Identifying Topical Shifts in Twitter Streams: An Integration of Non-negative Matrix Factorisation, Sentiment Analysis and Structural Break Models for Large Scale Data

Mattias Luber[1], Christoph Weisser[1,2](✉), Benjamin Säfken[1,2], Alexander Silbersdorff[1,2], Thomas Kneib[1,2], and Krisztina Kis-Katos[1,2]

[1] Georg-August-Universität Göttingen, Göttingen, Germany
christoph.weisser@oxon.org
[2] Campus-Institut Data Science (CIDAS), Göttingen, Germany

Abstract. We propose an integration of Non-negative Matrix Factorisation, Sentiment analysis and Structural Break Models to identify significant topical shifts on the social media platform Twitter. For the topic modelling, we compare Latent Dirichlet Allocation and Non-negative Matrix Factorization in terms of their applicability to short text documents. The extraction of sentiment is done by the rule-based VADER model. Structural breaks in the relative frequency and daily sentiments of topics over time are identified with the Bai-Perron model. Combining these methods, we provide a valuable and easy to use exploratory tool for social scientists to study the discourse on Twitter over time. Detecting statistically significant shifts in topics over time enables researchers to perform statistical inference and test hypotheses about the discourse on Twitter. The framework is implemented efficiently to ensure that it can be used on average consumer hardware in a reasonable amount of time. A case study with COVID-19 related tweets in the UK is provided. Our method is validated by linking the topical shifts to real world events by the use of the timestamps of the COVID-19 related tweets.

Keywords: Twitter · Social media · Topic model · Non-negative Matrix Factorisation · Sentiment analysis · Structural Break Models

1 Introduction

For research in the social sciences, the content of discussions on social media and on micro-blogs such as Twitter are highly relevant as they allow to capture shifts in public sentiment and discourse. We provide a user-friendly framework to model the discourse on Twitter by applying natural language processing methods to user-generated Twitter data. We compare Latent Dirichlet Allocation (LDA) [7]

J. Bright et al. (Eds.): MISDOOM 2021, LNCS 12887, pp. 33–49, 2021.
https://doi.org/10.1007/978-3-030-87031-7_3

and Non-negative Matrix Factorization (NMF) [9,19] in terms of their applicability to short text documents and find that the NMF leads to substantially better results. Sentiments are extracted with the rule-based VADER model [12] and shifts in the topical prevalence are identified endogenously with the Bai-Perron model [6]. We provide efficient implementations to ensure that our framework can be used on average consumer hardware in a reasonable amount of time. By combining these methods, we provide a tool for social scientist to study the discourse on Twitter over time. In particular, our approach allows users to detect statistically significant shifts in the discourse of topics.

Our framework is tested in a case study relying on COVID-19 related tweets in the UK. We identify topics that are linked to the context of COVID-19, determine their sentiment, and model their development over time. The results are validated by linking the topical shifts to real world events by the use of the timestamps of the COVID-19 related tweets.

The remainder of the paper is structured as follows. Section 2 discusses related work. Section 3.1 outlines the data collection and pre-processing. Since the data is streamed directly from Twitter, there are specific requirements to be met to enable a time efficient analysis with average consumer hardware. Section 3.2 introduces the sentiment analysis with the VADER model. The estimation of topics models and a comparison of the LDA and NMF is provided in Sect. 3.3. Section 4.3 outlines the Bai-Perron model. Results are presented and linked to real word events in Sect. 4. Section 5 discusses and concludes and provides suggestions for further research.

2 Related Work

Sentiment analysis in general is a well researched topic across multiple domains, but especially in context of social media data and micro-blogs. The proposed methods are quite diverse and range from lexicon-based, over rule-based approaches up to complex deep neural networks [22]. An extensive overview and comparison of different methodologies in particular for Twitter data can be found in [11] or [33].

For topic modelling, several NMF [9,19]- or LDA [7]-based approaches exists. However, not all of them are directly applicable for micro-blog data because of the specific challenges of short and sparse text. There are extensions that take these into account and it has been shown that this can indeed improve the estimates [29]. Despite its potential limitations, we build our framework on the basic configuration of NMF since it provides good results and furthermore is well-researched, robust, and intuitive to understand. More details on this decision can be found in the methodology section. A probabilistic alternative for short text would be the Dirichlet Multinomial Model [28].

On the framework level, current approaches are still often either build heuristically or require manual annotations, especially in the context of socioeconomic analysis. PoliTwi [20] as an example narrows down the concept of topics to single hashtags and then tries to track emerging political topics by simply visualizing their occurrence over time. In a similar manner, Adedoyin-Olowe et al. [1] use automated rule mining to detect events through trending hashtags. Yaqub et al. [27] associate keywords with sentiment scores to gather insight about the US elections 2016. Cases where manual annotations are required [33] are problematic as Twitter data is usually unlabeled and the process of labeling is generally extremely time consuming and not scalable.

Nevertheless, it also has been shown that the connection of sentiment analysis and topic modelling can provide valuable insights for better understanding the discussion on social media [2,4,18] and that topic models can be used to trace the change of events [26,32]. We extend those insights by applying the Bai-Perron model [5,6] for detecting structural changes in the topics discussed on Twitter. The Bai-Perron model is specifically designed to identify multiple structural breaks in a time series, without a-priori information about the break point. Thereby we can not only detect the presence of an event or change points in sentiment, but also provide statistical confidence intervals for their time-points.

In a broader context the modelling of the Twitter discourse is a quite diverse field, which also includes fundamentally different approaches like frame detection [13], network analysis [25] or evaluation of disinformation campaigns[14]. An overview about the current literature on Twitter analysis can be found in Antonakaki et al. [3].

3 Methodology

An overview of the framework is provided in Fig. 1. After collecting tweets via the Twitter-API, the data are pre-processed such that hyperlinks and mentions are removed and the texts were set to lowercase. After that, the workflow splits into two major branches, which are topic modelling and sentiment analysis. Topic modelling is used to identify the most influential topics that are discussed on Twitter and delivers word-clouds for each of them. In parallel to that, the sentiment analysis annotates each tweet with a sentiment score. These two types of outcomes are aggregated into a time-series of daily relative frequencies of topics and their average daily sentiment, Finally, the structural break models are run on each topic separately to detect statistically significant shifts in their frequency and the related sentiments. Finally, the user is provided with valuable information about the topics, their occurrence, as well as with time-points of structural breaks in the frequency of topics as well as in topic-specific sentiment. Each component is explained in more detail in the following sections.

3.1 Data Collection and Preprocessing

The data was collected with the Twitter API in the time period between 25th October 2020 and 14th January 2021 with the python package Tweepy [21] and a geo-location filter for the United Kingdom. Tweepy connects to the Twitter sample stream, which provides access to a random 1% subsample of all tweets published in the given area in real time. At the time point of data collection,

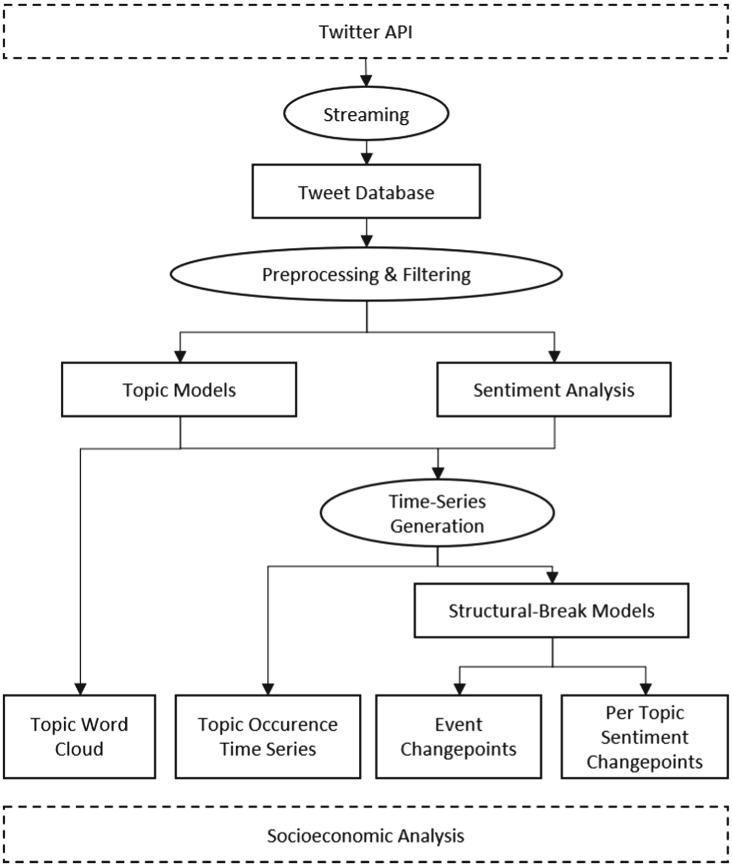

Fig. 1. Visualization of the overall workflow of our framework

no additional filters were applied to ensure that the data is as complete and unbiased as possible. The tweets were accessed as they were published. This was explicitly intended to ensure that later deletions or modifications of tweets would not affect the analysis.

One key difference between the study of Twitter data and classical text mining is that the documents can have fundamentally different properties that have to be taken into account. Users on social media platforms often use a distinct vocabulary that contains more slang, emojis, and hashtags than standard language does. Furthermore, especially on micro-blogs like Twitter, the documents tend to be extremely short.

This gives rise to new challenges as most of the algorithms for text analysis were designed for classical documents and are not guaranteed to work on non-standard language. To compensate for that, only 'extended tweets' were included in the analysis, which are at least 160 characters long. In addition, retweets and replies were excluded to get the true user-generated content only.

After data collection, a lexicon-based filtering is applied to extract the share of tweets that are associated with COVID-19. Only posts are included that contain at least one word of the lexicon in Fig. 2. The lexicon is designed heuristically by identifying words and hashtags that are often used in the context of the COVID-19 pandemic. It contains words and hashtags that are expected to be used neutrally, but also most exclusively within the context of the pandemic. For instance, words like "stay" and "safe" would not be good choices because they are too generic while the hashtag 'staysafe' is nowadays almost exclusively used in association with COVID-19 and can as such be used for filtering. After pre-processing and applying the filtering procedure, about 71.000 COVID-19 related tweets remain. This is about 10% of the original data.

Fig. 2. A wordcloud of the lexicon, which was used for filtering COVID-19 related tweets.

3.2 Sentiment Analysis

The streamed data does not contain true sentiment labels, so that supervised machine learning based models cannot be used. Instead, we implement the rule-based VADER model [12]. The outcome of the prediction is a compound score, ranging on a continuous scale from the most negative (-1) over neutral (0) to the most positive (1).

VADER works with lists of positive and negative words, but in contrast to other rule-based models, it also takes basic grammatical and syntactical features into account. For example, some amplifying words like "very" or "extremely" would not change the sentiment, but increase its intensity. "But" on the other hand, is used to detect sentiment changes. Here the weights of preceding words are lowered while the upcoming words are amplified. The biggest advantage of VADER is probably, that it is designed with social media applications in mind and hence it can deal with common slang phrases and emoticons. Moreover, it does not require any training steps and despite its simplicity, it scores surprisingly well across various domains. The original paper states that correlations up to 0.88 with user ratings could be observed and in a self-conducted evaluation on the SemEval2018 dataset [17] of labeled tweets a correlation of round about 0.7 is reached. This is indeed in a similar range of the more complex models, which score between 0.7 and 0.8.

Sentiment in the context of this analysis does not necessarily reflect the opinion towards a certain topic. It rather gives an indication if the discussion about a certain topic has heated up or is loaded with strong positive or negative emotions.

3.3 Topic Modelling

Topic modelling describes the extraction of underlying topics as hidden semantic structures in text documents. Two algorithms, which are often recommended for this tasks, are Latent Dirichlet Allocation (LDA) [7] and Non-negative-Matrix-Factorization (NMF) [9,19]. Both work with a Bag-of-Words representation of documents, which means that a document is fully defined by its words, while the semantic structure is neglected. LDA models topics as hidden variables and follows a probabilistic approach. By contrast, NMF is essentially a matrix factorization that tries to minimize a reconstruction error.

In the following, the applicability of both algorithms is briefly outlined with particular attention to the special characteristics of micro-blogs and tweets. Since the NMF turned out to work better for the analysis of COVID-19 related tweets, the further evaluation is carried out with this model only.

Latent Dirichlet Allocation. For topic modelling, LDA is often considered the established state of the art since it yields good results and furthermore comes with an implicit measure of uncertainty due to its probabilistic perspective. LDA assumes a data generating process, where each document can be described by a distribution over k latent topics, and each topic is a distribution over the vocabulary. New documents are then generated by repeatedly drawing a topic first and then sampling a word from its distribution.

While LDA performs well on longer texts it shows shortcomings on short and sparse documents such as tweets. Chen et al. [29] conducted a comprehensive exploratory study for various topic modelling algorithms and find that LDA indeed leads to worse results for short and spare text. A reason is the extremely sparse document-feature matrices for short and spare text which makes word-word co-occurrences harder to estimate. Especially for the probabilistic approach this compromises the ability of the model to capture the underlying structure since the estimates are affected by a high variance [29]. Furthermore, one key assumption of LDA is that each document can contain more then one topic. However, due to the short and sparse nature of tweets, it is reasonable to suppose that they are mostly mono-thematic.

To overcome the sparsity, various pooling procedures are proposed, where several tweets are aggregated into pseudo-documents, e.g., based on their hashtags [16]. Within our analysis, such a hashtag-pooling would interfere with the chosen filtering approach since tweets are collected based on certain hashtags like “covid19” or “coronavirus”.

Those hashtags occur in a substantial amount of posts. If hashtags are then used for the aggregation, some extremely large documents are created, while the vast majority is still very short. Besides the highly skewed document length, this results in an unnaturally imbalanced number of topics per document since general purpose hashtags are used with basically every subtopic, while really specific ones are still mono-thematic.

Non-negative-Matrix-Factorization. NMF treats topic modelling as a matrix factorization problem. It approximates the data matrix of observed documents X as the product of a coefficient matrix W and a component matrix H, i.e., $X \approx W * H$. The matrices W and H are derived by minimizing the reconstruction error in respect to the Frobenius norm with the additional constraint that all coefficients have to be non-negative. Documents are hence modelled as linear combinations of the components, and due to the non-negativity of the coefficients, the components become interpretable as topics.

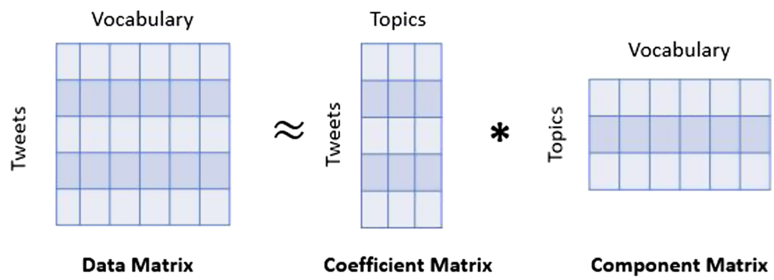

Fig. 3. Top-level visualization of the Non-negative Matrix Factorization algorithm.

Note that NMF also makes the assumption that tweets can contain a mixture of topics. However, we find that NMF performs substantially better than LDA in terms of the extreme sparsity of short mono-thematic documents, since it can be initialized with sparse components [8] and has clustering properties [10,15].

3.4 Time Series Generation

By default, topic coefficients as well as sentiment scores are defined on the level of individual tweets.

To calculate a metric that measures the daily prevalence of a topic over time, all rows from the coefficient matrix of the NMF are normalized, (with the sum of the coefficients in a row) such that the coefficients add up to 1. This makes the scoring comparable between different tweets. After that, a post is said to contain a sufficient amount of a topic i if the corresponding coefficient exceeds a threshold of 0.3. For each topic, the coefficients that fulfill this criterion are aggregated per day to show at which times a certain topic was mostly discussed. Setting this threshold can help to improve the signal in the daily topic prevalence. This is because after the normalization, tweets which can not be assigned to any topic at all end up with approximately uniformly distributed coefficients across all topics. Aggregating coefficients without filtering would therefore lead to a high noise ratio in the time-series caused by unassigned tweets.

This approach was chosen in accordance to the assumptions made by the NMF algorithm, which explicitly allows a document to contain more the one topic. As an example, if any tweet consists of two topics in equal shares, it now contributes to both related topic time series in that ratio. A totally monothematic post would respectively contribute to its time series with its full coefficient.

The sentiment scores are binned into "negative", "neutral" and "positive" and here the fraction of negative tweets within each day is used as a time-series. The series are once derived globally across all tweets and in addition for each topic separately.

3.5 Structural Break Models

The structural break model by Bai-Peron [6] was developed to identify multiple change-points in time series data. The general idea is to fit a linear model for each segment between two change-points and to derive the optimal number and location of the breaks based on the overall residual sum of squares (RSS). Even if the true number of changes would be known in advance (which is not the case in most applications and also not in our analysis) the derivation of the optimal placement would still be computationally expensive, since m breaks and n time-points imply $\binom{n}{k}$ possible combinations. For an unknown number of break-points this gets even more problematic. As a solution the estimation is done by the help of dynamic programming to ensure an efficient implementation [31].

Since this model is essentially approximating the time series by piecewise linear functions, additional breakpoints would always lead to an improvement of the total RSS. To compensate for that, the optimal number of break-points are determined via the Bayesian Information Criterion (BIC) which penalizes additional model complexity. As a result a new break is only introduced, if it is justified by an sufficient improvement in terms of RSS [30].

One of the biggest advantages of the Bai-Peron model is that each breakpoint comes with a statistical confidence interval, which allows a quantification of the uncertainty and hypothesis testing. Furthermore, as the segments between two breakpoints can be interpreted as linear models, the slope and intercept coefficients can be used to quantify the impact and development of structural changes [5].

4 Application on COVID-19 Related Tweets

Since the outbreak of COVID-19 in early 2020, the pandemic has certainly been one of the major topics in almost every aspect of life, but it especially affected public opinion and discourse on healthcare, politics and society. While the effect of measures and restrictions on the spread of the virus can be more or less directly assessed by analysing changes in the number of reported infections, the public perceptions of those measures are harder to capture directly.

Our framework can help to measure the topic-specific public opinion in Twitter data and help to assess the public acceptance if certain policy measures. This is relevant not only for socioeconomic research purposes, but also for pandemic monitoring as acceptance matters for behaviour [23,24].

4.1 Encoding Sentiment

The sentiment for each tweet is predicted with VADER and the resulting compound scores are binned into three classes. Scores between $[-1, -0.3)$ are classified as negative, $[-0.3, 0.3]$ as neutral and $(0.3, 1]$ as positive. Then the fractions of negative as well as positive tweets are calculated per day and tracked over time to visualize how the sentiment developed.

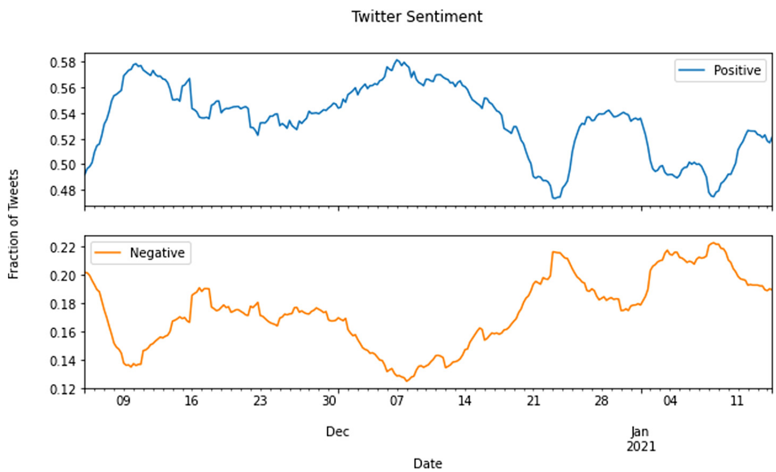

Fig. 4. The fraction of negative and positive Covid-related tweets per day. Especially during December a strong increase in negativity can be observed.

4.2 Establishing Topics

The evaluation of the NMF-based topic model was primarily done by plotting word-clouds of the highest scoring words per topic of the component matrix of the NMF. Twitter is used for commercial purposes to a great extent, and despite the filtering, still quite a lot of advertisements were contained in the data. However, those clustered together in shopping related topics and did not impair the quality of the topics of interest. Some of the identified topics are clearly related to COVID-19, such as the topics "lockdown", "vaccine" or "tier system" as shown in Fig. 5, others reflected some general events that took place during that time, like e.g. New Years Eve and Christmas.

Fig. 5. Selected NMF-topics visualized as wordclouds. Some of them are clearly event related, while others reflect a more continuous discussion about e.g. restrictions in context of COVID-19.

Figure 6 reveals that different topics hold different characteristics. While some are clearly event related and mainly discussed around certain time points, others reflect a more continuous discussion. In context of the evaluation, the time series visualization can be used as a consistency check for the topic model results. In general, the evaluation of topic models is not trivial, since no such thing as a ground truth exists. However, in this case we know the major historical events during the investigated time period. If events can be clearly linked to topics that are related to those events, this shows that the discussion actually indeed centered around those events. For example, if the "New Year" topic would have been discussed at any other time-point then the actual New Year's Eve, it would be highly questionable that the model actually captured the real life concept behind it. For the topics on "Vaccine", "Lockdown" and "Tier System", this evaluation is done in detail in Sect. 4.3 in combination with Bai-Perron models that are used to detect events and structural breaks.

4.3 Event Detection with Structural Break Models

To formalize the process of event detection in the generated time series, the Bai-Perron model is applied with a relatively small trimming factor of 0.05. Thus,

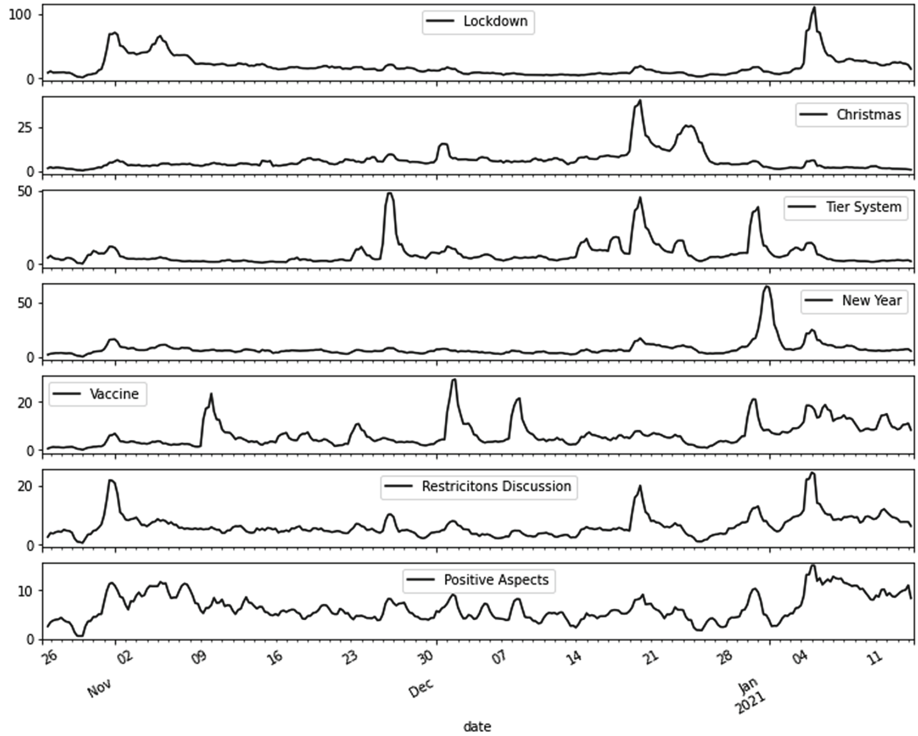

Fig. 6. The occurrence of some selected topics along the timeline. The occurrence is measured by the number of tweets, which had a corresponding normalized coefficient of at least 0.3. The y-axis reflects the summed coefficients per topic for those tweets within a six hour window and is smoothed over a day.

the model is able to frame event-related peaks quite closely and therefore yields good results in regard to their detection.

For the evaluation of the event detection, the breakpoints are of special interest. As mentioned, an in depth investigation is done for the "Vaccine", "Lockdown" and "Tier System" topics, since here the major real-world events during that time are well captured by classical media sources and therefore can be verified.

The complete comparison is listed in the following table. The results show, that indeed for the topics and break-points a clear linkage to a corresponding real-world event can be drawn. For the "Lockdown" topic, the events are centered around the announcement and first implementation of lock-downs. For the "Vaccine" topic, important press releases and vaccine approvals are linked to structural breaks in the topic. For the "Tier System" topic, each major tier level change was detected as a separate event.

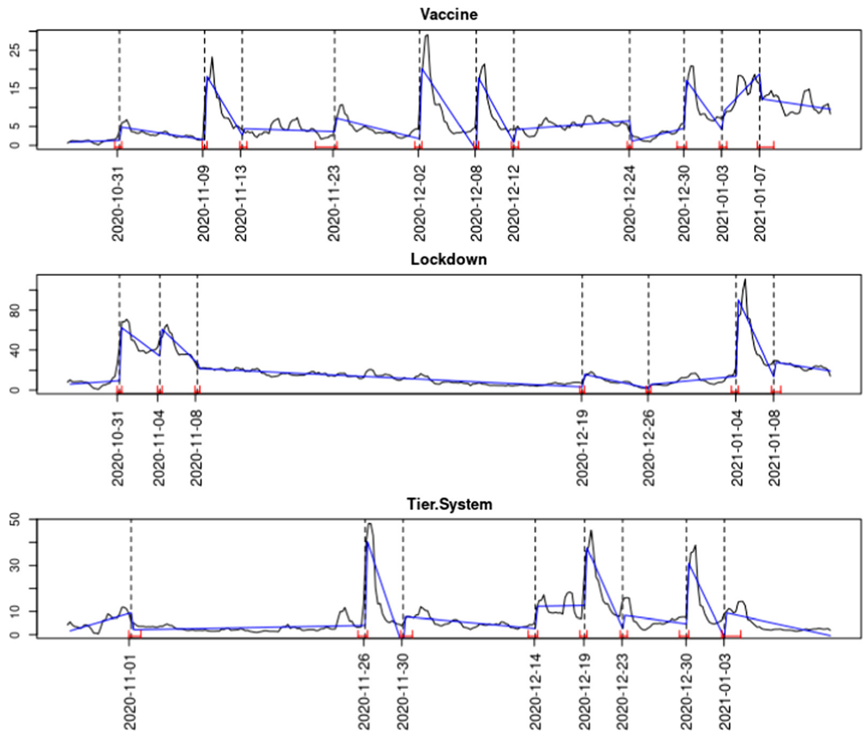

Fig. 7. Event detection for event related topics with Bai-Perron models.

4.4 Detecting Sentiment Changes in User Discussions

To track and identify major change-points in the sentiment of Covid-related discussions, the Bai-Perron model is applied to each topic's sentiment timeline, which represents the fraction of negative tweets per day. Through the integration of topic modelling, sentiment analysis and structural break detection, the framework is now able to not only detect whenever a major shift in the public opinion took place, but also can separate the shifts across different topics. This is useful for the socioeconomic analysis since it can provide a more detailed quantification of the reception of various events within the Twitter active population.

An example can be seen in Fig. 8 where the fraction of negative tweets is displayed for different topics. It can be observed that an event around the 8th to 10th of November lead to an increase in negative tweets for the discussions about politics and 'covidiots', but the sentiment in the restriction discussion remain unchanged. On the other hand, between the 9th and 20th of December, the sentiments for all these topics moved more synchronously, therefore the related events impacted those topics in a similar manner.

Topics	Detected breakpoints	Related real-world events	
"Vaccine"	2020-10-31	2020-10-31	Press conference, Johnson expects vaccine in first quarter of 2021
	2020-11-09 - 2020-11-13	2020-11-09	Pfizer and BioNTech published press release that stated a 90% effectiveness of their candidate
	2020-11-23	2020-11-23	AstraZeneca publishes a press release that stated a 70% efficacy of their candidate
	2020-12-02 - 2020-12-08	2020-12-02	First vaccine against COVID-19 was approved by the Medicines and Healthcare products Regulatory Agency (MHRA) in the United Kingdom
	2020-12-08 - 2020-12-12	2020-12-08	First patient receives a shot of the BioNTech vaccine
	2020-12-24		Christmas
	2020-12-30 - 2021-01-03	2020-12-30	MHRA approves the AstraZeneca vaccine candidate
	2021-12-03 - 2021-01-07	2021-01-04	First patient receives a shot of AstraZeneca's vaccine
	2021-01-07 - END		Wider rollout of the vaccine program
"Lockdown"	2020-12-31 - 2020-11-04	2020-10-31	Boris Johnson announces the second national lockdown
	2020-11-04 - 2020-11-08	2020-11-04	Second national lockdown takes place
	2020-11-08 - 2020-12-19		Long tail of lockdown discussion
	2020-12-19 - 2020-12-26	2020-12-19	Johnson announces that tight restrictions also hold during Christmas
	2021-01-04 - 2021-01-08	2021-01-04	The third lockdown for England and Scotland is announced
"Tier system"	2020-11-01	Late Oct 2020	Various areas reach another tier level within the old 3-level system
	2020-11-26 - 2020-11-30	2020-11-26	Introduction of the new 4-level tier system
	2020-12-14 - 2020-12-19	2020-12-14	London, south and west Essex, and south Hertfordshire are announced to enter Tier 3
	2020-12-19 - 2020-12-23	2020-12-19	Johnson announces that major parts of areas England will enter tier 4
	2020-12-30 - END	2020-12-30	Press conference announces that various other areas are entering tier 4 as well

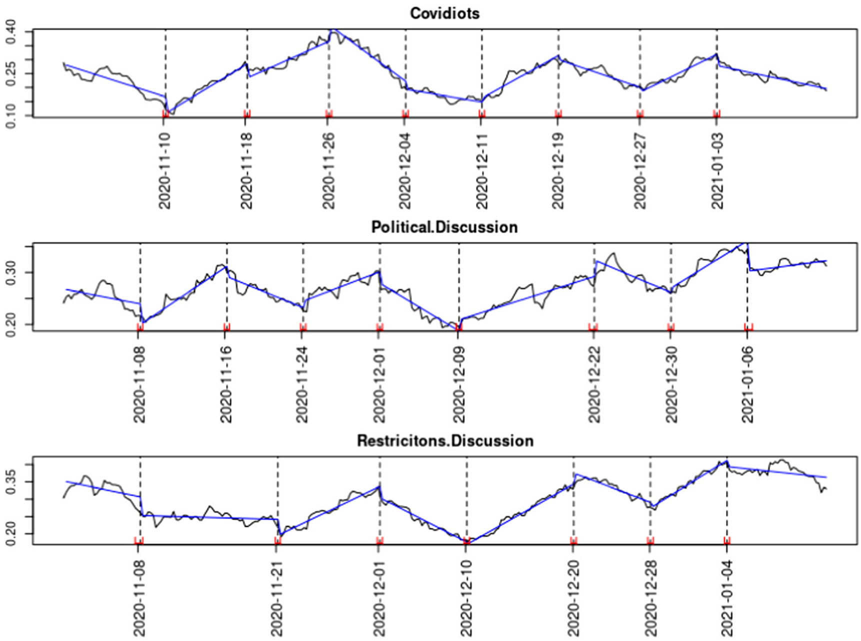

Fig. 8. Detection of sentiment break-point in the daily fraction of negative tweets per topic. An increasing slope reflects an increasing negativity of the sentiment within the topic.

5 Conclusion

Our framework provides an easy to use tool for social scientists to study the discourse on Twitter over time. In particular, it allows the user to detect statistically significant shifts in the sentiment and occurrence of topics. We find that the non-probabilistic NMF algorithm is more suitable for topic modelling on micro-blog document, since it is able to extract topics directly from the original tweets. In contrast, the probabilistic LDA would require the generation of larger pseudo documents via tweet pooling. Furthermore, the properties of NMF in respect to shortness and sparseness are better reflecting the mono-thematic structure of the data.

In a case study about COVID-19 in the UK, we are able extract COVID-19 related topics and their sentiment to gain insights into the discourse during the pandemic. Further, we showed, that with Bai-Perron models the outcomes of the topic models can be used to detect significant shifts in the topic occurrence, that can be matched very well with historical events. In combination with sentiment analysis the same model can help to detect and quantify significant structural breaks in the sentiments per topic.

References

1. Adedoyin-Olowe, M., Gaber, M.M., Dancausa, C.M., Stahl, F., Gomes, J.B.: A rule dynamics approach to event detection in twitter with its application to sports and politics. Expert Syst. Appl. **55**, 351–360 (2016)
2. Andry, A., Wirawan, R., Adhi, N.D.D., Farhan, R., Siti, S.: Dynamic large scale data on twitter using sentiment analysis and topic modeling. In: 2018 6th International Conference on Information and Communication Technology (ICoICT), pp. 254–258 (2018). https://doi.org/10.1109/ICoICT.2018.8528776
3. Antonakaki, D., Fragopoulou, P., Ioannidis, S.: A survey of twitter research: data model, graph structure, sentiment analysis and attacks. Expert Syst. Appl. **164**, 114006 (2021)
4. Bahja, M., Lycett, M.: Identifying patient experience from online resources via sentiment analysis and topic modelling. In: Proceedings of the 3rd IEEE/ACM International Conference on Big Data Computing, Applications and Technologies, BDCAT 2016, New York, NY, USA, pp. 94–99. Association for Computing Machinery (2016). https://doi.org/10.1145/3006299.3006335
5. Bai, J., Perron, P.: Estimating and testing linear models with multiple structural changes. Econometrica **66**(1), 47–78 (1998)
6. Bai, J., Perron, P.: Computation and analysis of multiple structural change models. J. Appl. Economet. **18**(1), 1–22 (2003). https://doi.org/10.1002/jae.659
7. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. J. Mach. Learn. Res. **3**(null), 993–1022 (2003)
8. Boutsidis, C., Gallopoulos, E.: SVD based initialization: a head start for nonnegative matrix factorization. Pattern Recogn. **41**(4), 1350–1362 (2008)
9. Févotte, C., Idier, J.: Algorithms for nonnegative matrix factorization with the beta-divergence. CoRR abs/1010.1763 (2010)
10. Ding, C., He, X., Simon, H.D.: On the equivalence of nonnegative matrix factorization and spectral clustering. In: Proceedings of the 2005 SIAM International Conference on Data Mining (SDM), pp. 606–610. https://doi.org/10.1137/1.9781611972757.70
11. Giachanou, A., Crestani, F.: Like it or not: a survey of twitter sentiment analysis methods. ACM Comput. Surv. **49**(2) (2016). https://doi.org/10.1145/2938640
12. 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, vol. 8, no. 1 (2014). https://ojs.aaai.org/index.php/ICWSM/article/view/14550
13. Johnson, K., Jin, D., Goldwasser, D.: Modeling of political discourse framing on twitter. In: Proceedings of the International AAAI Conference on Web and Social Media, vol. 11, no. 1, May 2017. https://ojs.aaai.org/index.php/ICWSM/article/view/14958
14. Keller, F.B., Schoch, D., Stier, S., Yang, J.: Political astroturfing on twitter: how to coordinate a disinformation campaign. Polit. Commun. **37**(2), 256–280 (2020)
15. Lu, H., Fu, Z., Shu, X.: Non-negative and sparse spectral clustering. Pattern Recogn. **47**(1), 418–426 (2014)

16. Mehrotra, R., Sanner, S., Buntine, W., Xie, L.: Improving LDA topic models for microblogs via tweet pooling and automatic labeling. In: Jones, G.J., Sheridan, P., Kelly, D., de Rijke, M., Sakai, T. (eds.) Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, New York, NY, USA, pp. 889–892. ACM (2013). https://doi.org/10.1145/2484028.2484166
17. Mohammad, S.M., Bravo-Marquez, F., Salameh, M., Kiritchenko, S.: SemEval-2018 task 1: affect in tweets. In: Proceedings of International Workshop on Semantic Evaluation (SemEval-2018), New Orleans, LA, USA (2018)
18. Patil, P.P., Phansalkar, S., Kryssanov, V.V.: Topic modelling for aspect-level sentiment analysis. In: Kulkarni, A.J., Satapathy, S.C., Kang, T., Kashan, A.H. (eds.) Proceedings of the 2nd International Conference on Data Engineering and Communication Technology. AISC, vol. 828, pp. 221–229. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-1610-4_23
19. Pedregosa, F., et al.: Scikit-learn: machine learning in Python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
20. Rill, S., Reinel, D., Scheidt, J., Zicari, R.V.: PoliTwi: early detection of emerging political topics on twitter and the impact on concept-level sentiment analysis. Knowl.-Based Syst. **69**, 24–33 (2014)
21. Roesslein, J.: Tweepy: Twitter for Python! (2020). https://github.com/tweepy/tweepy
22. Severyn, A., Moschitti, A.: Twitter sentiment analysis with deep convolutional neural networks. In: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2015, New York, NY, USA, pp. 959–962. Association for Computing Machinery (2015). https://doi.org/10.1145/2766462.2767830
23. Siegrist, M., Luchsinger, L., Bearth, A.: The impact of trust and risk perception on the acceptance of measures to reduce COVID-19 cases. Risk Anal. (2021). https://doi.org/10.1111/risa.13675
24. Siegrist, M., Zingg, A.: The role of public trust during pandemics. Eur. Psychol. **19**(1), 23–32 (2014). https://doi.org/10.1027/1016-9040/a000169
25. Soares, F.B., Recuero, R., Zago, G.: Influencers in polarized political networks on twitter. In: Proceedings of the 9th International Conference on Social Media and Society, SMSociety 2018, New York, NY, USA, pp. 168–177. Association for Computing Machinery (2018). https://doi.org/10.1145/3217804.3217909
26. Suri, P., Roy, N.R.: Comparison between LDA & NMF for event-detection from large text stream data. In: 2017 3rd International Conference on Computational Intelligence and Communication Technology (CICT), pp. 1–5. IEEE (09022017-10022017). https://doi.org/10.1109/CIACT.2017.7977281
27. Yaqub, U., Chun, S.A., Atluri, V., Vaidya, J.: Analysis of political discourse on twitter in the context of the 2016 US presidential elections. Gov. Inf. Q. **34**(4), 613–626 (2017)
28. Yin, J., Wang, J.: A Dirichlet multinomial mixture model-based approach for short text clustering. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 233–242 (2014)
29. Chen, Y., Zhang, H., Liu, R., Ye, Z., Lin, J.: Experimental explorations on short text topic mining between LDA and NMF based schemes. Knowl.-Based Syst. **163**, 1–13 (2019)
30. Zeileis, A., Kleiber, C., Krämer, W., Hornik, K.: Testing and dating of structural changes in practice. Comput. Stat. Data Anal. **44**, 109–123 (2003)

31. Zeileis, A., Leisch, F., Hornik, K., Kleiber, C.: strucchange: an R package for testing for structural change in linear regression models. J. Stat. Softw. **7**(2), 1–38 (2002). http://www.jstatsoft.org/v07/i02/
32. Zhou, X., Chen, L.: Event detection over twitter social media streams. VLDB J. **23**(3), 381–400 (2013). https://doi.org/10.1007/s00778-013-0320-3
33. Zimbra, D., Abbasi, A., Zeng, D., Chen, H.: The state-of-the-art in twitter sentiment analysis: a review and benchmark evaluation. ACM Trans. Manage. Inf. Syst. **9**(2) (2018). https://doi.org/10.1145/3185045