REGULAR PAPER

Kostiantyn Kucher · Rafael M. Martins · Carita Paradis · Andreas Kerren

StanceVis Prime: visual analysis of sentiment and stance in social media texts

Received: 12 February 2020 / Revised: 25 May 2020 / Accepted: 16 June 2020 / Published online: 25 August 2020
© The Author(s) 2020

Abstract Text visualization and visual text analytics methods have been successfully applied for various tasks related to the analysis of individual text documents and large document collections such as summarization of main topics or identification of events in discourse. Visualization of sentiments and emotions detected in textual data has also become an important topic of interest, especially with regard to the data originating from social media. Despite the growing interest in this topic, the research problem related to detecting and visualizing various stances, such as rudeness or uncertainty, has not been adequately addressed by the existing approaches. The challenges associated with this problem include the development of the underlying computational methods and visualization of the corresponding multi-label stance classification results. In this paper, we describe our work on a visual analytics platform, called StanceVis Prime, which has been designed for the analysis of sentiment and stance in temporal text data from various social media data sources. The use case scenarios intended for StanceVis Prime include social media monitoring and research in sociolinguistics. The design was motivated by the requirements of collaborating domain experts in linguistics as part of a larger research project on stance analysis. Our approach involves consuming documents from several text stream sources and applying sentiment and stance classification, resulting in multiple data series associated with source texts. StanceVis Prime provides the end users with an overview of similarities between the data series based on dynamic time warping analysis, as well as detailed visualizations of data series values. Users can also retrieve and conduct both distant and close reading of the documents corresponding to the data series. We demonstrate our approach with case studies involving political targets of interest and several social media data sources and report preliminary user feedback received from a domain expert.

Keywords Stance visualization · Sentiment visualization · Text visualization · Stance analysis · Sentiment analysis · Opinion mining · Visualization · Interaction · NLP · Text analytics

---

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s12650-020-00684-5>) contains supplementary material, which is available to authorized users.

K. Kucher (✉) · R. M. Martins · A. Kerren
Department of Computer Science and Media Technology, Linnaeus University, Växjö, Sweden
E-mail: kostiantyn.kucher@lnu.se

C. Paradis
Centre for Languages and Literature, Lund University, Lund, Sweden

1 Introduction

The recent years have demonstrated how massively available digital communication channels, such as social media, affect world politics and shape the agenda in multiple spheres of life. The understanding of phenomena occurring in the corresponding data is therefore interesting and important for decision makers, researchers, and the general public. Some of the most interesting aspects of human communication to analyze in such data are related to various expressions of subjectivity in social media document texts, such as sentiments, opinions, and emotions (Pang and Lee 2008; Mohammad 2016; Zhang et al. 2018). The analysis of stance-taking in texts (Englebretson 2007) can provide even further insights about the subjective position of the speaker, for instance, AGREEMENT or DISAGREEMENT with a certain topic (Chen and Ku 2016; Mohammad et al. 2016, 2017), or expression of CERTAINTY and PREDICTION (Simaki et al. 2017).

However, the manual analysis of texts and the manual examination of raw output of computational text analyses do not scale up to the amount of data produced by social media, which can range from hundreds to millions of messages per day, depending on the topic or target of interest. Besides the traditional close reading task, support for distant reading (Jänicke et al. 2015) is required to make sense of such data. Information visualization and visual analytics approaches have been therefore applied successfully to address this challenge for social media data (Chen et al. 2017). In particular, text visualization (Kucher and Kerren 2015) and visual text analytics (Liu et al. 2019) methods can support various tasks related to the analysis of individual text documents and large document collections such as summarization of main topics or identification of events in discourse (Dou and Liu 2016), using the techniques developed for time-oriented data visualization (Aigner et al. 2011), when necessary.

Visualization of sentiments and emotions detected in textual data has also become an important topic of interest in the visualization community (Kucher et al. 2018a). Multiple existing *sentiment visualization* techniques address the tasks of visualizing polarity (Cao et al. 2012), opinions (Wu et al. 2014, 2010), and emotions (Zhao et al. 2014) detected in text data. However, the related task of *stance visualization* has not enjoyed the same level of support by the existing approaches so far. One of the main challenges of stance visualization compared to sentiment visualization is related to the need to accommodate the visual design to (1) a specific definition of stance and (2) a data format produced by (3) a specific computational method, which are all decided on by the users (for instance, domain experts in computational linguistics). Several existing techniques support stance visualization for temporal text data to a certain extent; however, they either make use of a very limited set of stance categories/aspects (Diakopoulos et al. 2014; El-Assady et al. 2016; Kucher et al. 2016), or interpret such categories as mutually exclusive (Martins et al. 2017). Visualization of multiple non-exclusive stance categories at the same time (e.g., produced as an output of a multi-label rather than multi-class classification task) for social media texts is, thus, still an open challenge.

Fig. 1 Overview of our approach: (1) the data from several streaming text sources is tracked according to the list of targets of interest and saved into a database; (2) the retrieved documents are processed with sentiment and stance classifiers at the utterance/sentence level; (3) the classification result counts are used to produce multiple data series; and (4) the user is provided with means for visual analysis of text and time data via interactive visual representations

In this paper [based on our previous poster abstract, see Kucher et al. (2018b)], we present our work on a visual analytics platform, called StanceVis Prime, which has been designed to support visual analysis of sentiment and stance in temporal text data from social media. Our approach was developed in collaboration with domain experts in linguistics as part of a larger research project on stance analysis titled StaViCTA. Thus, our work follows the definition of stance categories defined by the linguists and uses a custom classifier for 12 stance categories developed by the collaborating experts in computational linguistics.

Compared to the existing works on stance visualization, our approach aims to support the following (see Fig. 1):

- consumption of data from several social media sources;
- classification of both sentiment polarity and multiple non-exclusive stance categories at the utterance/sentence level;
- visual analysis of data series for various sentiment and stance categories at multiple levels of granularity, including both values and similarity between the series; and
- support for distant and close reading of corresponding text document sets augmented with multi-label classification results, including export of document lists processed and annotated by the users.

The main scientific contributions of this paper are the following:

- an analysis of the workflow and user tasks related to visual analysis of sentiment and stance in social media texts;
- a design study involving temporal and textual data analysis methods in order to represent sentiment and stance; and
- a visual analytics solution supporting exploratory analysis of sentiment and stance classification results for temporal text data from multiple sources.

The rest of this paper is organized as follows. In the next section, we briefly introduce the necessary background information on sentiment and stance analysis, and then we discuss the related work relevant to our text and time data visualization approaches. Section 3 presents the analysis of the workflow and user tasks guiding our design. We briefly describe the overall architecture of our implementation in Sect. 4 and then discuss our design decisions for the visualization components in Sect. 5. We demonstrate our approach with several case studies in Sect. 6 and discuss preliminary user feedback as well as some other aspects of our work in Sect. 7. Finally, we conclude this paper and outline the directions for future work in Sect. 8.

2 Background and related work

In order to highlight the existing approaches for sentiment and stance visualization, we should at first briefly discuss the underlying computational methods. We also outline the existing time-varying data visualization techniques relevant to our visual analysis workflow.

2.1 Sentiment and stance analysis

Sentiment analysis is a well-established task in computational linguistics, which has traditionally been applied for the analysis of customer reviews and, later, in social media analytics. The existing surveys (Pang and Lee 2008; Mohammad 2016; Zhang et al. 2018) describe a multitude of existing approaches for sentiment classification at various levels of granularity, from words to complete documents, and with various categories, from POSITIVE/NEUTRAL/NEGATIVE to multidimensional emotion models. For our purposes, the detection of POSITIVE and NEGATIVE sentiment at the level of utterances/sentences in written text data is sufficient. We have used an existing rule-based classifier called VADER (Hutto and Gilbert 2014) designed for short social media texts in our work. We use the information about utterances with normalized POSITIVITY and/or NEGATIVITY scores above a certain threshold (currently set to 0.3), and consider the rest as NEUTRAL with regard to sentiment. Our approach is not specific to VADER, though, and could use a better performing deep learning classifier (Lai et al. 2015; Zhang et al. 2018) in the future to achieve better classification quality/performance.

In contrast to sentiment, stance-taking has been studied in linguistics (Biber and Finegan 1989; Englebretson 2007; Glynn and Sjölin 2014) with regard to the subjective position of the speaker that might not imply POSITIVE/NEGATIVE polarity or other emotions. The existing computational approaches usually focus on AGREEMENT or DISAGREEMENT on a certain topic (Chen and Ku 2016; Mohammad et al. 2016, 2017), and

only a few works take a wider view of stance aspects/categories into account, such as NECESSITY and VOLITION (Simaki et al. 2017). In this work, we follow the approach taken in our interdisciplinary project, where the researchers in linguistics defined the stance categories of interest (see Table 1) and the experts in computational linguistics implemented a custom stance classifier (Skeppstedt et al. 2016b, 2017). Classification is carried out at the utterance level in a multi-label fashion, i.e., one utterance can be labeled with multiple stance categories simultaneously. Classification results are used for further analysis and visualization.

2.2 Sentiment and stance visualization

Sentiment and stance visualization problems can be treated as part of a more general text visualization (Kucher and Kerren 2015) field. The existing techniques addressing its various tasks and aspects are covered by several surveys, including the works on topic- and time-oriented visual text analytics (Dou and Liu 2016), techniques supporting close and distant reading (Jänicke et al. 2015), social media visual analytics (Chen et al. 2017), and visual text mining (Liu et al. 2019). More specifically, the existing sentiment visualization techniques are discussed in a recent survey (Kucher et al. 2018a), including, among others, multiple approaches that support sentiment visualization in temporal text data: for example, Whisper (Cao et al. 2012) visualizes polarity of tweets as part of a real-time Twitter monitoring task; OpinionFlow (Wu et al. 2014) indicates the polarity of public opinions on particular topics; PEARL (Zhao et al. 2014) represents the emotions detected in tweets by an individual user over time with a stacked graph; SocialHelix (Cao et al. 2015) focuses on the visual analysis of sentiment divergence between the user communities extracted from Twitter data, represented visually with a helix-like visual metaphor; and IdeaFlow (Wang et al. 2016a) supports visualization of sentiment towards specific ideas, which are identified in temporal text data and represented with flow maps.

Several approaches are also relevant to the stance visualization task, e.g., Lingoscope (Diakopoulos et al. 2014) visualizes the language use of *acceptors* and *sceptics* of a certain topic in blogs. ConToVi (El-Assady et al. 2016) supports visualization of debate transcripts using categories beyond sentiment, such as CERTAINTY, ELOQUENCE, and POLITENESS. Our previous approach uVSAT (Kucher et al. 2016) focuses on data series based on the markers of sentiment, emotions, and stance categories such as CERTAINTY and UNCERTAINTY in blogs and forums. All of these works make use of only a limited number of stance-related categories. ALVA (Kucher et al. 2017) overcomes this limitation, but focuses on data annotation and classifier training stages rather than on analyses of data sets collected from social media over time. Another relevant approach is StanceXplore (Martins et al. 2017), which supports visualization of multiple stance categories detected in social media data, however, (1) it treats such categories in a mutually exclusive way for visualization purposes, (2) it does not support sentiment analysis and visualization, and (3) it is limited to the data from a single data source (Twitter). Regarding this last limitation, support for visual comparison of text data from several data sources was previously demonstrated by the approaches such as TopicPanorama (Wang et al. 2016b) for the purposes of topic analysis, but not for stance analysis. In contrast to the existing works, our contribution discussed in this paper is designed specifically for visual analysis of multiple non-exclusive stance categories as well as sentiment categories in the data from several social media sources.

2.3 Visualization of time-varying data

One of the tasks of our work is to represent the change of public sentiment and stance over time (Aigner et al. 2011). ThemeRiver (Havre et al. 2000) is a classic example of temporal text data visualization, which uses a stacked area graph metaphor (Byron and Wattenberg 2008) to represent a flow of themes/topics

Table 1 Data categories used in our work

Type	Categories	Description
General	Document Count	#documents detected in the source data for a target of interest
Sentiment	POSITIVITY, NEGATIVITY	#occurrences of sentiment in utterances detected with VADER (Hutto and Gilbert 2014)
Stance	AGREEMENT, CERTAINTY, CONCESSION & CONTRARIENESS, CONTRAST, DISAGREEMENT, HYPOTHETICALS, NEED & REQUIREMENT, PREDICTION, RUDENESS, SOURCE OF KNOWLEDGE, TACT, UNCERTAINTY	#occurrences of stance in utterances detected with a custom stance classifier (Skeppstedt et al. 2016b, 2017)

detected in text data. In a similar fashion, Dörk et al. (2010) use a stacked graph to represent topics in the streaming Twitter data, and TextFlow (Cui et al. 2011) combines a stacked graph with glyphs and line plots to represent topic evolution. RankExplorer (Shi et al. 2012) uses a stacked graph to support exploration of rank changes over time in sets of time series. MultiStream (Cuenca et al. 2018) adapts ThemeRiver to support interactive exploration of hierarchies of multiple time series using a nonlinear time axis. Our work also uses a stacked graph metaphor augmented with additional cue labels representing sentiment and stance data, as discussed in Sect. 5.3.

The analysis of multiple data series may focus on the task of finding similarities in how the series develop through time. In such cases, similarity computations themselves generate sequential data that must be analyzed, and the main patterns of interest are how similarities between pairs or groups of data series change and how their relationships evolve through time. Storyline techniques, such as the ones discussed by Tanahashi and Ma (2012), Liu et al. (2013), and Silvia et al. (2016) visualize each entity in the data as a timeline that converges (and diverges) with other timelines during periods of more (or less) interaction. Storygraph by Shrestha et al. (2013) also represents the actors' dynamics as timelines using the horizontal time axis, but it focuses on actual spatial distances and movement instead of abstract similarities. Further possibilities include the approach from SocialHelix (Cao et al. 2015) that highlights the divergence of a pair of data series by using a helix-like layout with two strands—however, that visualization approach was only demonstrated for two data series at a time. In order to represent lead-lag relationships between multiple data objects over time, IdeaFlow (Wang et al. 2016a) uses flow map representations computed by laying out directed acyclic graphs and then filtering, clustering, and smoothing the resulting visual representation. However, this approach requires a number of complex computational steps to prepare the input lead-lag data. While adapting this approach for our task of visual analysis of multiple sentiment and stance data might be considered as part of future work, we currently focus on a different solution discussed below instead.

Dimensionality reduction (DR) is an effective technique for the visualization of similarities between groups of entities, and it has also been applied in the context of time-varying data sets. In the more usual scenario, where time-varying multidimensional data are projected into 2D and visualized with interactive scatterplots, possible approaches are to project all time steps at once, then visualize time with trajectories [as in the work by Bernard et al. (2012)]; or to generate one such 2D projection per time step, then present them in a sequence, as done by Alencar et al. (2012) and Rauber et al. (2016). Another solution is to project time steps into 1D, and use one axis (usually the horizontal one) for time. With temporal multidimensional scaling (TMDS), Jäckle et al. (2016) achieve this by arranging (but not explicitly connecting) sequences of 1D MDS projections using the horizontal axis. Crnovrsanin et al. (2009) propose a similar approach, this time with line segments connecting the time steps of the same series, but, as with Storygraph (Shrestha et al. 2013), this technique focuses on the spatial movement of actors.

Two of our views discussed in Sect. 5.2 are related to these works, with further customizations both in the underlying algorithm and the presentation. In contrast to the previous work, we use an efficient implementation (Martins and Kerren 2018) of dynamic time warping (DTW) (Berndt and Clifford 1994; Esling and Agon 2012) in order to compute meaningful distances between the data series, which are then used as input for a DR technique that generates 1D and 2D projections at each time step.

3 Requirements analysis

The work described in this paper was carried out as the last stage of an interdisciplinary research project titled StaViCTA, which was dedicated to stance analysis of written text data. Besides the researchers in visualization, project members included researchers in linguistics and computational linguistics. These domain experts provided the definition of stance categories used in our work and developed a multi-label stance classifier for 12 categories listed in Table 1. As discussed in Sect. 2.1, our collaborators followed an approach rooted in the research on discourse analysis and stance-taking in linguistics (Simaki et al. 2017) in order to select a variety of fine-grained aspects of stance, which were used as the foundation for the stance categories studied within StaViCTA. Our previous visualization contributions within the project supported the stages of data collection (Kucher et al. 2016) as well as data annotation and active learning for training a stance classifier (Kucher et al. 2017), among other tasks. As research and development of several versions of the stance classifier were reaching the completion within the scope of the project (Skeppstedt et al. 2016a, b, 2017), our next goal was set to applying our stance classifier—complemented with a sentiment

classifier implementation—to large volumes of real-world text data retrieved from streaming data sources. In order to help users make use of the massive resulting classification data and extract new knowledge (Sacha et al. 2014) about the sentiment and stance usage in online social media, we have designed the visual analytic approach described in this paper. Thus, multiple discussions with our project members and our earlier experiences in the project have laid the foundation for the overall design of our StanceVis Prime approach, which is aimed to support exploratory analysis (Tukey 1977) of sentiment and stance in temporal text data.

Based on interviews with data analysis practitioners, Alspaugh et al. (2019) describe some of the core exploratory activities as (1) searching for new interesting phenomena, (2) comparing the data to the existing understanding, and (3) generating new analysis questions or hypotheses. These activities in general fit the overall requirements provided by our domain experts, who wanted to use an interactive visual analytics tool to access and explore large amounts of data retrieved from social media.

The overview of the workflow expected from our implementation is illustrated in Fig. 2: users would (1) start from the aggregated temporal data, (2) identify and select interesting time ranges, (3) retrieve and study the corresponding sets of documents, and then (4) focus on individual documents. Interested users should also be able to (5) export filtered and annotated sets of documents for further offline investigation with a focus on close reading (Jänicke et al. 2015) and for presentation/dissemination purposes (Pirulli and Card 2005; Chen et al. 2018). This workflow allows the users interested mostly in textual data to reach the resulting document texts in a rather straightforward way [“simple is good”—Russell (2016)]. Other users who are more interested in analyzing trends in social media could focus mostly on temporal data exploration using coordinated multiple views (Roberts 2007). The concrete list of user tasks for this workflow is as follows:

- T1** Investigate temporal trends in social media data from several data sources/domains for several targets of interest.
- T2** Investigate temporal trends with regard to sentiment and stance data series.
- T3** Investigate similarities of multiple data series over time.
- T4** Retrieve underlying documents for the specified time ranges for sets of targets/domains.
- T5** Summarize document sets with regard to text contents and sentiment & stance.
- T6** Engage in close reading of the classified text documents.
- T7** Export document lists for further offline investigation.

Fig. 2 The visual analysis workflow for StanceVis Prime according to the requirements discussed with the domain experts. The user should be able to conduct interactive analysis of temporal and textual data using the system and export certain textual data for further offline investigation outside of the scope of the system (denoted with a gray dashed edge)

4 Architecture

The backend of StanceVis Prime is designed around a data collection service implemented in Python which consumes text data streams from several data sources (see Fig. 1). Our implementation currently supports Twitter and Reddit; however, it could be easily extended to other data sources in the future. There are multiple targets of interest to be tracked, each defined by a list of key terms and phrases. In addition, the Reddit data stream is configured to include only comments from subforums relevant to our targets of interest. The current list of tracked targets focuses mostly on several key political actors, movements, and events, as these targets are interesting for our domain experts, for instance, *European Politics* and *Brexit*. The retrieved text documents are saved into MongoDB and put on the queue for classification.

Our implementation currently uses the VADER sentiment classifier (Hutto and Gilbert 2014) and a custom logistic regression-based (Hosmer et al. 2013) stance classifier (Skeppstedt et al. 2016b, 2017) developed with scikit-learn (Pedregosa et al. 2011) for 12 stance categories listed in Table 1. The stance classifier was trained on the text data that was (1) collected from blogs and forums, (2) annotated at the utterance/sentence level as part of the StaViCTA project (Simaki et al. 2017), and (3) later augmented with token-level annotations by our collaborators (Skeppstedt et al. 2016a) in order to improve the classification performance compared to the initial utterance-level classification. The data features used for stance classification are based on vector representations of individual tokens and their neighbors (context words) as well as their clustering (Skeppstedt et al. 2016b). Our collaborators chose the logistic regression approach as a standard, well-known alternative that provides quantifiable output for the classification decisions, thus contributing to the interpretability and eventual trust of the end users (Chatzimparmpas et al. 2020).

In the future, we could replace the classifiers used for specific subsets of categories, for instance, by using a deep learning-based classifier for AGREEMENT/DISAGREEMENT (Chen and Ku 2016). Alternatively, we could consider applying ensemble learning techniques (Sagi and Rokach 2018) to improve the classification performance. The classification occurs at the level of individual utterances/sentences, and the information about each detected sentiment and stance category is saved into the database alongside the documents. The stance classifier also reports classification decision confidence, and VADER reports normalized valence values for POSITIVITY and NEGATIVITY, which are also saved into the database.

The utterance classification results for each combination of target, domain, and category (e.g., *Brexit/Reddit/ PREDICTION*) are used to create the corresponding data series at the granularity of one second and several levels of aggregation (minute, hour, day). The document counts for each target/domain are also saved as data series to allow the users investigate all of the retrieved text documents, even including the ones with no sentiment or stance detected. Thus, each data series entry consists of the *timestamp*, the temporal *granularity* level (e.g., one second), the *number* of detected category cases (sentiment, stance, or document count) for this timestamp, and the corresponding classification *confidence* level (for sentiment and stance categories).

Our visualization frontend comprises a web-based application served with Flask, and it is implemented in JavaScript with D3 (Bostock 2011) and Rickshaw (Shutterstock Images 2011). Its components are discussed next.

5 Visualization methodology

In this section, we are going to discuss the design of visual representations used in StanceVis Prime for various parts of the workflow discussed above. We start with some general considerations affecting multiple views (cf. Fig. 3) and then discuss the specific data processing and representation concerns in detail. The supplementary material (Online Resource 1) includes a video demonstration of the visualization frontend of StanceVis Prime.

5.1 General design considerations

Our visualization approach was required to support multiple data series and text document representations associated with the classification results for sentiment and stance, as discussed in Sect. 4. This presented us with multiple design challenges, including the color coding consistency across multiple coordinated views (Roberts 2007). More specifically, the user could desire to investigate the data for N combinations of targets of interests and data domains, for instance, *Brexit/Reddit*, *Brexit/Twitter*, *Vaccination/Reddit*, etc. Each of

Fig. 3 The user interface in StanceVis Prime: **a** the loaded data series table; **b** data series similarity views; **c** the document count graph; **d** bar charts with data series values; and **e** document list views

such combinations is associated with one data series based on the retrieved document count, two series for sentiment categories, and twelve series for stance categories (cf. Table 1), which means that the user would typically work with dozens of data series simultaneously, and it would not be possible to encode each series with a unique hue. The sentiment categories of POSITIVITY and NEGATIVITY are usually encoded by green (or blue) and red hues by the existing sentiment visualization techniques (Kucher et al. 2018a), and our design had to incorporate this fact to match the user expectations, too.

The current color coding approach in our implementation handles targets/domains and data categories in an orthogonal fashion. First of all, each target is assigned with a unique hue using a qualitative scheme from ColorBrewer (Brewer et al. 2009), and the concrete target/domain combinations are then encoded with darker or brighter variations of that color. These colors are used in the visual representations of data series to facilitate the comparison tasks **T1–T3** between various targets/domains (see Sects. 5.2 and 5.3). Investigation of alternative encodings with, e.g., pattern/texture instead of color, is part of our future work. As for the data categories, we have assigned green and red colors to the respective sentiment categories, blue to all stance categories, and gray to the document count series. These encodings are used in the views described in the following subsections.

The visualization interface of StanceVis Prime is displayed in Fig. 3, with the panels (a–d) visible initially after the user logs in and selects the data to be loaded using a dialog. The table in Fig. 3a acts as a legend for the colors associated with each target/domain, and it also enumerates the corresponding data series with their abbreviated titles and sparkline plots (Tuft 2006). The category labels are also used in other parts of the interface to avoid the introduction of complicated glyphs for multiple data series. After loading the data, the user can proceed to data series exploration.

5.2 Representation of data series similarity

As described above, the data initially loaded by the user in StanceVis Prime consists of C data series (currently, 1 for document count + 2 for sentiment + 12 for stance) for each of N target/domain combinations. The number of visualized data series quickly becomes overwhelming for task **T3**. This was our motivation for introducing separate views with a focus on *similarity* rather than *values*. The corresponding pipeline is displayed in Fig. 4.

Fig. 4 The pipeline for data series *similarity* analysis and visualization in StanceVis Prime

The processing starts by computing dynamic time warping [DTW, see Berndt and Clifford (1994), Esling and Agon (2012)] distances between the pairs of data series over the available time range. We use a custom DTW implementation with a sliding window (Martins and Kerren 2018), and the size of the window is currently set to 30% of the data series length. It is then possible to analyze the distances between the series at each time step (temporal slice). Currently, we employ the multidimensional scaling [MDS, see Borg and Groenen (2005)] method to compute 1D and 2D projections. We have chosen MDS as a standard and reliable technique for this task, but it would be possible to extend our approach with other methods. Since MDS is invariant to transformations such as rotation, our next step is to align the projections from consecutive time steps. We have applied Procrustes analysis (Krzanowski 2000), a technique for detecting the optimal superimposition between a pair of data sets, to align subsequent temporal slices of both 1D and 2D projections in a consistent way.

The resulting 2D projection can be visualized with a standard scatterplot representation one temporal step at a time, as displayed in Fig. 5d, controlled by a slider displayed in Fig. 5c. 1D projection results could be represented over time with a line chart, for instance; however, such attempts result in a severely cluttered view. Therefore, we have decided to forgo the idea of using the exact 1D-projected values and instead drew inspiration from rank-based (Shi et al. 2012) and storyline-based (Tanahashi and Ma 2012; Silvia et al. 2016) representations. For each temporal step, we have computed a clustering of the 1D projections using DBSCAN (Ester et al. 1996). The algorithm is configured to detect clusters containing at least three items; the items not belonging to any cluster are labeled as outliers. To compute the resulting layout for this temporal step, we (1) order the 1D projection items by ascending value and (2) assign increasing y coordinates while (3) ensuring vertical space intervals proportional to intra-cluster, inter-cluster, and intra-outlier cases. The resulting layout is presented in Fig. 5e, g: lines represent individual data series, and dark rectangular blocks represent the detected clusters of data series.

By glancing at this representation, the user can identify the major groups of data series and time steps where the behavior of such groups and series changes drastically. For example, by investigating Fig. 5e, we discover that most of the data series related to *Trump/Reddit* (purple) demonstrated similar behavior during the first half of the currently explored time range. At the same time, the data series related to *Brexit/Reddit* demonstrated several patterns of behavior, including one rather stable cluster. By the timestamp marked in Fig. 5f, g, the two main clusters for both targets of interest merge, thus indicating the similar temporal pattern for the corresponding data series. From this particular example, we can observe that sentiment and stance data series for different targets of interest demonstrate similar behavior despite the patterns demonstrated by the corresponding document count series (bolder lines, faded out in Fig. 5). The user can

Fig. 5 Visualization of data series similarity in StanceVis Prime: **a** the loaded data series table (with support for filtering and brushing); **b** sparkline previews for data series; **c** a temporal slider; **d** a 2D projection view representing the data series at the current time step; **e** a temporal similarity view representing the data series' projections over time; **f** an indicator of the current time step; and **g** a rectangular representation of a data series cluster in the 1D projection. Here, the user has hovered over the cluster **g**, which caused all the data series not included in that cluster to fade out

learn further details from the tooltip and also use the 2D view in Fig. 5d to explore the relationships between the data series in addition to the representation in Fig. 5e. These views are coordinated with the others by brushing and linking, so the user can explore the similarity between data series to fulfill task **T3** and then switch to investigation of specific data series values, as discussed below.

5.3 Representation of data series values

While the previous steps focused on the *similarity* between the loaded data series, they did not provide a visual representation of their actual *values*. As mentioned in Sect. 4, each data series supported by our approach contains the timestamped values for a combination of target, domain, and category such as *Brexit/Reddit/DISAGREEMENT*. The values provide the information about the detected document counts or the number of corresponding stance or sentiment cases. The data series values are visualized according to the pipeline in Fig. 6 with the representations depicted in Fig. 7. A stacked graph (Havre et al. 2000; Byron and Wattenberg 2008) is used in the central part of the interface to represent the overall counts of processed documents for each target/domain over time (see Fig. 7a). The user is initially presented with an overview of the complete loaded data set and can then focus on a specific time interval using the range slider depicted in Fig. 7b, which might cause the change of data granularity (e.g., from days to hours). The main graph focuses mostly on the overall document counts rather than subjectivity categories, since it would require up to 2+12 additional plots per target/domain. However, the representation includes visual cues about the temporal points with relatively high amounts of detected subjectivity. For instance, a “Dis” label is displayed in Fig. 7d over the graph for *Brexit/Reddit*. It means that at the corresponding time step the value of the *Brexit/Reddit/DISAGREEMENT* data series was relatively high compared to the maximum value in this loaded series. By using the controls depicted in Fig. 7f, the user can adjust the minimal threshold for the relative level of subjectivity or hide such subjectivity cues altogether (as they could lead to visual clutter in some cases). This visual representation supports tasks **T1** and **T2**.

Fig. 6 The pipeline for data series *values* visualization in StanceVis Prime

Fig. 7 Visualization of data series values in StanceVis Prime: **a** a stacked graph representing document counts for target/domain combinations over time; **b** a range slider providing an overview for the complete loaded data set; **c** an indicator of the time step selected in Fig. 5c; **d** a subjectivity cue label indicating sentiment/stance; **e** a time range selection for document queries; **f** subjectivity cues controls and the document query button; **g** sentiment data series representations; **h** stance data series representations; and **i** a panel collapse button

To support **T2** further, our implementation also provides multiple small bar charts for separate sentiment and stance data series displayed in Fig. 7g, h. This specific visual encoding was selected for several reasons in contrast to the stacked graph representation presented above. First of all, our earlier design attempts at reusing the stacked graph technique for each of the multiple data series views revealed the issues of representing the rather sparse data, especially at the more detailed levels of temporal granularity such as one second. In these scenarios, data points could be missing for certain timestamps, as no corresponding sentiment or stance category occurrences were detected within the captured documents, thus resulting in non-contiguous regions for visual representations such as a stacked graph. Stacked bar charts were therefore selected as a representation that could address this issue while supporting the visual comparison of several data series at any given timestamp. The additional benefits of this choice include (1) visual indication of a different data type compared to the main stacked graph view, i.e., the number of utterances with sentiment or stance detected—as opposed to the number of captured documents; (2) support for representing the average classification confidence level for the data series at each timestamp with opacity; and (3) reliance on a simple, standard visual metaphor familiar to the users (Russell 2016). Investigation—and evaluation—of further alternatives for representing the sentiment and stance data series, either separately or combined, is part of our future work.

If the user is not interested in the details visualized with the bar charts, it is possible to collapse the panels to save screen space (see Fig. 7i). The details about individual entries for these charts can be revealed with a tooltip, including the information such as the average classification confidence for a specific category of stance, for instance.

Finally, the main stacked graph view also supports task **T4**: the user can request the documents corresponding to the selected time range and target/domain combinations (cf. Fig. 7e, f).

5.4 Representation of document lists

On user's request, the document texts with sentiment and stance classification results are retrieved from the database (see Fig. 8), and a new document list is displayed in the user interface, as displayed in Fig. 9. Additionally, a static thumbnail of the graph selection is created to serve as a snapshot (cf. Fig. 9b).

One part of task **T5** is concerned with summarization of text contents of such document lists. To address this, we have decided to provide the users with a list of key terms. Our implementation computes the counts of uni- and bigrams (Manning and Schütze 1999) present in the document texts and then processes them using TF-IDF weighting (Salton and Buckley 1988). Currently, up to 25 top key terms are returned for each document list and represented as an interactive list displayed in Fig. 9c. We used a simple list representation (with eventual wrapped lines/rows) instead of a word cloud with spatial layout, as similar effectiveness of these approaches for topic discovery tasks was demonstrated recently (Felix et al. 2018).

To represent the summary of category classification results for **T5**, our implementation includes an interactive list of statistics for sentiment and stance categories displayed in Fig. 9d. The total number of utterances with a specific category detected is computed over all document list texts and then displayed as a numeric label. Additionally, the average stance classification confidence and the average valence for detected POSITIVITY/NEGATIVITY are visualized with bar charts.

Fig. 8 The pipeline for document lists and text views representation in StanceVis Prime

Fig. 9 Document list representations in StanceVis Prime: **a** a header including the list description, close and export buttons, and a drag'n'drop indicator icon; **b** a static thumbnail copied from the representation displayed in Fig. 7a; **c** a list of salient uni- and bigrams discovered in the document texts; **d** statistics for sentiment and stance classification results in the document texts; **e** additional controls including an area for user notes, a text search field, and a sort button; **f** a document view with injected labels with utterance classification results; and **g** a tooltip with details for a selected document. Here, the user has changed the position of one of the lists with drag'n'drop and used various interactions in **c-e** and **g**

The user interactions for document lists include filtering (by key terms or detected categories), notes/annotations, text search (cf. Fig. 9e), sorting (e.g., by timestamp or by detected stance occurrences), reordering the lists for comparison, and removing them.

To support **T6**, the document lists provide access to the actual texts as seen in Fig. 9f, including links to the source social media posts. Sentiment and stance classification results are injected directly into the document view as labels with abbreviated category titles and classification confidence results. By hovering over a document, the user is presented with a tooltip displayed in Fig. 9g which includes additional details about the document timestamp, data domain, associated target(s) of interest, and detailed classification results. It is also possible to click a document element to make the tooltip persistent.

Finally, the user can use the export button located in the document list header (see Fig. 9a). In this case, a static version of the document list is exported as an HTML file, with user notes, filtering, and sorting preserved (cf. Fig. 10d, g). Exported lists can be used for further offline investigation, thus supporting **T7**.

6 Case study

In this section, we demonstrate the usage of StanceVis Prime with two case studies. In the first study, the information about public sentiment and stance on two different targets of interest is compared for Twitter data. In the second study, we focus on a single target of interest in two different data sources, Twitter and Reddit.

6.1 Case study a: Brexit and European politics on Twitter

In this case study, we have explored the Twitter data for summer 2018 available in StanceVis Prime. We have focused on two targets of interest tracked in our system: *Brexit* and *European Politics*. The initial time range of interest is the complete summer 2018; the resulting stacked graph, initially loaded at the granularity of one day, is displayed in Fig. 10a.

We can see that the volume of tweets for *Brexit* (in green) dominates this data set, and the maximum number of posts is created at a time step highlighted in Fig. 10b. This time step corresponds to July 9, 2018, and there are 742,139 tweets retrieved in our system for this data point. According to the subjectivity cues, multiple categories of sentiment and stance reach large values on this day, such as **NEGATIVITY**, **UNCERTAINTY**, and **POSITIVITY**.

As we focus on a shorter time range corresponding to this day and filter out the data series related to the other target, the representation changes to hourly values, as displayed in Fig. 10c. We can now see that a large number of tweets were produced starting around 16:00 (GMT+2). Exploration reveals that there are

Fig. 10 The Twitter data in case study A: **a** the data series view in StanceVis Prime for *Brexit* (green) and *European Politics* (red) for the complete summer 2018; **b** the data selection corresponding to July 9, 2018; **c** the updated data series view representing the hourly data for *Brexit* on July 9, 2018; **d** the exported document list with the tweets relevant to *Brexit* retrieved for the time period 16:00–16:20 on July 9, 2018; **e** the data selection corresponding to July 15, 2018; **f** the updated data series view representing the hourly data on July 15, 2018; and **g** the exported document list with the tweets relevant to *European Politics* retrieved for the time period 17:30–18:30 on July 15, 2018

79,711 tweets in total for this hour, containing 14,706 utterances with UNCERTAINTY, 12,992 with NEGATIVITY, and 11,279 with POSITIVITY.

After focusing on a shorter period (16:00–16:20) and retrieving the documents, we can find out that the main key terms for the corresponding document list were “secretary,” “johnson,” “resigns,” and so on. Apparently, this spike in the discussion of *Brexit* on Twitter was related to the resignation of the UK Foreign Secretary Boris Johnson. The related utterances include, for example, “*Chaos ensues*” (labeled with NEGATIVITY by our classifiers), “*If true it’s effectively the final nail in the coffin for May*” (POSITIVITY, HYPOTHETICALS, and UNCERTAINTY), and “*Could be good for Ireland if it leads to Theresa May having a stronger hand to push through a soft Brexit*” (POSITIVITY, HYPOTHETICALS, NEED & REQUIREMENT, and UNCERTAINTY). After interacting with the document list, we can export it for further offline exploration, with the result displayed in Fig. 10d.

Returning to the complete loaded data set, we can now focus on the data region where *European Politics* gained prominence (see the red band of the stacked graph in Fig. 10e), which corresponds to July 15, 2018, with 122,298 tweets in our retrieved data. Focusing further and switching to the granularity of one hour, we can see in Fig. 10f that sudden growth of interest for this target started around 18:00 (GMT+2). There are 13,945 tweets corresponding to this hour, with 7,990 utterances classified with NEGATIVITY, 2,177 with POSITIVITY, and 2,070 with the stance of SOURCE OF KNOWLEDGE. After focusing on the period of 17:30–18:30 and investigating the document list, we can find out that this case is related to the quote of POTUS Donald Trump naming European Union “a foe” in an interview. As we can see in the exported document list in Fig. 10g, a lot of expressions of sentiment were produced in the tweets, followed by several stance categories such as SOURCE OF KNOWLEDGE, NEED & REQUIREMENT, and CONCESSION & CONTRARIETY. The examples include utterances such as “*If you look at the thing they quoted, Trump clearly said that the European Union was an economic foe*” (CERTAINTY, HYPOTHETICALS, and SOURCE OF KNOWLEDGE), “*Man’s a lunatic*” (NEGATIVITY), and “*But the concern should be there just the same*” (NEED & REQUIREMENT).

By using StanceVis Prime in this case study, we were able to understand the temporal trends in public sentiment and stance on several targets of interest, retrieve and get an overview of the underlying text data for specific time ranges, and use the exported versions of document lists for close reading offline (cf. user tasks T1–T2 and T4–T7).

6.2 Case study B: European politics in Twitter data versus Reddit comments

In the second case study, we have compared the data on the same target of interest (*European Politics*) from two data sources, Twitter and Reddit. We have selected a time range of several days around September 9,

Fig. 11 The document lists for *European Politics* in case study B: **a** relevant tweets retrieved for the evening of September 9, 2018; **b** relevant Reddit comments retrieved for the evening of September 9, 2018; and **c** relevant Reddit comments retrieved for the morning of September 10, 2018

2018, when a general election took place in Sweden. After exploring the temporal data and downloading the document lists, we can confirm the expectation that the number of documents from Twitter for the same time range would be much larger than the number of Reddit comments on this subject, as seen in Fig. 11a, b with approximately 5000 tweets and 50 Reddit comments (however, we should note that the incoming Reddit data stream was initially filtered to a set of specific subforums, as mentioned in Sect. 4). Filtering the document lists on key terms related to the Swedish election reduces these numbers even further. The examples of relevant utterances found in tweets include “*Today is a great day in Sweden*” (POSITIVITY), “*I guess that will do Sweden*” (PREDICTION and UNCERTAINTY), and “*I understand none of it but hey I might live there some day so I should probably watch their election process*” (CONCESSION & CONTRARIENESS, CONTRAST, NEED & REQUIREMENT, and UNCERTAINTY). The relevant Reddit comments include utterances such as “*I heard there are reports of election fraud*” (NEGATIVITY and SOURCE OF KNOWLEDGE), “*Most votes for the party were clearly protest votes*” (POSITIVITY and CERTAINTY), and “*Sweden TV needs to ask the BBC for tips on how to report elections...*” (NEED & REQUIREMENT).

A more lively discussion of the Swedish elections on Reddit does not start until the morning of the next day, September 10, as seen in Fig. 11c. It is interesting to compare the documents between the data sources, though, as Reddit comments tend to contain much longer statements and arguments rather than short tweets, which might be interesting for further detailed exploration with close reading. For instance, some relevant utterances found in this case are “*However it would be completely illogical if the right parties with fewer mandates demands to govern anyway*” (DISAGREEMENT and HYPOTHETICALS), “*Sweden is extremely progressive, most parties would be seen as progressive parties in the US except the nationalist Sweden Democrats and the Christian Democrats*” (CONCESSION & CONTRARIENESS and CONTRAST), and “*As the statistics showed, most SD voters come from the Socialdemocrats which have been the most progressive party in Sweden*” (SOURCE OF KNOWLEDGE).

In summary, this case study demonstrated support for comparisons between data from several sources with a focus on retrieval, overview, and close reading of multiple document lists (cf. user tasks T4–T7).

7 Discussion

In this section, we report the preliminary expert user feedback received for StanceVis Prime and reflect upon various aspects of our implementation, including its scalability and generalizability.

7.1 Expert user review

While our approach has not been evaluated with a larger user study yet, we have relied on an expert review to collect valuable user feedback (Tory and Möller 2005). It was provided by one of our project collaborators, a postdoctoral researcher in computational linguistics, who had previously contributed to the development of the stance classifier used by our system’s backend. This expert user had prior knowledge about information visualization and visual analytics, but she had not been involved in the design and development of StanceVis Prime. During the session which lasted several hours, we (1) explained the

underlying design choices, (2) demonstrated the visual interface functionality, (3) allowed the user to engage in free data exploration with a think-aloud protocol, and (4) conducted a semi-structured interview using the ICE-T questionnaire by Wall et al. (2019) focusing on the value of visualization.

The expert decided to focus on the target of *Vaccination* rather than political targets, as she had previously worked on this topic in her own research. Based on the sparkline charts in the data loading dialog, the expert settled on the data from mid-August to early September 2018 from both Twitter and Reddit. When exploring the stacked graph view, the cues about NEGATIVITY and NEED & REQUIREMENT in the Twitter data for a specific day (August 26, 2018) caught the eye of the expert. She focused on this date, studied the detailed category charts, and loaded the corresponding documents (approximately 33,000 tweets). The expert used the category bar charts to focus on the documents with NEED & REQUIREMENT and/or NEGATIVITY and engaged in close reading of the resulting documents. After studying multiple texts, she concluded that (1) the classification results seemed reasonable; (2) there were a lot of urges and requests both to vaccinate and not to vaccinate in the tweets; and (3) NEGATIVITY did not co-occur with NEED & REQUIREMENT in the data.

The expert also investigated other time ranges with a focus on TACT and SOURCE OF KNOWLEDGE. Her insights included the discoveries that (1) many occurrences of TACT in this data are in fact explained by sarcasm; and (2) many references to SOURCE OF KNOWLEDGE include vague statements and opinions (“*people are saying that...*”) rather than facts and authorities (however, several references to organizations, such as CDC, were also found).

In general, the expert concluded that StanceVis Prime seemed useful for the tasks of investigating public sentiment and stance. As her expertise is in computational linguistics, she was particularly interested in exploring the classification results in documents corresponding to the peaks in the data series. The expert user also commended the possibility of exporting the document lists annotated with user notes for further analyses and close reading.

The expert also had some suggestions and notes on the current shortcomings, mostly related to the document list views. First of all, she noted a number of duplicate documents (retweets) in the Twitter data and suggested collapsing them in order to focus only on unique texts. Second, she noted that it would be useful to support “AND”-based category filters (in addition to the currently supported “OR”) in document list views in order to focus on particular combinations of categories, such as NEGATIVITY and NEED & REQUIREMENT mentioned above. Taking such selected combinations of categories into account when sorting the document lists was also potentially useful in the expert’s opinion. Finally, the delays associated with the document loading stage for larger data sets (more than 10,000 documents) were also perceived negatively. These comments and suggestions will be taken into account as part of the future work to improve StanceVis Prime.

7.2 Scalability and limitations

While our approach is designed to continuously retrieve and apply computational analyses to text documents from streaming data sources such as Twitter, it is currently unable to support real-time streaming data processing and visual analysis for several reasons. First of all, the current design of the backend (see Sect. 4) includes the data collection service, the database, and the classification service as separate components. After the incoming documents are received and preliminary filtered (typically at a rate up to several dozens of documents per second, depending on the targets of interest, time of day, ongoing events, etc.), they are stored in the database and put on the classification queue. The classifiers provided by our project collaborators are currently deployed separately from StanceVis Prime since they are used for multiple purposes as a RESTful web service. Unfortunately, this implies a significant overhead for data serialization and transfer, and one of our next steps for improving the performance of StanceVis Prime is going to be the direct integration of the classification codebase in our system. As for the scalability of the server-side computations for visual analysis tasks, StanceVis Prime uses efficient implementations of DTW, MDS, and DBSCAN algorithms for data series similarity analysis, as discussed in Sect. 5.2, and standard text pre-processing algorithms for document lists, as discussed in Sect. 5.4. Thus, as long as the indices are set up properly within the underlying database, the performance of server-side computations for the visualization frontend is not the main performance concern. Further performance measurements for individual server-side elements of our system can be considered part of our future work, though. Extension of the client-side elements of StanceVis Prime to support real-time streaming data visualization is also an interesting prospect,

however, as we discuss below in Sect. 7.3, such a real-time monitoring task might not be a priority for the users interested in long-term trends and underlying textual data.

With regard to the scalability of the visual interface, we should also note that the coordinated multiple views strategy has a drawback related to the overall screen area usage. The user interface of StanceVis Prime supports window resizing with regard to width, but it is designed with a vertical workflow in mind, from the source data selection and listing to the data series charts and the document lists, which results in vertical scrolling on smaller monitors. One possible alternative solution would be to introduce separate windows or tabs for separate views, similar to document list tabs in uVSAT (Kucher et al. 2016). However, it could arguably disrupt the user's mental map and would also make comparison between multiple document lists more difficult. For the time being, we intend our implementation to be used on desktop monitors rather than laptop screens. Displaying some of the detailed views on demand only is also an option (cf. Fig. 7i).

Additionally, the visualization of data series similarity discussed in Sect. 5.2 can be subject to visual clutter depending on the number of displayed data series and the output of MDS and DBSCAN algorithms. This representation could be improved in the future using the methods introduced for storyline visualizations (Tanahashi and Ma 2012; Silvia et al. 2016). While the 2D projection view displayed in Fig. 5d is currently limited to MDS, we could allow the user to change the underlying dimensionality reduction algorithm to reduce the clutter, for instance. Our data series representations could also adapt some of the aspects of MultiStream (Cuenca et al. 2018) for hierarchical multiscale series. We are also aware of the risks of the current color coding scheme used for targets/sources and categories in these representations, however, the expert user feedback did not include any complaints in this regard.

Finally, the web browser rendering performance for our visual interface depends on the loaded data set size, especially for document lists. Thus, the workflow currently recommended to the users is to focus on moderate time ranges when requesting the documents, i.e., the scale of minutes or hours rather than days or months.

7.3 Overall utility and generalizability

Based on the range of user tasks supported by StanceVis Prime (see Sect. 3), we expect its components and views to have different value to different users: for instance, the expert feedback discussed in Sect. 7.1 confirms our expectation that experts in linguistics and computational linguistics would ultimately be interested in accessing and studying text data when using our tool. For example, the said expert was mainly interested in several specific sentiment and stance categories as well as the underlying documents, rather than investigation of numerous data series over an extended period of time. The representations focusing on temporal data and data series similarity could be more interesting to the users interested in overall trends in public sentiment and stance, such as brand managers or political scientists, and it would be a part of our future work to collaborate with such users. Further interactive control over the parameters of computational analyses, such as DTW, should be provided to such users, too. While our case studies in Sect. 6 did not make use of the DTW-related views, the example in Sect. 5.2 demonstrated how these views could help the users analyze the similarities between sentiment and stance patterns across several targets of interest or data sources. These are exactly the insights associated with the user task **T3**, and we would like to continue the investigation of such analyses as part of our future work.

We foresee several ways to generalize our approach. First of all, StanceVis Prime was initially designed to support multiple data sources: it is already possible to access and compare the data from the two currently supported social media platforms. Our approach can also be generalized to other temporal text data sources and, with some modifications, to static corpora. Second, we are currently using a specific set of sentiment and stance categories which could be extended or replaced as long as text analyses are carried out at the utterance/sentence level: for instance, our stance classification pipeline could be extended with another classifier which achieves better classification results for a limited set of categories such as *AGREEMENT* and *DISAGREEMENT* (Mohammad et al. 2017) by using deep learning (Chen and Ku 2016; Lai et al. 2015; Liu et al. 2017) or ensemble learning approaches (Sagi and Rokach 2018). Finally, our research so far has focused only on texts in English, and it would be interesting to apply StanceVis Prime to text data in other languages once the respective sentiment and stance classifiers are available.

8 Conclusions and future work

In this paper, we have discussed our work on StanceVis Prime, a visual analytics platform for social media texts supporting sentiment and stance analysis. Our approach is based on the definition of stance categories and a classifier developed as part of an interdisciplinary project on stance analysis. Based on the requirements of our project collaborators, StanceVis Prime allows the users to investigate temporal trends, get an overview about document sets, study individual texts, and export processed documents for further offline exploration. We demonstrate our approach with two case studies and expert user feedback from one of our collaborators. Further analysis of the data already collected with StanceVis Prime (190 million documents so far) in collaboration with experts in sociolinguistics is one of the next steps in our work. Another step is evaluation with a full-fledged user study.

This contribution opens up additional opportunities for future work on visual analysis of sentiment and stance as well as other aspects of temporal text data. Our approach could be enriched with additional analyses and representations, e.g., by using geospatial information (Martins et al. 2017), named entity recognition (El-Assady et al. 2016), topic modeling (Dou and Liu 2016; Wang et al. 2016b), and anomaly detection (Cao et al. 2016). While StanceVis Prime supports the export of some of the analytical findings as annotated document lists, our approach could be extended with further means to externalize and present analytical results, thus following the suggestions by Chen et al. (2018) on bridging the gap between visual analytics and storytelling. Visual analysis of the public discourse structure (Cao et al. 2015; El-Assady et al. 2018; Lu et al. 2018) enriched by sentiment and stance data is also an interesting prospect. Combining such an approach with the analyses discussed above would make it possible to shift the focus of the workflow from the initial selection of targets of interests—as currently supported by StanceVis Prime—towards the monitoring and investigation of patterns and themes emerging in the incoming data in a dynamic way. Finally, the methods of predictive analytics (Lu et al. 2017) could be introduced to forecast public sentiment and stance on specific targets of interest in social media.

Acknowledgements Open access funding provided by Linnaeus University. The authors would like to thank Maria Skeppstedt for her contributions to the implementation of computational text analyses, as well as for providing her feedback for StanceVis Prime. This work was partially supported by the framework grant “The Digitized Society—Past, Present, and Future” from the Swedish Research Council (Vetenskapsrådet) [Grant Number 2012–5659].

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aigner W, Miksch S, Schumann H, Tominski C (2011) Visualization of time-oriented data. Springer, Berlin. <https://doi.org/10.1007/978-0-85729-079-3>
- Alencar AB, Börner K, Paulovich FV, de Oliveira MCF (2012) Time-aware visualization of document collections. In: Proceedings of the 27th annual ACM symposium on applied computing, ACM, SAC’12, pp 997–1004. <https://doi.org/10.1145/2245276.2245469>
- Alspaugh S, Zokaei N, Liu A, Jin C, Hearst MA (2019) Futzing and moseying: interviews with professional data analysts on exploration practices. *IEEE Trans Vis Comput Graphics* 25(1):22–31. <https://doi.org/10.1109/TVCG.2018.2865040>
- Bernard J, Wilhelm N, Scherer M, May T, Schreck T (2012) TimeSeriesPaths: projection-based explorative analysis of multivariate time series data. *J WSCG* 20(2):97–106
- Berndt DJ, Clifford J (1994) Using dynamic time warping to find patterns in time series. In: Proceedings of the AAAI workshop on knowledge discovery in databases, AAAI Press, KDD’94, pp 359–370
- Biber D, Finegan E (1989) Styles of stance in English: lexical and grammatical marking of evidentiality and affect. *Interdiscip J Study Discourse* 9(1):93–124. <https://doi.org/10.1515/text.1.1989.9.1.93>
- Borg I, Groenen PJF (2005) Modern multidimensional scaling: theory and applications. Springer, Berlin. <https://doi.org/10.1007/0-387-28981-X>
- Bostock M (2011) D3—data-driven documents. <https://d3js.org/>. Accessed 28 July 2020
- Brewer C, Harrower M, The Pennsylvania State University (2009) ColorBrewer 2.0—color advice for cartography. <http://colorbrewer2.org/>. Accessed 28 July 2020

- Byron L, Wattenberg M (2008) Stacked graphs: geometry & aesthetics. *IEEE Trans Vis Comput Graphics* 14(6):1245–1252. <https://doi.org/10.1109/TVCG.2008.166>
- Cao N, Lin YR, Sun X, Lazer D, Liu S, Qu H (2012) Whisper: tracing the spatiotemporal process of information diffusion in real time. *IEEE Trans Vis Comput Graphics* 18(12):2649–2658. <https://doi.org/10.1109/TVCG.2012.291>
- Cao N, Lu L, Lin YR, Wang F, Wen Z (2015) SocialHelix: visual analysis of sentiment divergence in social media. *J Vis* 18(2):221–235. <https://doi.org/10.1007/s12650-014-0246-x>
- Cao N, Shi C, Lin S, Lu J, Lin YR, Lin CY (2016) TargetVue: visual analysis of anomalous user behaviors in online communication systems. *IEEE Trans Vis Comput Graphics* 22(1):280–289. <https://doi.org/10.1109/TVCG.2015.2467196>
- Chatzimpampas A, Martins RM, Jusufi I, Kucher K, Rossi F, Kerren A (2020) The state of the art in enhancing trust in machine learning models with the use of visualizations. *Comput Graphics Forum* 39(3):713–756. <https://doi.org/10.1111/cgf.14034>
- Chen WF, Ku LW (2016) UTCNN: a deep learning model of stance classification on social media text. In: Proceedings of the 26th international conference on computational linguistics—technical papers, ACL, COLING 2016, pp 1635–1645
- Chen S, Lin L, Yuan X (2017) Social media visual analytics. *Comput Graphics Forum* 36(3):563–587. <https://doi.org/10.1111/cgf.13211>
- Chen S, Li J, Andrienko G, Andrienko N, Wang Y, Nguyen PH, Turkay C (2018) Supporting story synthesis: bridging the gap between visual analytics and storytelling. *IEEE Trans Vis Comput Graphics*. <https://doi.org/10.1109/TVCG.2018.2889054>
- Crnovrsanin T, Muelder C, Correa C, Ma KL (2009) Proximity-based visualization of movement trace data. In: Proceedings of the IEEE symposium on visual analytics science and technology, VAST'09, pp 11–18. <https://doi.org/10.1109/VAST.2009.5332593>
- Cuenca E, Sallaberry A, Wang FY, Poncelet P (2018) MultiStream: a multiresolution streamgraph approach to explore hierarchical time series. *IEEE Trans Vis Comput Graphics* 24(12):3160–3173. <https://doi.org/10.1109/TVCG.2018.2796591>
- Cui W, Liu S, Tan L, Shi C, Song Y, Gao Z, Qu H, Tong X (2011) TextFlow: towards better understanding of evolving topics in text. *IEEE Trans Vis Comput Graphics* 17(12):2412–2421. <https://doi.org/10.1109/TVCG.2011.239>
- Diakopoulos N, Zhang AX, Elgesem D, Salway A (2014) Identifying and analyzing moral evaluation frames in climate change blog discourse. In: Proceedings of the eighth international AAAI conference on weblogs and social media, AAAI, ICWSM'14, pp 583–586
- Dörk M, Gruen D, Williamson C, Carpendale S (2010) A visual backchannel for large-scale events. *IEEE Trans Vis Comput Graphics* 16(6):1129–1138. <https://doi.org/10.1109/TVCG.2010.129>
- Dou W, Liu S (2016) Topic- and time-oriented visual text analysis. *IEEE Comput Graphics Appl* 36(4):8–13. <https://doi.org/10.1109/MCG.2016.73>
- El-Assady M, Gold V, Acevedo C, Collins C, Keim DA (2016) ConToVi: multi-party conversation exploration using topic-space views. *Comput Graphics Forum* 35(3):431–440. <https://doi.org/10.1111/cgf.12919>
- El-Assady M, Sevastjanova R, Keim D, Collins C (2018) ThreadReconstructor: modeling reply-chains to untangle conversational text through visual analytics. *Comput Graphics Forum* 37(3):351–365. <https://doi.org/10.1111/cgf.13425>
- Englebretson R (ed) (2007) Stancetaking in discourse: subjectivity, evaluation, interaction, pragmatics & beyond new series, vol 164. John Benjamins, Amsterdam. <https://doi.org/10.1075/pbns.164>
- Esling P, Agon C (2012) Time-series data mining. *ACM Comput Surv* 45(1):12:1–12:34. <https://doi.org/10.1145/2379776.2379788>
- Ester M, Kriegel HP, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the second international conference on knowledge discovery and data mining, AAAI Press, KDD'96, pp 226–231
- Felix C, Franconeri S, Bertini E (2018) Taking word clouds apart: an empirical investigation of the design space for keyword summaries. *IEEE Trans Vis Comput Graphics* 24(1):657–666. <https://doi.org/10.1109/TVCG.2017.2746018>
- Glynn D, Sjölin M (eds) (2014) Subjectivity and epistemicity: corpus, discourse, and literary approaches to stance. Lund studies in English. Lund University Press, Lund
- Havre S, Hetzler B, Nowell L (2000) ThemeRiver: visualizing theme changes over time. In: Proceedings of the IEEE symposium on information visualization, IEEE, InfoVis'00, pp 115–123. <https://doi.org/10.1109/INFVIS.2000.885098>
- Hosmer DW Jr, Lemeshow S, Sturdivant RX (2013) Applied logistic regression. Wiley, Hoboken. <https://doi.org/10.1002/9781118548387>
- Hutto C, Gilbert E (2014) VADER: a parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the eighth international AAAI conference on weblogs and social media, AAAI, ICWSM'14
- Jäckle D, Fischer F, Schreck T, Keim DA (2016) Temporal MDS plots for analysis of multivariate data. *IEEE Trans Vis Comput Graphics* 22(1):141–150. <https://doi.org/10.1109/TVCG.2015.2467553>
- Jänicke S, Franzini G, Cheema MF, Scheuermann G (2015) On close and distant reading in digital humanities: a survey and future challenges. In: Proceedings of the EG/VGTC conference on visualization—STARs, The Eurographics Association, EuroVis'15. <https://doi.org/10.2312/eurovisstar.20151113>
- Krzanowski WJ (2000) Principles of multivariate analysis. Oxford statistical science series. Oxford University Press, Oxford
- Kucher K, Kerren A (2015) Text visualization techniques: taxonomy, visual survey, and community insights. In: Proceedings of the 8th IEEE Pacific visualization symposium, IEEE, PacificVis'15, pp 117–121. <https://doi.org/10.1109/PACIFICVIS.2015.7156366>
- Kucher K, Schamp-Bjerede T, Kerren A, Paradis C, Sahlgren M (2016) Visual analysis of online social media to open up the investigation of stance phenomena. *Inf Vis* 15(2):93–116. <https://doi.org/10.1177/1473871615575079>
- Kucher K, Paradis C, Sahlgren M, Kerren A (2017) Active learning and visual analytics for stance classification with ALVA. *ACM Trans Interact Intell Syst* 7(3):141–1431. <https://doi.org/10.1145/3132169>
- Kucher K, Paradis C, Kerren A (2018a) The state of the art in sentiment visualization. *Comput Graphics Forum* 37(1):71–96. <https://doi.org/10.1111/cgf.13217>

- Kucher K, Paradis C, Kerren A (2018b) Visual analysis of sentiment and stance in social media texts. In: Poster abstracts of the EG/VGTC conference on visualization, The Eurographics Association, EuroVis'18, pp 49–51. <https://doi.org/10.2312/eurp.20181127>
- Lai S, Xu L, Liu K, Zhao J (2015) Recurrent convolutional neural networks for text classification. In: Bonet B, Koenig S (eds) Proceedings of the twenty-ninth AAAI conference on artificial intelligence, AAAI, AAAI'15
- Liu S, Wu Y, Wei E, Liu M, Liu Y (2013) StoryFlow: tracking the evolution of stories. *IEEE Trans Vis Comput Graphics* 19(12):2436–2445. <https://doi.org/10.1109/TVCG.2013.196>
- Liu J, Chang WC, Wu Y, Yang Y (2017) Deep learning for extreme multi-label text classification. In: Proceedings of the 40th international ACM SIGIR conference on research and development in information retrieval, ACM, SIGIR'17, pp 115–124. <https://doi.org/10.1145/3077136.3080834>
- Liu S, Wang X, Collins C, Dou W, Ouyang F, El-Assady M, Jiang L, Keim DA (2019) Bridging text visualization and mining: a task-driven survey. *IEEE Trans Vis Comput Graphics* 25(7):2482–2504. <https://doi.org/10.1109/TVCG.2018.2834341>
- Lu Y, Garcia R, Hansen B, Gleicher M, Maciejewski R (2017) The state-of-the-art in predictive visual analytics. *Comput Graphics Forum* 36(3):539–562. <https://doi.org/10.1111/cgf.13210>
- Lu Y, Wang H, Landis S, Maciejewski R (2018) A visual analytics framework for identifying topic drivers in media events. *IEEE Trans Vis Comput Graphics* 24(9):2501–2515. <https://doi.org/10.1109/TVCG.2017.2752166>
- Manning CD, Schütze H (1999) Foundations of statistical natural language processing. MIT Press, Cambridge
- Martins RM, Kerren A (2018) Efficient dynamic time warping for big data streams. In: Proceedings of the 3rd workshop on real-time & stream analytics in big data & stream data management at IEEE Big Data'18, pp 2924–2929. <https://doi.org/10.1109/BigData.2018.8621878>
- Martins RM, Simaki V, Kucher K, Paradis C, Kerren A (2017) StanceXplore: visualization for the interactive exploration of stance in social media. In: Proceedings of the 2nd workshop on visualization for the digital humanities, VIS4DH'17
- Mohammad SM (2016) Sentiment analysis: detecting valence, emotions, and other affectual states from text. In: Meiselman HL (ed) Emotion measurement. Woodhead Publishing, Sawston, pp 201–237. <https://doi.org/10.1016/B978-0-08-100508-8.00009-6>
- Mohammad SM, Kiritchenko S, Sobhani P, Zhu X, Cherry C (2016) SemEval-2016 task 6: detecting stance in tweets. In: Proceedings of the international workshop on semantic evaluation, SemEval'16
- Mohammad SM, Sobhani P, Kiritchenko S (2017) Stance and sentiment in tweets. *ACM Trans Internet Technol* 17(3):26:1–26:23. <https://doi.org/10.1145/3003433>
- Pang B, Lee L (2008) Opinion mining and sentiment analysis. *Found Trends Inf Retr* 2(1–2):1–135. <https://doi.org/10.1561/1500000011>
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É (2011) Scikit-learn: machine learning in Python. *J Mach Learn Res* 12:2825–2830
- Pirolli P, Card S (2005) The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In: Proceedings of the international conference on intelligence analysis, vol 5
- Rauber PE, Falcão AX, Telea AC (2016) Visualizing time-dependent data using dynamic t-SNE. In: Short papers of the EG/VGTC conference on visualization, The Eurographics Association, EuroVis'16. <https://doi.org/10.2312/eurovisshort.20161164>
- Roberts JC (2007) State of the art: coordinated & multiple views in exploratory visualization. In: Proceedings of the fifth international conference on coordinated and multiple views in exploratory visualization, IEEE, CMV'07, pp 61–71. <https://doi.org/10.1109/CMV.2007.20>
- Russell DM (2016) Simple is good: Observations of visualization use amongst the Big Data digerati. In: Proceedings of the international working conference on advanced visual interfaces, ACM, AVI'16, pp 7–12. <https://doi.org/10.1145/2909132.2933287>
- Sacha D, Stoffel A, Stoffel F, Kwon BC, Ellis G, Keim DA (2014) Knowledge generation model for visual analytics. *IEEE Trans Vis Comput Graphics* 20(12):1604–1613. <https://doi.org/10.1109/TVCG.2014.2346481>
- Sagi O, Rokach L (2018) Ensemble learning: a survey. *WIREs Data Min Knowl Discov* 8(4):e1249. <https://doi.org/10.1002/widm.1249>
- Salton G, Buckley C (1988) Term-weighting approaches in automatic text retrieval. *Inf Process Manag* 24(5):513–523. [https://doi.org/10.1016/0306-4573\(88\)90021-0](https://doi.org/10.1016/0306-4573(88)90021-0)
- Shi C, Cui W, Liu S, Xu P, Chen W, Qu H (2012) RankExplorer: visualization of ranking changes in large time series data. *IEEE Trans Vis Comput Graphics* 18(12):2669–2678. <https://doi.org/10.1109/TVCG.2012.253>
- Shrestha A, Miller B, Zhu Y, Zhao Y (2013) Storygraph: extracting patterns from spatio-temporal data. In: Proceedings of the ACM SIGKDD workshop on interactive data exploration and analytics, ACM, IDEA'13, pp 95–103. <https://doi.org/10.1145/2501511.2501525>
- Shutterstock Images, LLC (2011) Rickshaw: a JavaScript toolkit for creating interactive time-series graphs. <https://github.com/shutterstock/rickshaw>. Accessed 28 July 2020
- Silvia S, Etemadpour R, Abbas J, Huskey S, Weaver C (2016) Visualizing variation in classical text with force directed storylines. In: Proceedings of the 1st workshop on visualization for the digital humanities, VIS4DH'16
- Simaki V, Paradis C, Skeppstedt M, Sahlgren M, Kucher K, Kerren A (2017) Annotating speaker stance in discourse: the Brexit blog corpus. *Corpus Linguist Linguist Theory*. <https://doi.org/10.1515/cllt-2016-0060>
- Skeppstedt M, Paradis C, Kerren A (2016a) PAL, a tool for pre-annotation and active learning. *J Lang Technol Comput Linguist* 31(1):91–110
- Skeppstedt M, Sahlgren M, Paradis C, Kerren A (2016b) Active learning for detection of stance components. In: Proceedings of the workshop on computational modeling of people's opinions, personality, and emotions in social media at COLING'16, ACL, PEOPLES'16, pp 50–59

- Skeppstedt M, Simaki V, Paradis C, Kerren A (2017) Detection of stance and sentiment modifiers in political blogs. In: Proceedings of the international conference on speech and computer. Springer, SPECOM'17, pp 302–311. https://doi.org/10.1007/978-3-319-66429-3_29
- Tanahashi Y, Ma KL (2012) Design considerations for optimizing storyline visualizations. *IEEE Trans Vis Comput Graphics* 18(12):2679–2688. <https://doi.org/10.1109/TVCG.2012.212>
- Tory M, Möller T (2005) Evaluating visualizations: do expert reviews work? *IEEE Comput Graphics Appl* 25(5):8–11. <https://doi.org/10.1109/MCG.2005.102>
- Tufte ER (2006) *Beautiful evidence*. Graphics Press, Cheshire
- Tukey JW (1977) *Exploratory data analysis*. Addison-Wesley Publishing Company, Boston
- Wall E, Agnihotri M, Matzen L, Divis K, Haass M, Endert A, Stasko J (2019) A heuristic approach to value-driven evaluation of visualizations. *IEEE Trans Vis Comput Graphics* 25(1):491–500. <https://doi.org/10.1109/TVCG.2018.2865146>
- Wang X, Liu S, Chen Y, Peng TQ, Su J, Yang J, Guo B (2016a) How ideas flow across multiple social groups. In: Proceedings of the IEEE conference on visual analytics science and technology, IEEE, VAST'16, pp 51–60. <https://doi.org/10.1109/VAST.2016.7883511>
- Wang X, Liu S, Liu J, Chen J, Zhu J, Guo B (2016b) TopicPanorama: a full picture of relevant topics. *IEEE Trans Vis Comput Graphics* 22(12):2508–2521. <https://doi.org/10.1109/TVCG.2016.2515592>
- Wu Y, Wei F, Liu S, Au N, Cui W, Zhou H, Qu H (2010) OpinionSeer: interactive visualization of hotel customer feedback. *IEEE Trans Vis Comput Graphics* 16(6):1109–1118. <https://doi.org/10.1109/TVCG.2010.183>
- Wu Y, Liu S, Yan K, Liu M, Wu F (2014) OpinionFlow: visual analysis of opinion diffusion on social media. *IEEE Trans Vis Comput Graphics* 20(12):1763–1772. <https://doi.org/10.1109/TVCG.2014.2346920>
- Zhang L, Wang S, Liu B (2018) Deep learning for sentiment analysis: a survey. *Wiley Interdiscip Rev Data Min Knowl Discov* 8(4):e1253. <https://doi.org/10.1002/widm.1253>
- Zhao J, Gou L, Wang F, Zhou M (2014) PEARL: an interactive visual analytic tool for understanding personal emotion style derived from social media. In: Proceedings of the IEEE conference on visual analytics science and technology, IEEE, VAST'14, pp 203–212. <https://doi.org/10.1109/VAST.2014.7042496>