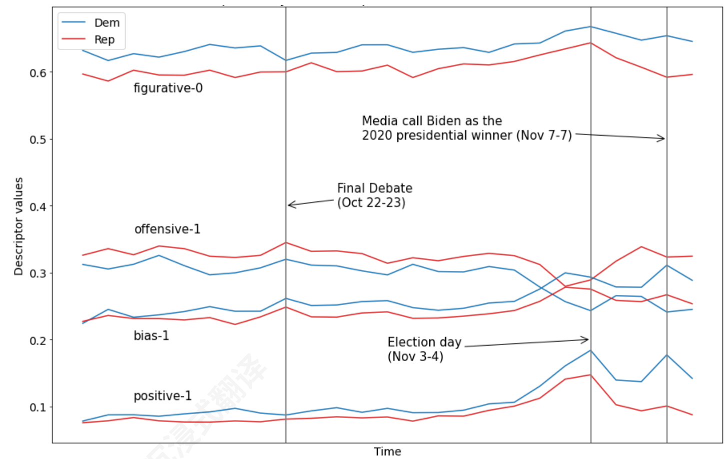


**图5** 使用所提出的描述符和均值聚合策略从2020年数据中构建的4D时间序列的逐通道逐日值，分别针对民主党（Dems）和共和党（Reps）。每个事件给出的两个日期是该事件（第一个）和Twitter中的相应反应日期（第二个）。

图 3 描绘了从2020-10-15到2020-11-08期间发布的推文数量，分别针对民主党人和共和党人。同样，在最重要的活动中，两党的Twitter流量都增加了。然而，2020年的图中观察到的事件较少，这是由于整个数据集较小，与2016年的数据集相比。总的来说，直到11月7日之前，关于特朗普的推文更多。在那一天，媒体宣布拜登为2020年总统选举的获胜者，Biden的Twitter流量激增，大幅超过特朗普的推文数量。这显然是有道理的，因为拜登的胜利是第一次正式宣布。

在建立了原始的推特流量模式后，使用所提出的描述符和均值聚合策略构建的4D时间序列的并发每日值分别描绘在图 4中，分别针对2016年美国大选的每个政党（党派归属用颜色编码）。在这个图中，如在许多后续的图中一样，我们用两个相反的类别标签（例如，“正面”或“比喻”）中的一个来标记每个时间序列，然后是推文分类器输出值，该值表示此标签（例如，一个完全且无疑被分类为比喻/正面的推文，分别被比喻性/情感分类器分配了0/1的值）。很明显，时间序列在所有四个维度和整个时间范围内都保持稳定的类别：它们的类别是负面的（在情感上），无偏的，非攻击性的和字面的。这表明了公众对竞争对手政治家的总体立场，这反映了一种评判性和愤慨（负面 + 字面的）但同时也受过教育（无偏 + 非攻击性的）的公众。

图 5 展示了2020年美国大选每个政党的平均每日4D描述符。同样，时间序列保持了相同的稳定类别，与

2016 data along all four dimensions and across the entire time range for both parties. An interesting conclusion that can be drawn by comparing the plots of the 2016 and the 2020 elections is that the public seems to have a rather fixed stance towards the competing parties, carried over from one election period to the next one, with the winner determined by a small margin/difference.

Moreover, by visually inspecting Figures 2, 4 and 3, 5 for the 2016 and 2020 datasets, respectively, a correlation can be identified between the occurrence of crucial events and abrupt changes (spikes) in the number of tweets or public opinion. As expected, this reaction in Twitter takes place the day after the event.

By comparing the timeseries of the two parties in figure 4, one can observe that tweets about Republicans are less negative, less unbiased and less literal, while there is no clearly distinguishable difference between the two parties concerning offensiveness. These observations shed new light to the election results of November 8th. A less negative opinion is clearly an advantage in itself for Republicans, but combining it with a more biased opinion reflects the possibility that there were more Trump's partisans active in Twitter. Given that partisans are decided voters that do not easily change their opinion, these conclusions drawn from analysing the constructed timeseries paint the picture of a significant Republican advantage, by only using public Twitter data.

Interestingly, posts about Trump appear to be less literal. Despite common perceptions that figurative language is most often used to express negative opinions, tweets about Republicans are on average less negative. A possible explanation is that figurativeness doesn't reflect on the voters' decision and is mainly being used by Twitter users to attract higher attention. Therefore, our analysis indicates that if a voter is clearly against a candidate, it is more probable for them to be straightforward in their comments.

Corresponding conclusions can be drawn from Figure 5 regarding the 2020 US elections. One can observe that the tweets about Democrats are less negative, less unbiased, more literal and less offensive. The less negative and less unbiased attributes can be interpreted as beneficial factors for the Democratic party (like in 2016). The main difference in the 2020 data are the less offensive and more literal tweets that seemingly further contribute to Democratic dominance, as figurative language usually implies negativity [40].

Principal Component Analysis (PCA) was exploited for applying dimensionality reduction to the 4D mean Republican/Democrat timeseries, so that they can be visualized in 2D plots. The 2D descriptor points per party are presented in Figures 6, 7 for the 2016 dataset and 8, 9 for the 2020 dataset. Here, outlying data points correspond to the spikes of the original time-domain plots of Figures 2, 4 and 3, 5. Thus, outliers in PCA Figures indicate the occurrence of crucial events. This visualization can help us identify incidents that significantly affect public opinion, but does not immediately provide us with corresponding information about the semantic descriptor values.

Springer Nature 2021 LaTeX template

通过推文的集体语义分析进行舆论监测 19

2016年数据在所有四个维度上跨越整个时间范围，涵盖双方。通过比较2016年和2020年选举的图表，可以得出一个有趣的结论：公众对竞争党派似乎持有相当固定的立场，这种立场从一次选举周期延续到下一次，而获胜者则由微小的差距/差异决定。

此外，通过直观检查2016年和2020年数据集的图 2， 4 和 3， 5 ，可以识别出关键事件的发生与推文数量或公众舆论的急剧变化（峰值）之间的相关性。正如预期的那样，这种反应发生在事件发生后的第二天。

通过比较图 4中的两党时间序列，可以观察到关于共和党的推文负面性较低、偏见较少且更直白，而两党在攻击性方面没有明显可区分的差异。这些观察为11月8日的选举结果提供了新的见解。负面意见较少本身对共和党就是一个优势，但结合更偏见的意见反映了特朗普的支持者在Twitter上可能更为活跃。鉴于支持者决定投票且不易改变意见，从分析构建的时间序列中得出的结论表明，仅使用公共Twitter数据，共和党具有显著优势。

有趣的是，关于特朗普的帖子似乎不那么直白。尽管普遍认为比喻性语言通常用来表达负面观点，但关于共和党的推文平均而言不那么负面。一个可能的解释是，比喻性语言并不反映选民的决定，而主要是被推特用户用来吸引更多关注。因此，我们的分析表明，如果一个选民明确反对某个候选人，他们更有可能直言不讳地发表评论。

可以从图 5 关于2020年美国大选得出相应的结论。可以观察到，关于民主党的推文不那么负面、不那么有偏见、更直白、不那么冒犯。不那么负面和不那么有偏见的属性可以被解释为对民主党有利因素（就像2016年一样）。2020年数据的主要区别是更不冒犯和更直白的推文，这似乎进一步有助于民主党的主导地位，因为比喻性语言通常暗示负面 [40]。

主成分分析（PCA）被用于对4D平均共和党/民主党时间序列进行降维，以便在2D图中进行可视化。每个政党的2D描述符点在图 6、 7 中呈现，用于2016数据集， 8、 9中呈现，用于2020数据集。在这里，异常数据点对应于图 2 、 4 和 3 、 5 中原始时域图的尖峰。因此，PCA 图中的异常点表明关键事件的发生。这种可视化可以帮助我们识别显著影响公众意见的事件，但不会立即提供相应的语义描述符值信息。

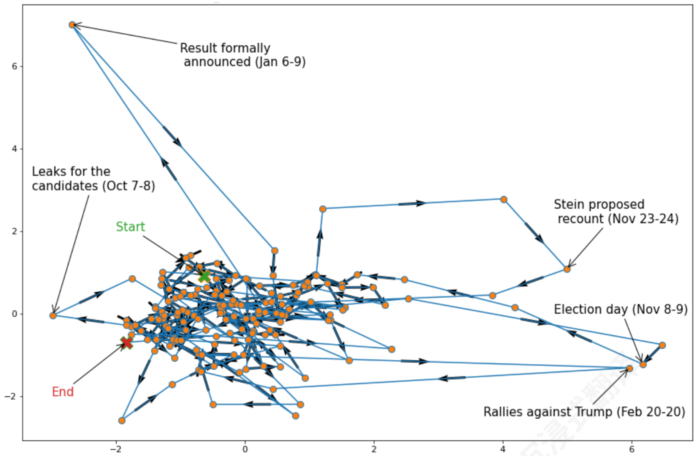


**Fig. 6** PCA-based 2D visualization of the constructed 4D timeseries for the Democrats, using a mean aggregation strategy, across the entire 2016 dataset time range (163 days). The two dates given per event are the date of that event (first) and the date of the respective reaction in Twitter (second).

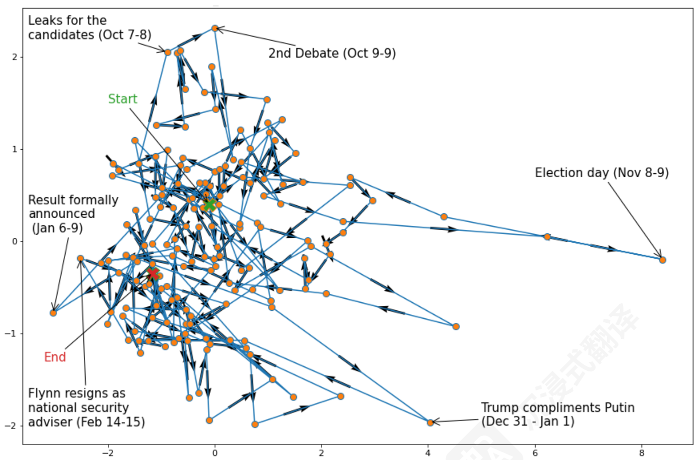


**Fig. 7** PCA-based 2D visualization of the constructed 4D timeseries for the Republicans, using a mean aggregation strategy, across the entire 2016 dataset time range (163 days). The two dates given per event are the date of that event (first) and the date of the respective reaction in Twitter (second).

However, the following observations can be made based on the outliers. Regarding the 2016 data we can tell from Figures 6, 7 that the events influencing public opinion about the Democrats the most were the elections themselves, the formal announcement of the results and the leaks about the candidates. The respective events for the Republicans are identical, with the addition of the second candidate debate and the compliments made by Donald Trump on



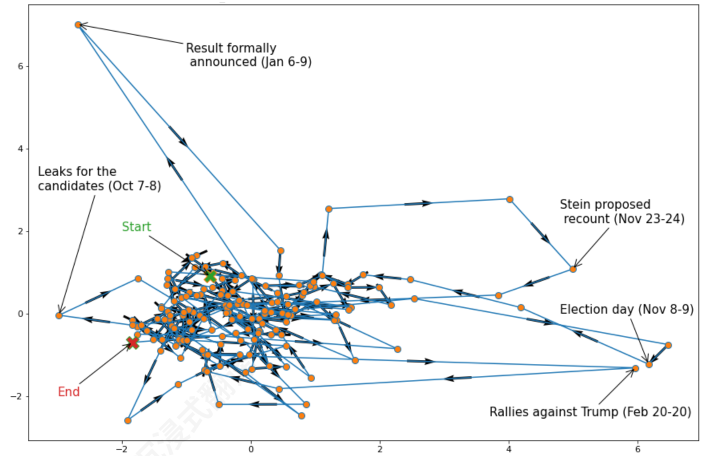


**图6** 基于PCA的2D可视化，展示了为民主党构建的4D时间序列，使用均值聚合策略，跨越整个2016数据集的时间范围（163天）。每个事件给出的两个日期是该事件（第一个）和Twitter中的相应反应日期（第二个）。

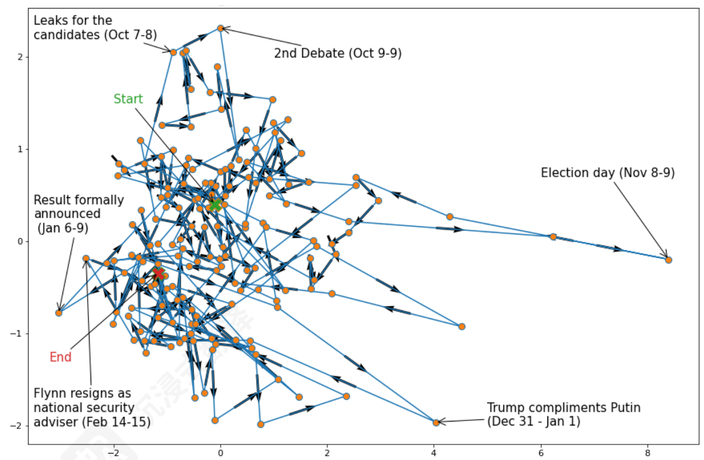


**图7** 基于PCA的2D可视化，展示了为共和党构建的4D时间序列，使用均值聚合策略，跨越整个2016数据集的时间范围（163天）。每个事件给出的两个日期是该事件（第一个）和Twitter中的相应反应日期（第二个）。

然而，根据异常值可以进行以下观察。关于2016年的数据，我们可以从图 6、 7 中看出，影响公众对民主党舆论最的事件是选举本身、结果正式宣布以及关于候选人的泄露。共和党的相应事件是相同的，增加了第二次候选人辩论以及唐纳德·特朗普对

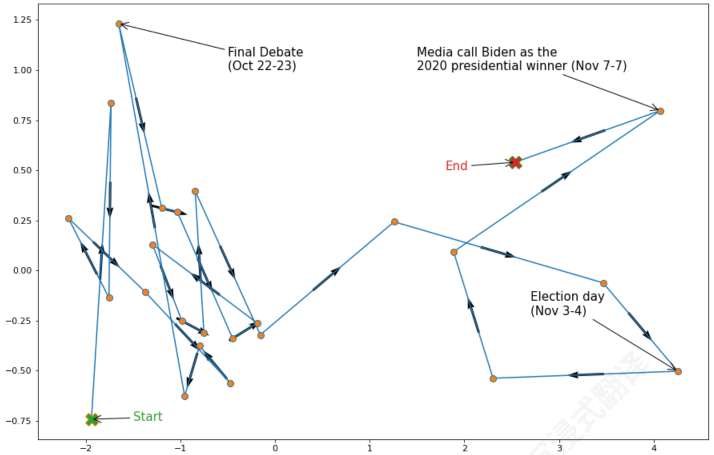


**Fig. 8** PCA-based 2D visualization of the constructed 4D timeseries for the Democrats, using a mean aggregation strategy, across the entire 2020 dataset time range (25 days). The two dates given per event are the date of that event (first) and the date of the respective reaction in Twitter (second).

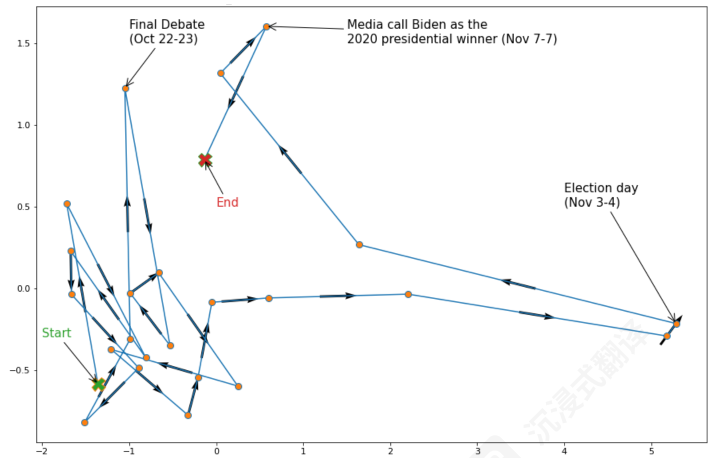


**Fig. 9** PCA-based 2D visualization of the constructed 4D timeseries for the Republicans, using a mean aggregation strategy, across the entire 2020 dataset time range (25 days). The two dates given per event are the date of that event (first) and the date of the respective reaction in Twitter (second).

President Putin. This possibly reflects the increased relevance of national security concerns in the US public discourse. Regarding the 2020 data, Figures 8, 9 confirm the three crucial events indicated by the spikes in Figures 3, 5.

Given the explanatory power of specific dates shown to be semantic outliers in the constructed timeseries, a different way to exploit the proposed public opinion description mechanism was also investigated: to focus on individual salient dates. In the context of this paper and given the previously discussed



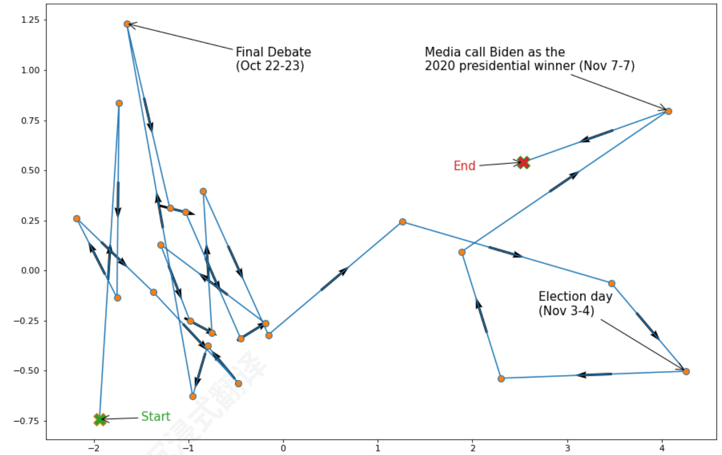


**图8** 基于PCA的2D可视化，展示了为民主党构建的4D时间序列，使用均值聚合策略，覆盖整个2020年数据集的时间范围（25天）。每个事件给出的两个日期分别是该事件日期（第一个）和Twitter中的相应反应日期（第二个）。

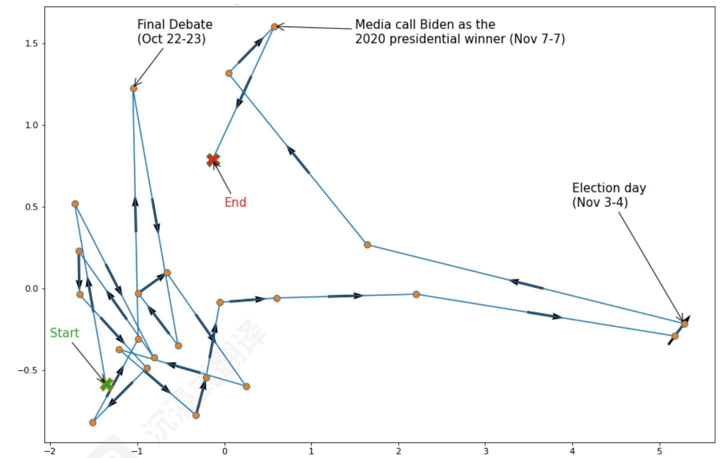


**Fig. 9** PCA-based 2D visualization of the constructed 4D timeseries for the Republicans, using a mean aggregation strategy, across the entire 2020 dataset time range (25 days). The two dates given per event are the date of that event (first) and the date of the respective reaction in Twitter (second).

普京总统。这可能与美国公共话语中国家安全问题的相关性增加有关。关于2020年的数据，图 8、9 证实了图 3、5中的三个关键事件。

鉴于特定日期在构建的时间序列中被证明为语义异常值，解释力，一种不同的方法也被研究：专注于单个显著的日期。在本文的背景下，鉴于之前讨论的内容，还研究了利用所提出的舆论描述机制的不同方式。

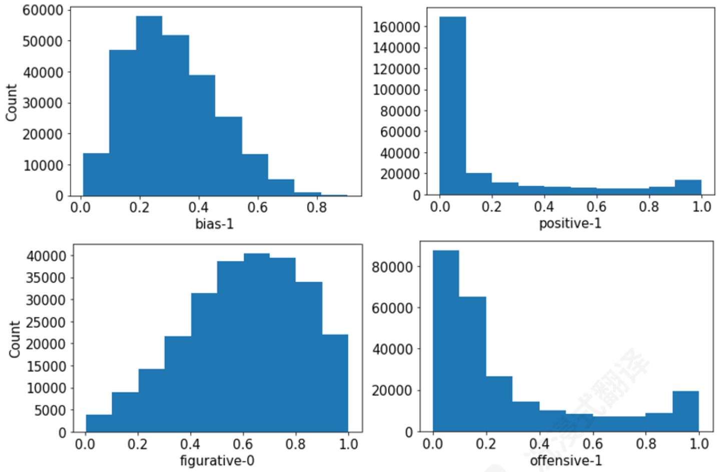


**Fig. 10** Histograms of the four descriptor dimensions, depicting how the number of tweets is distributed over the DNN classifier outputs. These histograms concern the Democrats on Nov 9, 2016 (the day after election).

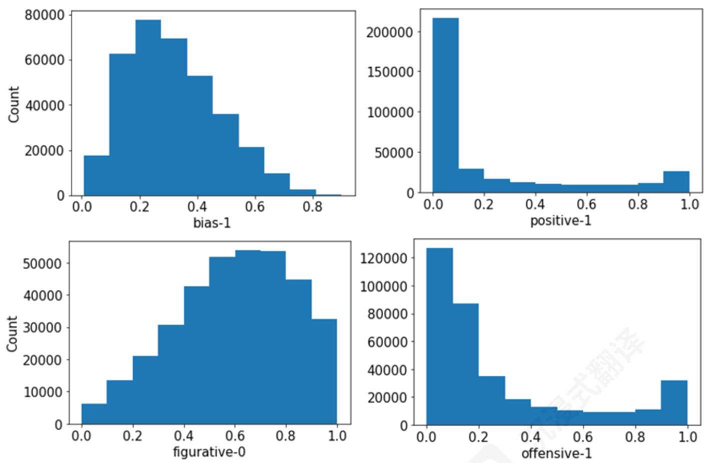


**Fig. 11** Histograms of the four descriptor dimensions, depicting how the number of tweets is distributed over the DNN classifier outputs. These histograms concern the Republicans on Nov 9, 2016 (the day after election).

observations, the day after the elections (November 9, 2016 and November 4, 2020) was selected as the target date.

Figures 10 and 11 show how the tweets posted on November 9, 2016 were distributed along the four descriptor dimensions, separately for the two parties and before any aggregation strategy was applied. These 10-bin histograms show the distribution of the number of tweets (vertical axis) over the semantic values of each of the four descriptor dimensions (horizontal axis). The employed semantic values (outputted by the 4 pretrained DNN classifiers) were

Springer Nature 2021 LATEX模板



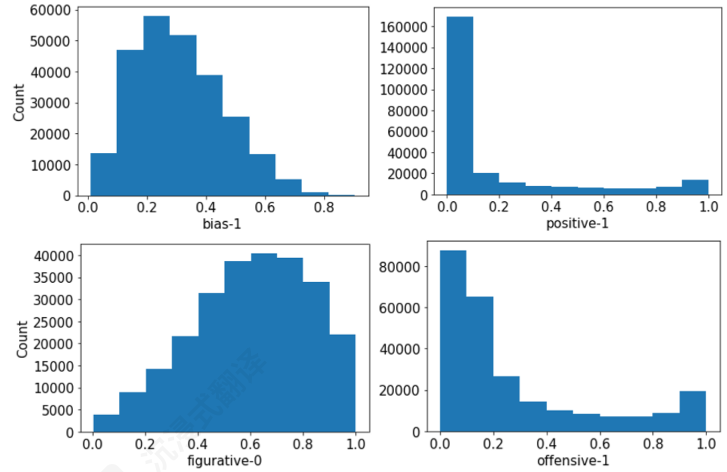


**图10** 四个描述维度直方图，描绘推文数量如何在DNN分类器输出上分布。这些直方图涉及2016年11月9日（大选后一天）的民主党人。

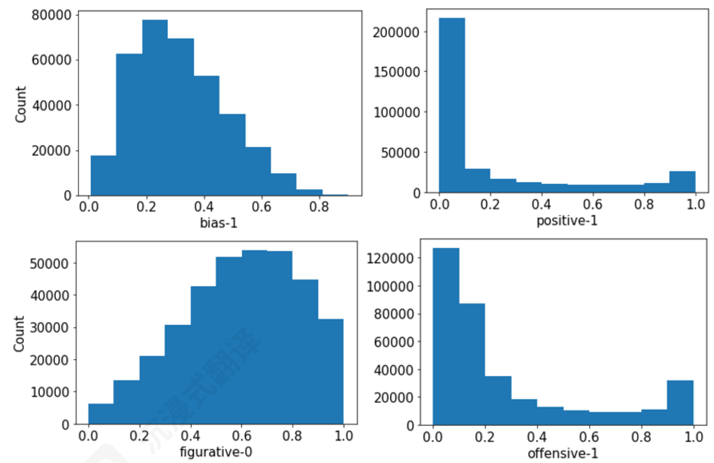


**图11** 四个描述维度直方图，描绘推文数量如何在DNN分类器输出上分布。这些直方图涉及2016年11月9日（大选后一天）的共和党人。

观察结果，大选后（2016年11月9日和2020年11月4日）被选为目标日期。

图 10 和 11 显示2016年11月9日发布的推文如何在四个描述维度上分布，分别针对两个政党，且在应用任何聚合策略之前。这些10箱直方图显示了推文数量（垂直轴）在四个描述维度的语义值（水平轴）上的分布。所使用的语义值（由4个预训练DNN分类器输出）是

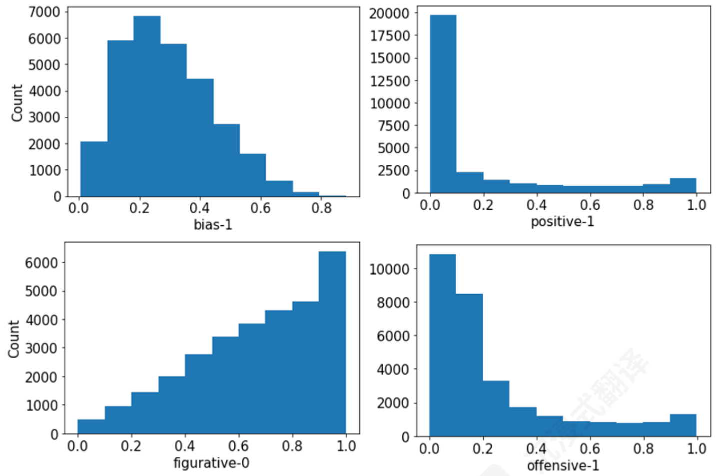


**Fig. 12** Histograms of the four descriptor dimensions, depicting how the number of tweets is distributed over the DNN classifier outputs. These histograms concern the Democrats on Nov 4, 2020 (the day after the election).

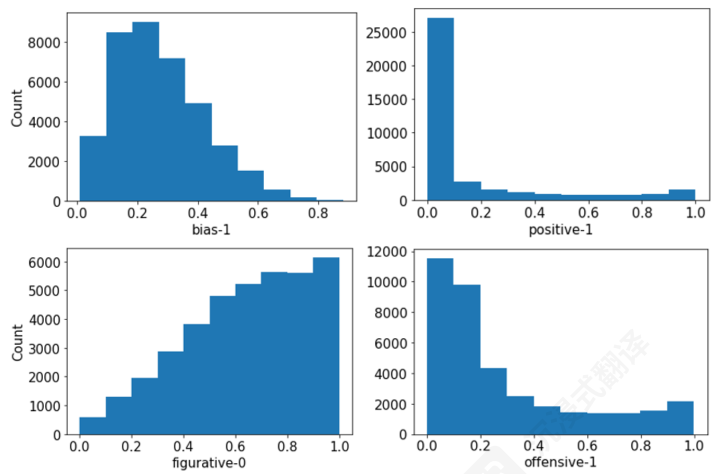


**Fig. 13** Histograms of the four descriptor dimensions, depicting how the number of tweets is distributed over the DNN classifier outputs. These histograms concern the Republicans on Nov 4, 2020 (the day after the election).

real numbers in the interval [0,1]. Therefore, the horizontal axis has not been normalized in range: each of the 10 bins corresponds to a subrange of length 0.1.

The histograms are almost identical for Democrats and Republicans, an observation compatible with the behaviour captured in Figure 4. Moreover, similar histogram shapes can be discerned for the bias-figurativeness and for the polarity-offensiveness features. Bias and figurativeness have approximately shifted normal distributions, implying that the mean aggregation strategy is



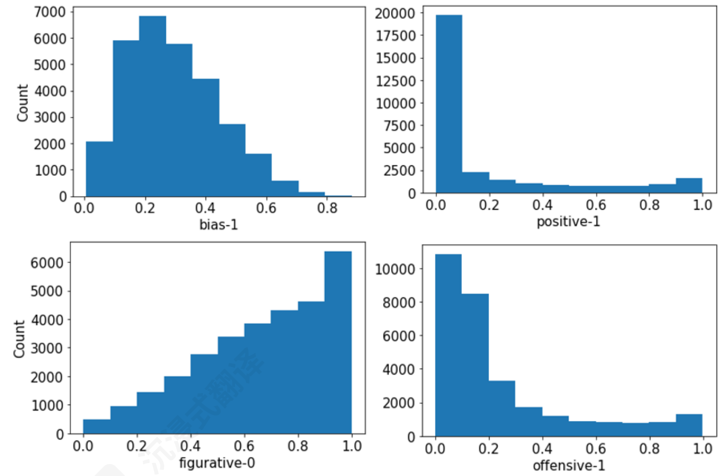


**图12** 四个描述维度上的直方图，描绘了推文数量如何在DNN分类器输出上分布。这些直方图涉及2020年11月4日（大选次日）的民主党人。

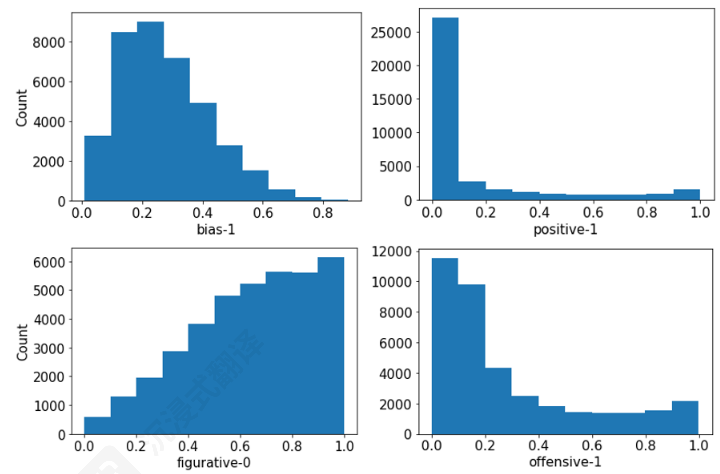


**图13** 四个描述维度上的直方图，描绘了推文数量如何在DNN分类器输出上分布。这些直方图涉及2020年11月4日（大选次日）的共和党人。

实数位于区间 [0,1]中。因此，水平轴未在此范围内进行归一化：每个10个bin对应一个长度为0。1的子区间。

民主党人和共和党人的直方图几乎相同，这一观察结果与图4中捕获的行为一致。此外，对于偏见-形象性和情感-攻击性特征，可以观察到相似的直方图形状。偏见和形象性呈近似正态分布，这意味着均值聚合策略是

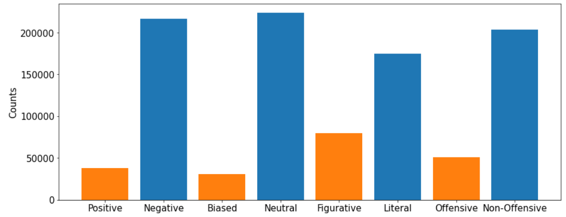


**Fig. 14** Number of tweets concerning Democrats, separately for each class of the four descriptor dimensions, on November 9, 2016. The two colors distinguish between the opposite classes of each semantic dimension.

indeed a good choice during timeseries construction. In contrast, polarity and offensiveness histograms are significantly more polarized in shape, rendering the mean aggregation strategy less reliable in their case for that particular date.

The fact that the majority of tweets fall within the interval [0,0.2] for the polarity and offensiveness dimensions, means that they have been clearly classified as rather negative and non-offensive: a fully negative/positive and absolutely non-offensive/offensive tweet would be characterized by a value of 0/1 in both dimensions, respectively. These histograms paint the picture of a public that is carping and complaining in the aftermath of the elections, yet avoids the use of offensive language. Concerning the relative absence of intermediate values lying within the range [0.2,0.8], we can say that classification regarding polarity and offensiveness was straightforward and the respective models pretty confident. This implies that indeed most tweets were clearly negative or positive, as well as clearly non-offensive or offensive, without many users being neutral in these respects. In contrast, classification regarding bias and figurative attributes does not lead to such polarized results. This is because the DNN models have trouble classifying these tweets as pure instances of a specific class (e.g., the "figurative" or the "literal" class), leading to intermediate values near 0.5. This implies that most users were rather neutral with regard to these semantic dimensions. Still, we can clearly see that the majority of tweets tend to be non-biased and literal.

Similar histogram shapes are observed for the respective tweet distributions of November 4, 2020 for both Democrats 12 and Republicans 13. This was no surprise given the similarity of Figures 4 and 5. However, there is a noticeable difference in the figurative elements of 2016 and 2020, where the distribution is shifted right, indicating more literal language used on that day's tweets. This can be confirmed by comparing Figures 4 and 5.

Finally, Figures 14, 15 and 16, 17 depict the number of tweets per party, separately for each class of the four descriptor dimensions, on the election days November 9, 2016 and November 4, 2020. This visualization provides a glimpse to the non-dominant classes that disappeared when constructing the timeseries using the mean aggregation strategy.



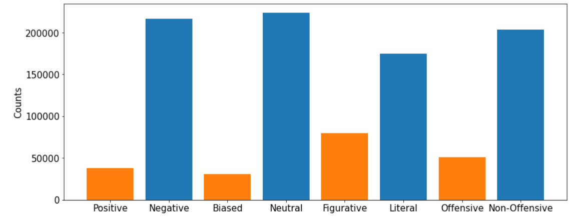


**图14** 关于民主党人的推文数量，分别针对四个描述维度的每个类别，在2016年11月9日。两种颜色区分每个语义维度的相反类别。

确实是在构建时间序列时一个好的选择。相比之下，极性和国民性直方图在形状上明显更两极分化，使得均值聚合策略在那一天对于它们来说不太可靠。

大多数推文在极性和国民性维度上落在区间 [0,0.2]内，这意味着它们已被明确分类为相当负面和非冒犯性的：一个完全负面/正面和绝对非冒犯性/冒犯性的推文分别在两个维度上被表征为0/1的值。这些直方图描绘了一幅在选举后抱怨和挑剔的公众形象，但避免了使用冒犯性语言。关于在范围 [0.2,0.8]，内缺乏中间值，我们可以说极性和国民性的分类很简单，相应的模型相当自信。这意味着确实大多数推文被明确地归类为负面或正面，以及明确地归类为非冒犯性或冒犯性，而没有许多用户在这些方面保持中立。相比之下，关于偏见和比喻属性的分类并没有导致如此两极分化的结果。这是因为DNN模型难以将这些推文分类为特定类别的纯实例（例如，“比喻”或“直白”类别），导致接近0.5的中间值。这意味着大多数用户在这些语义维度上相当中立。尽管如此，我们可以清楚地看到大多数推文倾向于非偏见和直白。

对于2020年11月4日的民主党人和共和党人的推文分布，观察到相似直方图形状12 和13。鉴于图4 和5的相似性，这并不令人意外。然而，2016年和2020年的图形元素存在明显差异，其中分布向右移动，表明那天推文中使用了更多直白的语言。这可以通过比较图4和5来证实

最后，图14、15 和16、17 分别描绘了2016年11月9日和2020年11月4日选举日每个政党的推文数量，并按四个描述维度的每个类别单独显示。这种可视化提供了一窥在构建使用均值聚合策略的时间序列时消失的非主导类别的机会

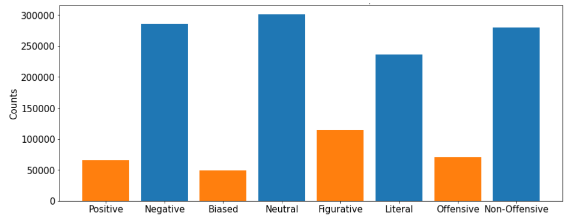


**Fig. 15** Number of tweets concerning Republicans, separately for each class of the four descriptor dimensions, on November 9, 2016. The two colors distinguish between the opposite classes of each semantic dimension.

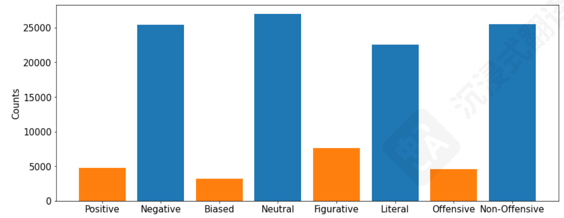


**Fig. 16** Number of tweets concerning Democrats, separately for each class of the four descriptor dimensions, on November 4, 2020. The two colors distinguish between the opposite classes of each semantic dimension.

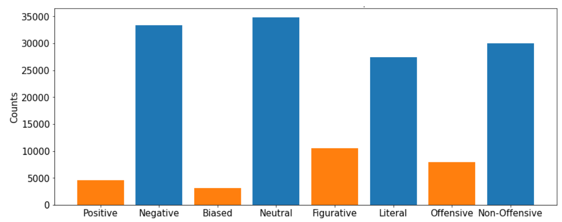


**Fig. 17** Number of tweets concerning Republicans, separately for each class of the four descriptor dimensions, on November 4, 2020. The two colors distinguish between the opposite classes of each semantic dimension.

## 4.4 Analysis 3: Poll/election results prediction

To validate the correlation of the proposed mechanism/descriptor with public opinion actually captured in political polls and election results, an additional set of experiments were conducted using: a) the 2016 USA Presidential Elections dataset, and b) actual, national-level poll results from that pre-election period. The goal was to assess said correlation through estimating the predictive ability of the timeseries that are constructed by using the proposed mechanism.

The following two assumptions were made: a) the final result of a multi-day poll is considered valid for all days during which the census was being conducted, and b) the mean value per day was obtained for different polls covering overlapping periods. Thus, a dataset of mean daily national poll results for the



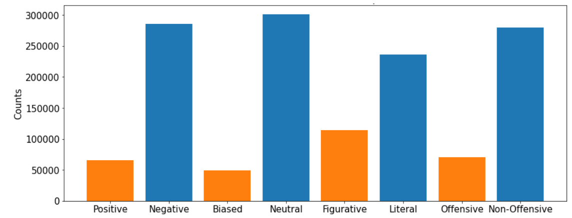


**图15** 关于共和党的推文数量，分别针对四个描述维度的每个类别，在2016年11月9日。两种颜色区分每个语义维度的相反类别。

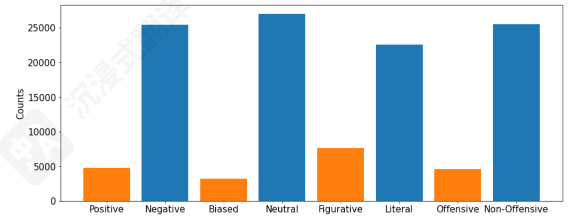


**图16** 关于民主党的推文数量，分别针对四个描述维度的每个类别，在2020年11月4日。两种颜色区分每个语义维度的相反类别。

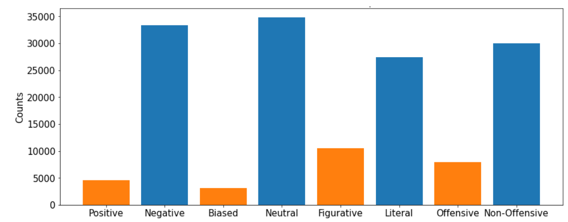


**图17** 关于共和党的推文数量，分别针对四个描述维度的每个类别，在2020年11月4日。两种颜色区分每个语义维度的相反类别。

## 4.4 分析3：民意调查/选举结果预测

为验证所提出的机制/描述符与政治民意调查和选举结果中实际捕捉到的公众舆论的相关性，进行了额外的实验，使用了：a) 2016年美国总统选举数据集，以及 b) 该预选期间的实际全国性民意调查结果。目标是通过对使用所提出机制构建的时间序列的预测能力进行估计，来评估这种相关性。

做出了以下两个假设：a) 多日民意调查的最终结果被认为对所有进行普查期间的天数都有效，以及 b) 对涵盖重叠时期的不同民意调查，获得了每日的均值。因此，为

Democratic and Republican parties was constructed, covering the period from 2016-08-30 up to the election day.

The day-by-day timeseries constructed using the proposed mechanism (under the element-wise vector mean aggregation strategy) and the poll results were temporally aligned for the three-month period prior to the elections. A 7-day sliding window was shifted through these 3 months with a 1-day step, so as to derive the following **x**-**y** pair for each such window ($\mathbf{x} \in \mathbb{R}^{56}$, $\mathbf{y} \in \mathbb{R}^{2}$). The dimensionality of **x** is given by the number of parties (2) times the descriptor timeseries dimensionality (4) times the days covered by the window (7). The 2 entries of the respective vector **y** are the poll results (one vote percentage per party) of the day following the current temporal window. Notably, *the actual election results were employed instead of polls for the last window*. The outcome of this process was a regression dataset for learning to map public opinion descriptors constructed according to the proposed mechanism to poll/election results.

Overall, 55 **x**-**y** pairs were contained in this dataset. A single-hidden-layer MultiLayer Perceptron (MLP) was trained as the regression model, using a random 80%/20% training/test split. The temporally last data point (for the time window leading to the election day) was manually selected to be in the test set. 5-fold cross-validation in the training set was employed for manual hyperparameter tuning. Data point sampling was randomized during training, while preprocessing included only min-max normalization. The model was implemented in PyTorch, using an Adam optimizer and a Mean Square Error loss function. Optimal batch size, number of hidden neurons and learning rate were found to be 16, 16 and 0.04, respectively, while training proceeded for 40 epochs.

The exact same process was then repeated from scratch, but using only the polarity dimension from the timeseries derived through the proposed mechanism. This was done in order to emulate previous Twitter/NLP-based public opinion quantification methods from the existing literature, which only consider polarity, and compare against them. Thus, this reduced dataset contained modified $\tilde{\mathbf{x}}$-**y** pairs, where $\tilde{\mathbf{x}} \in \mathbb{R}^{14}$. Prediction results are shown in Table 6, using the MAPE metric for evaluating test accuracy. As it can be seen, the error achieved by using the proposed mechanism is very low (under 5%) and significantly lower compared to the case where only the polarity dimension is exploited.

Two conclusions can be drawn from this: a) during its training, the employed MLP successfully discovered strong correlations between the outcomes of the proposed mechanism and the poll/election results, and b) a reduced version of the proposed mechanism that stands for previous methods found in the literature (taking only polarity into account) performs worse than our method.

To further validate these results, we separately computed the normalized Pearson correlation (with values within the real range $[-1, 1]$) between each of the 8 timeseries for this 3-month period (4 timeseries per party) and the



民主党和共和党构建，涵盖2016-08-30至选举日的时间段。

使用所提出机制（在元素向量均值聚合策略下）构建的每日时间序列与民意调查结果在选举前三个月内进行了时间对齐。在这3个月内，使用7天滑动窗口以1天步长移动，以便为每个这样的窗口导出以下**x**-**y** 对（$\mathbf{x} \in \mathbb{R}^{56}$, $\mathbf{y} \in \mathbb{R}^{2}$）。**x** 的维度由政党的数量（2）乘以描述符时间序列的维度（4）乘以窗口覆盖的天数（7）给出。相应的向量**y** 的2个条目是当前时间窗口之后当天的民意调查结果（每个政党一个投票百分比）。值得注意的是，最后窗口使用的是实际选举结果而不是民意调查。该过程的结果是一个回归数据集，用于学习将根据所提出机制构建的舆论描述符映射到民意调查/选举结果。

总体而言，该数据集包含55 **x**-**y** 对。使用随机80%/20%的训练/测试分割，训练了一个单隐藏层的多层感知器（MLP）作为回归模型。时间上最后的数据点（用于选举日前的时间窗口）被手动选择为测试集。在训练集中采用了5折交叉验证进行手动超参数调整。训练期间数据点采样是随机的，而预处理仅包括最小-最大归一化。该模型使用PyTorch实现，采用Adam优化器和均方误差损失函数。最佳批大小、隐藏神经元数量和学习率分别为16、16和0.04，训练进行了40个epoch。

然后，从零开始重复了完全相同的流程，但仅使用通过所提出机制导出的时间序列的极性维度。这是为了模拟现有文献中仅考虑极性的Twitter/NLP-based公众舆论量化方法，并与它们进行比较。因此，这个缩减的数据集包含修改的 $\tilde{\mathbf{x}}$-**y** 对，其中 $\tilde{\mathbf{x}} \in \mathbb{R}^{14}$。预测结果如表6所示，使用MAPE指标评估测试精度。如所见，使用所提出机制实现的误差非常低（低于5%），并且与仅利用极性维度的情况相比显著更低。

可以得出两个结论：a) 在其训练过程中，所使用的MLP成功地发现了所提出的机制的结果与民意调查/选举结果之间的强相关性，以及 b) 所提出的机制的简化版本（仅考虑极性）比我们提出的方法更差。

为了进一步验证这些结果，我们分别计算了这3个月期间（每个政党4个时间序列）的8个时间序列中的每一个与之间的标准化皮尔逊相关系数（值在真实范围内 $[-1, 1]$）。

**Table 6** Poll/election results prediction accuracy, evaluated using the MAPE metric (percentage, lower is better).

| Method | MAPE |
|---|---|
| Full proposed mechanism | 4.17 % |
| Polarity-only mechanism | 6.53 % |

respective poll timeseries (one per party). As it can be seen in Fig. 18, the majority of the derived timeseries are strongly correlated with polls, either positively or negatively, with correlation values away from 0 at least for one of the two parties (in most cases for both). Note that in this Figure, as in many previous ones, we label each timeseries by one of the two opposite class labels it encodes (e.g., "positive" or "figurative") followed by the tweet classifier output value which implies this label. E.g., a tweet fully and undoubtedly classified as figurative/positive, has been assigned a value of 0/1 by the figurativeness/polarity classifier, respectively. In contrast, a tweet fully and undoubtedly classified as literal/negative, would have been assigned a value of 1/0 by the figurativeness/polarity classifier, respectively. Party affiliation is color-coded.

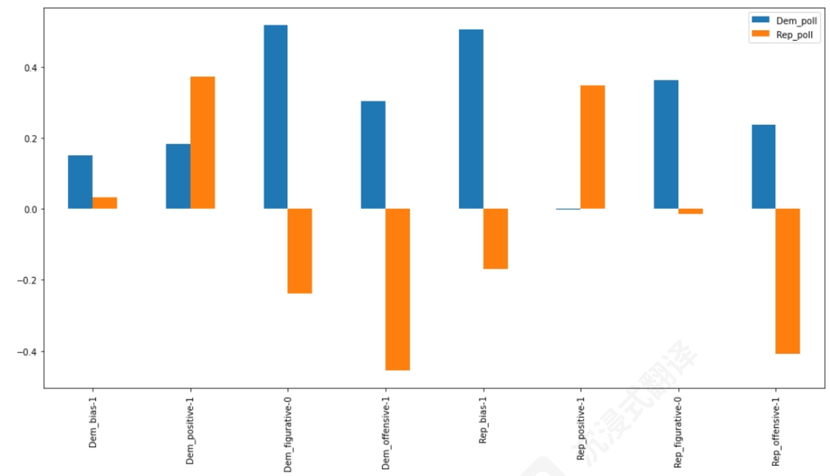


**Fig. 18** Pearson correlation between each of the 8 timeseries derived by the proposed mechanism (4 per party) with the respective poll timeseries, for the 3-month period before the 2016 US presidential elections.

## 5 Discussion

The evaluation presented in Section 4 indicates that the proposed mechanism for automated public opinion monitoring through Twitter is a very powerful tool, able to provide valuable information for more efficient decision-making. The **multidimensional nature** of the presented descriptor conveys rich insights (analyzed in Section 4) that are not typically captured by existing

Springer Nature 2021 LaTeX template



**表6** 民意/选举结果预测准确度，使用MAPE指标评估（百分比，越低越好）。

| 方法 | MAPE |
|---|---|
| Full prop osed mechanism | 4.17 % |
| Polarity-only mechanism | 6.53 % |

相应的投票时间序列（每个政党一个）。如图 18 所示，大多数派生的时间序列与投票结果高度相关，无论是正相关还是负相关，至少对于两个政党中的一个（在大多数情况下对于两个政党）的相关值偏离0。请注意，在这个图中，就像在许多之前的图中一样，我们通过它编码的两个相反的类别标签之一（例如，“正面”或“比喻”）为每个时间序列打标签，然后是推文分类器输出值，该值意味着这个标签。例如，一个完全且无疑被分类为比喻/正面的推文，已经被比喻性/极性分类器分别分配了0/1的值。相反，一个完全且无疑被分类为字面/负面的推文，将被比喻性/极性分类器分别分配1/0的值。政党归属按颜色编码。

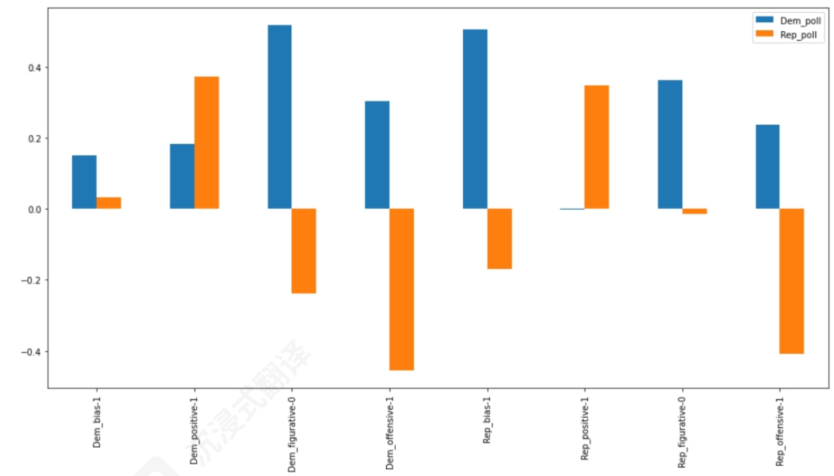


**图** 在2016年美国总统选举前3个月的时期内，所提出的机制（每个政党4个）派生的8个时间序列与相应的投票时间序列之间的皮尔逊相关性。

## 5 讨论

第 4 章中提出的评估表明，通过 Twitter 进行的自动公众舆论监测所提出的机制是一种非常强大的工具，能够为更有效的决策提供有价值的信息。The 4 多维度特性 **所提出的描述符传达了丰富的见解（在 第 4 章中分析），这些见解通常不会被现有**

relevant methods, which only exploit sentiment (and, rarely, also bias). This is shown quantitatively in Subsection 4.4, but also through the qualitative insights extracted in Subsection 4.3. Moreover, unlike the vast majority of previously published methods, the proposed mechanism relies on state-of-the-art **DNN-based NLP tools**, a fact which guarantees enhanced accuracy in comparison to existing comparable approaches. Finally, Subsection 4.4 indicates that the proposed descriptor can indeed be exploited for **successful prediction of future poll/election results**, with its multidimensional opinion semantics giving it an advantage over previous similar approaches.

To succinctly demonstrate the usefulness of the proposed mechanism in political analysis, the most important findings extracted by applying it to the datasets of Section 4 (concerning the US presidential elections of 2016 and 2020) are summarized below:

- Public opinion concerning Democrats was less stable and less predictable, in comparison to public opinion about Republicans.
- However, in general, the public has an overall relatively stable stance towards the competing politicians: judgmental and indignant (negative + literal) but simultaneously educated (unbiased + non-offensive).
- The timeseries derived through the proposed mechanism paint a rather accurate picture of the favored candidate. This is shown both through visualizing/inspecting the timeseries and through exploiting them for learning to quantitatively predict poll/election outcomes.
- The winning party is referenced during the pre-election period in tweets that are jointly less negative + less offensive + more biased. Strong partisan presence in Twitter seems to be heavily correlated with high vote percentages.
- Crucial events do directly lead to abrupt changes in the daily number of tweets (which is to be expected), but also in public opinion. This indicates that committed/stable partisan supporters are always a minority in the American Twitter. Moreover, post hoc visualizations of the timeseries automatically derived through the proposed mechanism can actually showcase which events were the most crucial to shifts in public opinion.

A few of these findings verify similar conclusions previously drawn in the existing literature: [20] (the public has an overall relatively stable, negative stance towards the competing politicians, during a specific pre-election period, while the public sentiment timeseries paint a rather accurate picture of the favored candidate) and [13] [18] [20] (crucial events directly lead to abrupt changes in public opinion). However, the majority of our findings for the US presidential elections of 2016 and 2020, as detailed in Section 4, are original contributions of this paper. Most importantly though, the proposed mechanism is not tied to these specific elections. *It is a fully generic and almost fully automated method*, that allows interested users to easily extract similar insights *for any time period*.



相关方法，这些方法仅利用情感（很少情况下也利用偏见）。这在子节 4.4中定量地显示，但也通过子节 4.3中提取的定性见解。此外，与先前发表的大多数方法不同，所提出的机制依赖于最先进的**基于DNN的自然语言处理工具**，这一事实保证了与现有可比方法的准确性提升。最后，子节 4.4 表明，所提出的描述符确实可以用于**成功预测未来的民意调查/选举结果**，其多维度的舆论语义使其比以前类似方法更具优势。

为了简洁地展示所提出机制在政治分析中的实用性，将应用于第4 (关于2016年和2020年美国总统选举)的数据集所提取的最重要发现总结如下：

- 关于民主党人的舆论不如共和党人的舆论稳定和可预测。
- 然而，总的来说，公众对竞争政客的立场相对稳定：评判性和愤慨（负面 ＋字面）但同时又受教育（无偏见 ＋ 非冒犯）。
- 通过所提出的机制推导出的时间序列描绘了被支持候选人的相当准确的图像。这既通过可视化/检查时间序列显示，也通过利用它们来学习定量预测民意调查/选举结果。
- 获胜党派在选举前期间被提及的推文在共同负面 ＋ 较少攻击性 ＋更偏袒。在Twitter中的强烈党派存在似乎与高投票百分比高度相关。
- 关键事件确实直接导致每日推文数量（这是可以预料的）发生急剧变化，但也导致公众舆论的变化。这表明在美国Twitter中，忠诚/稳定的党派支持者始终是少数。此外，通过所提出的机制自动推导出时间序列的事后可视化实际上可以展示哪些事件对公众舆论的转变最为关键。

其中一些发现验证了现有文献中先前得出的相似结论： [20](在特定的选举前时期，公众对竞争政客的整体态度相对稳定且负面，而公众情绪时间序列则能相当准确地描绘出最受欢迎的候选人)和 [13] [18] [20] (关键事件直接导致公众舆论的急剧变化)。然而，我们关于2016年和2020年美国总统选举的大多数发现，如第4节所述，是本文的原创贡献。最重要的是，所提出的机制并非与这些特定选举绑定。它是一种完全通用且几乎完全自动化的方法，允许感兴趣的用户轻松提取类似见解针对任何时间段。

## 6 Conclusions

Automated public opinion monitoring using social media is a very powerful tool, able to provide interested parties with valuable insights for more fruitful decision-making. Twitter has gained significant attention in this respect, since people use it to express their views and politicians use it to reach their voters.

This paper presented a novel, automated public opinion monitoring mechanism, consisting of a composite, quantitative, semantic descriptor that relies on NLP algorithms. A four-dimensional vector, i.e., an instance of the proposed descriptor, is first extracted for each tweet independently, quantifying text polarity, offensiveness, bias and figurativeness. Subsequently, the computed descriptors are summarized across multiple tweets, according to a desired aggregation strategy (e.g., arithmetic mean) and aggregation target (e.g., a specific time period). This can be exploited in various ways; for example, aggregating the tweets of each days separately allows us to construct a multivariate timeseries which can be used to train a forecasting AI algorithm, for day-by-day public opinion predictions.

In order to evaluate the usefulness of the proposed mechanism, it was applied to the large-scale 2016/2020 US Presidential Elections tweet datasets. The resulting succinct public opinion descriptions were successfully employed to train a DNN-based public opinion forecasting model with a 7-day forecasting horizon. Moreover, the constructed timeseries were thoroughly inspected in a qualitative manner in order to deduce insights about public opinion during a heated pre/post-election period. Finally, a set of regression experiments verified: a) the importance of the multidimensional opinion semantics captured in the derived timeseries, and b) the correlation of these timeseries with "ground-truth" public opinion, as captured in actual political polls and election results.

## Acknowledgment

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951911 (AI4Media). This publication reflects only the authors' views. The European Commission is not responsible for any use that may be made of the information it contains.

The authors would like to thank Georgios Chatziparaskevas for his technical aid in the evaluation of the proposed mechanism.

## Competing Interests

The authors have no competing interests to declare that are relevant to the content of this article.

Springer Nature 2021 LaTeX template

*Public opinion monitoring through collective semantic analysis of tweets* 29

## 6 结论

利用社交媒体进行自动化舆情监测是一种非常强大的工具，能够为相关方提供有价值的洞察，以促进更富有成效的决策。在这方面，推特受到了广泛关注，因为人们用它来表达观点，而政治家则用它来接触选民。

本文提出了一种新颖的自动化舆情监测机制，该机制由一个基于NLP算法的复合定量语义描述符组成。首先，为每条推文独立提取一个四维向量，即所提出描述符的实例，量化文本极性、攻击性、偏见和形象性。随后，根据所需的聚合策略（例如，算术平均值）和聚合目标（例如，特定时间段），对计算出的描述符在多条推文之间进行汇总。这可以以多种方式利用；例如，分别聚合每天推文，使我们能够构建一个多变量时间序列，该时间序列可用于训练预测人工智能算法，以进行每日舆情预测。

为了评估所提出机制的有效性，将其应用于大规模的 2016/2020 美国总统选举推文数据集。由此产生的简洁公众舆论描述成功用于训练一个基于 DNN 的公众舆论预测模型，预测范围为 7 天。此外，构建的时间序列在定性方面进行了彻底检查，以推断在激烈的选举前/后期间公众舆论的见解。最后，一组回归实验验证了：a) 所得时间序列中捕获的多维舆论语义的重要性，以及 b) 这些时间序列与“真实”公众舆论的相关性，这些舆论体现在实际政治民意调查和选举结果中。

## 致谢

导致这些结果的研究得到了欧盟“地平线 2020”研究和创新计划的资助，资助协议号为 951911（AI4Media）。本出版物仅反映作者的观点。欧洲委员会对其中包含信息的任何使用不承担责任。

作者感谢 Georgios Chatziparaskevas 在评估所提出的机制方面提供的技术援助。

## 竞争利益

作者没有需要声明的与本文内容相关的竞争利益。

## Data availability

The datasets generated during the current study are available from the corresponding author on reasonable request. The raw 2016 US Presidential Elections tweet dataset is publicly available at https://www.kaggle.com/paulrohan2020/2016-usa-presidential-election-tweets61m-rows.

## References


[1] El Barachi, M., AlKhatib, M., Mathew, S., Oroumchian, F.: A novel sentiment analysis framework for monitoring the evolving public opinion in real-time: Case study on climate change. Journal of Cleaner Production, 127820 (2021)

[2] Romero, D.M., Meeder, B., Kleinberg, J.: Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on Twitter. In: Proceedings of the International Conference on World Wide Web (2011)

[3] Baumgartner, J.C., Mackay, J.B., Morris, J.S., Otenyo, E.E., Powell, L., Smith, M.M., Snow, N., Solop, F.I., Waite, B.C.: Communicator-in-chief: How Barack Obama Used New Media Technology to Win the White House. Lexington Books, ??? (2010)

[4] Lorenzi, D., Vaidya, J., Shafiq, B., Chun, S., Vegesna, N., Alzamil, Z., Adam, N., Wainer, S., Atluri, V.: Utilizing social media to improve local government responsiveness. In: Proceedings of the Annual International Conference on Digital Government Research (2014)

[5] Ravi, K., Ravi, V.: A survey on opinion mining and sentiment analysis: tasks, approaches and applications. Knowledge-based systems **89**, 14–46 (2015)

[6] Ramteke, J., Shah, S., Godhia, D., Shaikh, A.: Election result prediction using Twitter sentiment analysis. In: Proceedings of the IEEE International Conference on Inventive Computation Technologies (ICICT) (2016)

[7] Agarwal, A., Singh, R., Toshniwal, D.: Geospatial sentiment analysis using Twitter data for UK-EU referendum. Journal of Information and Optimization Sciences **39**(1), 303–317 (2018)

[8] Chatfield, C.: Time-series Forecasting. CRC Press, ??? (2000)

[9] Bright, L.F., Sussman, K.L., Wilcox, G.B.: Facebook, trust and privacy in an election year: Balancing politics and advertising. Journal of Digital & Social Media Marketing **8**(4), 332–346 (2021)






## 数据可用性

本研究期间生成的数据集可在合理请求下从通讯作者处获取。原始的2016年美国总统选举推文数据集可在
https://www.kaggle.com/paulrohan2020/2016-usa-presidential-election-tweets61m-rows。

## 参考文献


[1] El Barachi, M., AlKhatib,M., Mathew, S., Oroumchian, F.: 一种用于实时监测气候变化演变舆论的新型情感分析框架：案例研究. 清洁生产杂志, 127820 (2021)

[2] Romero, D.M., Meeder, B., Kleinberg, J.: 不同主题间信息传播机制的差异：推文中的习语、政治标签和复杂传染. 国际万维网会议论文集 (2011)

[3] Baumgartner, J.C., Mackay, J.B., Morris, J.S., Otenyo, E.E., Powell, L., Smith, M.M., Snow, N., Solop, F.I., Waite, B.C.: Communicator-in-chief: How Barack Obama Used New Media Technology to Win the White House. Lexington Books, ??? (2010)

Lorenzi, D., Vaidya, J., Shafiq, B., Chun, S., Vegesna, N., Alzamil, Z., Adam, N., Wainer, S., Atluri, V.: 利用社交媒体提升地方政府响应能力。在：第 annual International Conference on Digital Government Research 会议论文集 (2014)

[5] Ravi, K., Ravi, V.: 关于观点挖掘和情感分析的调查：任务、方法和应用。Knowledge-basedsystems **89**, 14–46 (2015)

Ramteke, J., Shah, S., Godhia, D., Shaikh, A.: 使用推特情感分析预测选举结果。在：IEEE 国际创新计算技术会议论文集 (ICICT) (2016)

[7]Agarwal, A., Singh, R., Toshniwal, D.: 基于Twitter数据的英国-欧盟公投地理空间情感分析。信息与优化科学杂志 **39**(1),303–317 (2018)

[8] Chatfield, C.: 时间序列预测。CRC出版社，??? (2000)

[9] Bright, L.F., Sussman, K.L., Wilcox, G.B.: 选举年中的Facebook、信任和隐私：平衡政治与广告。数字与社会媒体营销杂志 **8**(4), 332–346 (2021)

[10] Lee, J., Lim, Y.-S.: Gendered campaign tweets: the cases of Hillary Clinton and Donald Trump. Public Relations Review **42**(5), 849–855 (2016)

[11] Buccoliero, L., Bellio, E., Crestini, G., Arkoudas, A.: Twitter and politics: Evidence from the US presidential elections 2016. Journal of Marketing Communications **26**(1), 88–114 (2020)

[12] Grimaldi, D.: Can we analyse political discourse using twitter? evidence from Spanish 2019 presidential election. Social Network Analysis and Mining **9**(1), 1–9 (2019)

[13] Cornfield, M.: Empowering the party-crasher: Donald J. Trump, the first 2016 GOP presidential debate, and the Twitter marketplace for political campaigns. Journal of Political Marketing (2017)

[14] Heredia, B., Prusa, J., Khoshgoftaar, T.: Exploring the effectiveness of Twitter at polling the United States 2016 presidential election. In: Proceedings of the IEEE International Conference on Collaboration and Internet Computing (CIC) (2017)

[15] Grimaldi, D., Cely, J.D., Arboleda, H.: Inferring the votes in a new political landscape: the case of the 2019 Spanish presidential elections. Journal of Big Data **7**(1), 1–19 (2020)

[16] Tavoschi, L., Quattrone, F., D'Andrea, E., Ducange, P., Vabanesi, M., Marcelloni, F., Lopalco, P.L.: Twitter as a sentinel tool to monitor public opinion on vaccination: an opinion mining analysis from September 2016 to August 2017 in Italy. Human Vaccines & Immunotherapeutics **16**(5), 1062–1069 (2020)

[17] Wang, M., Wu, H., Zhang, T., Zhu, S.: Identifying critical outbreak time window of controversial events based on sentiment analysis. PLOS One **15**(10), 0241355 (2020)

[18] Shi, W., Wang, H., He, S.: Sentiment analysis of Chinese microblogging based on sentiment ontology: a case study of '7.23 Wenzhou Train Collision'. Connection Science **25**(4), 161–178 (2013)

[19] Onyenwe, I., Nwagbo, S., Mbeledogu, N., Onyedinma, E.: The impact of political party/candidate on the election results from a sentiment analysis perspective using# anambradecides2017 tweets. Social Network Analysis and Mining **10**(1), 1–17 (2020)

[20] Yaqub, U., Chun, S.A., Atluri, V., Vaidya, J.: Sentiment based analysis of tweets during the US presidential elections. In: Proceedings of the Annual International Conference on Digital Government Research (2017)

Springer Nature 2021 LATEX template



[10] Lee, J., Lim, Y.-S.: 性别化竞选推文：希拉里·克林顿和唐纳德·特朗普的案例。公共关系评论 **42**(5), 849–855 (2016)

[11] Buccoliero, L., Bellio, E., Crestini, G., Arkoudas, A.: 推特与政治：来自2016年美国总统大选的证据。营销传播杂志 **26**(1), 88–114 (2020)

[12] Grimaldi, D.: 我们能利用推特分析政治话语吗？来自2019年西班牙总统大选的证据。社交网络分析与管理 **9**(1),1–9 (2019)

[13] Cornfield, M.: 赋能捣乱者：唐纳德·J·特朗普、2016年首次共和党总统辩论以及政治竞选的推特市场。政治营销杂志 (2017)

[14] Heredia, B., Prusa, J., Khoshgoftaar, T.: 探索推特在2016年美国总统选举中的投票效果。在：IEEE国际协作与互联网计算会议论文集（CIC）（2017）

[15] Grimaldi, D.,Cely,J.D.,Arboleda, H.: 推断新的政治格局中的投票情况：以2019年西班牙总统选举为例。大数据杂志 **7**(1), 1–19 (2020)

[16] Tavoschi, L., Quattrone, F., D'Andrea, E., Ducange, P., Vabanesi, M., Marcelloni, F., Lopalco, P.L.: 推特作为监测疫苗接种公众舆论的哨兵工具：2016年9月至2017年8月在意大利的意见挖掘分析。人类疫苗与免疫治疗 **16**(5), 1062–1069 (2020)

[17] 王, M., 吴,H.,张, T., 朱, S.: 基于情感分析的争议性事件关键爆发时间窗口识别。PLOS One**15**(10), 0241355 (2020)

[18]石, W., 王, H., 何, S.: 基于情感本体的中文微博情感分析：以“7.23 温州火车碰撞”为例。连接科学 **25**(4),161–178 (2013)

[19] Onyenwe, I., Nwagbo, S., Mbeledogu, N., Onyedinma, E.: 从情感分析视角探讨政党/候选人对选举结果的影响——以#ana mbradecides2017推文为例。社交网络分析与管理 **10**(1), 1–17 (2020)

[20] Yaqub, U., Chun, S.A., Atluri, V., Vaidya, J.: 基于情感分析的美国总统大选期间推文分析。在：数字政府研究年度国际会议论文集 (2017)

[21] Kraaijeveld, O., De Smedt, J.: The predictive power of public Twitter sentiment for forecasting cryptocurrency prices. Journal of International Financial Markets, Institutions and Money **65**, 101188 (2020)

[22] Qi, F., Yang, C., Liu, Z., Dong, Q., Sun, M., Dong, Z.: Openhownet: An open sememe-based lexical knowledge base. arXiv preprint arXiv:1901.09957 (2019)

[23] Zhang, S., Wei, Z., Wang, Y., Liao, T.: Sentiment analysis of Chinese micro-blog text based on extended sentiment dictionary. Future Generation Computer Systems **81**, 395–403 (2018)

[24] Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., Kappas, A.: Sentiment strength detection in short informal text. Journal of the American society for information science and technology **61**(12), 2544–2558 (2010)

[25] Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Proceedings of the Advances in Neural Information Processing Systems (NIPS) (2013)

[26] Hutto, C., Gilbert, E.: Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the International AAAI Conference on Web and Social Media (2014)

[27] Maqsood, H., Mehmood, I., Maqsood, M., Yasir, M., Afzal, S., Aadil, F., Selim, M.M., Muhammad, K.: A local and global event sentiment based efficient stock exchange forecasting using deep learning. International Journal of Information Management **50**, 432–451 (2020)

[28] Kordonis, J., Symeonidis, S., Arampatzis, A.: Stock price forecasting via sentiment analysis on Twitter. In: Proceedings of the Pan-Hellenic Conference on Informatics (2016)

[29] Oussalah, M., Zaidi, A.: Forecasting weekly crude oil using Twitter sentiment of US foreign policy and oil companies data. In: Proceedings of the IEEE International Conference on Information Reuse and Integration (IRI) (2018)

[30] Arias, M., Arratia, A., Xuriguera, R.: Forecasting with Twitter data. ACM Transactions on Intelligent Systems and Technology (TIST) **5**(1), 1–24 (2014)

[31] Kiran, R., Kumar, P., Bhasker, B.: OSLCFit (organic simultaneous LSTM and CNN Fit): a novel deep learning based solution for sentiment polarity classification of reviews. Expert Systems with Applications **157**, 113488 (2020)

Springer Nature 2021 LaTeX template



[21] Kraaijeveld, O., De Smedt, J.: 预测公共推特情绪对预测加密货币价格的影响。国际金融市场、机构和货币 **65**, 101188 (2020)

[22]Qi, F., Yang, C., Liu, Z., Dong, Q., Sun, M., Dong, Z.: Open- hownet: 基于开放语义素的词汇知识库。arXiv 预印本 arXiv:1901.09957 (2019)

[23]Zhang, S., Wei, Z., Wang, Y., Liao, T.: 基于扩展情感词典的中文微博文本情感分析。未来计算机系统 **81**,395–403 (2018)

[24] Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., Kappas, A.: 短非正式文本中的情感强度检测。美国信息科学和技术协会杂志 **61**(12), 2544–2558 (2010)

[25] Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean,J.: 分布式词和短语的表示及其组合性. In: Neural Information Processing Systems (NIPS) (2013) 的进展会议论文集

[26] Hutto, C., Gilbert, E.: Vader: 一种简约的基于规则的社交媒体文本情感分析模型. In: 国际人工智能联合会网络与社交媒体会议论文集 (2014)

[27] Maqsood, H.,Mehmood,I.,Maqsood,M., Yasir, M., Afzal, S., Aadil, F., Selim, M.M., Muhammad, K.: 基于本地和全局事件情感的高效股票交易所预测使用深度学习. 国际信息管理杂志 **50**, 432–451 (2020)

[28]Kordonis, J., Symeonidis, S., Arampatzis, A.: 通过推特情感分析进行股票价格预测. In: 泛希腊信息学会议论文集 (2016)

[29] Oussalah, M., Zaidi, A.: 预测每周原油价格，使用美国外交政策和石油公司数据的推特情绪。在：IEEE 国际信息复用与集成会议论文集 (IRI) (2018)

[30]Arias, M., Arratia, A., Xuriguera, R.: 使用推特数据的预测。ACM 智能系统与技术交易 (TIST) **5**(1), 1–24 (2014)

[31] Kiran, R., Kumar, P., Bhasker,B.:OSLCFit (有机同时 LSTM 和 CNN Fit)：一种基于深度学习的用于评论情感极性分类的新解决方案。专家系统应用**157**, 113488 (2020)

[32] Zampieri, M., Malmasi, S., Nakov, P., Rosenthal, S., Farra, N., Kumar, R.: Semeval-2019 task 6: Identifying and categorizing offensive language in social media (offenseval). arXiv preprint arXiv:1903.08983 (2019)

[33] Ling, J., Klinger, R.: An empirical, quantitative analysis of the differences between sarcasm and irony. In: European Semantic Web Conference (2016). Springer

[34] Pennington, J., Socher, R., Manning, C.D.: Glove: Global vectors for word representation. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) (2014)

[35] Hewamalage, H., Bergmeir, C., Bandara, K.: Recurrent Neural Networks for time series forecasting: Current status and future directions. International Journal of Forecasting **37**(1), 388–427 (2021)

[36] Orabona, F., Tommasi, T.: Training deep networks without learning rates through coin betting. Proceedings of the Advances in Neural Information Processing Systems (NIPS) **30**, 2160–2170 (2017)

[37] Hutter, F., Hoos, H.H., Leyton-Brown, K.: Sequential model-based optimization for general algorithm configuration. In: Proceedings of the International Conference on Learning and Intelligent Optimization (2011). Springer

[38] Claveria, O., Monte, E., Torra, S.: Data preprocessing for neural network-based forecasting: does it really matter? Technological and Economic Development of Economy **23**(5), 709–725 (2017)

[39] Hyndman, R.J., Koehler, A.B.: Another look at measures of forecast accuracy. International Journal of Forecasting **22**(4), 679–688 (2006)

[40] Tayal, D.K., Yadav, S., Gupta, K., Rajput, B., Kumari, K.: Polarity detection of sarcastic political tweets. In: Proceedings of the International Conference on Computing for Sustainable Global Development (INDIACom) (2014). IEEE





[Zampieri, M., Malmasi, S., Nakov, P., Rosenthal, S., Farra, N., Kumar, R.: Semeval-2019 任务 6：识别和分类社交媒体中的攻击性语言（offenseval）。arXiv 预印本 arXiv:1903.08983 (2019)

Ling, J., Klinger, R.: 针对讽刺与反讽差异的经验性定量分析. 在: 欧洲语义网会议 (2016). 斯普林格

Pennington, J., Socher, R., Manning, C.D.: Glove: 全局词向量表示。在：自然语言处理经验方法会议论文集（EMNLP）（2014）

Hewamalage, H., Bergmeir, C., Bandara, K.: 循环神经网络在时间序列预测中的应用：现状与未来方向。国际预测杂志 **37**(1), 388–427 (2021)

[36] Orabona, F., Tommasi, T.: 通过硬币投注训练深度网络而不需要学习率. 神经信息处理系统进展会议论文集 (NIPS)**30**, 2160–2170 (2017)

[37] Hutter, F., Hoos, H.H., Leyton-Brown, K.: 基于序列模型的通用算法配置优化. 国际学习与智能优化会议论文集 (2011). Springer

[38] Claveria, O., Monte,E.,Torra, S.: 基于神经网络的预测数据预处理：真的重要吗？ 经济技术与经济发展 **23**(5), 709–725 (2017)

[39]Hyndman, R.J., Koehler, A.B.: 再论预测准确度指标. 国际预测学杂志 **22**(4), 679–688 (2006)

[40]Tayal, D.K., Yadav, S., Gupta, K., Rajput, B., Kumari, K.: 嘲讽性政治推文的情感检测. 国际可持续全球发展计算会议论文集 (INDIACom) (2014). IEEE