


Public Security Sentiment Analysis on Social Web: A Conceptual Framework for the Analytical Process and a Research Agenda

Victor Diogho Heuer de Carvalho, Universidade Federal de Alagoas, Brazil

 <https://orcid.org/0000-0003-2369-7317>

Ana Paula Cabral Seixas Costa, Universidade Federal de Pernambuco, Brazil

ABSTRACT

This article presents (1) the results of a literature review on social web mining and sentiment analysis on public security; (2) the idea of a framework for the analytical process involved in the literature review themes; and (3) a research agenda with a perspective for future studies, considering some elements of the analytical process. The literature review was based on searches of five databases: Scopus, IEEE Xplore, Web of Science, ScienceDirect, and Springer Link. Search strings were applied to retrieve literature material of four kinds, without defining an initial time milestone, to get the historical register of publications associated with the main thematic. After some filtering, primary and secondary findings were separated, enabling the identification of elements for the framework. Finally, the research agenda is presented, containing a set of three research artifacts related to the proposed framework.

KEYWORDS

Analytical Process, Framework, Literature Review, Public Security, Research Agenda, Sentiment Analysis, Social Web, Web Mining

INTRODUCTION

The great amount of information available through the social web signals to organizations an analytical perspective aligned with a social need to view human behavior to understand preferences, opinions, emotions, and feelings (Bjurstrom, 2015). The emergence of the concept of Big Data, related to an enormous amount of data in several formats and retrievable by various sources (Poletto, Carvalho, & Costa, 2017), only tends to reinforce this need. It also makes evident the tendency to combine artificial intelligence with data science (Sapountzi & Psannis, 2018).

While there are enormous amounts and varieties of data possible to retrieve and use—for instance, on decision-making (Tien, 2013)—the mining process is not trivial and requires a suitable technological toolbox for different purposes (Dobre & Xhafa, 2014).

In the organizational context, public services have many advantages to derive from both the massive amount of information and the analytical tools to ensure that stakeholders' expectations can

DOI: 10.4018/IJDSST.2021010101

This article published as an Open Access Article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

be met within the appropriate time, and generating processes are satisfactorily attended (Charalabidis, Koussouris, & Ramfos, 2011).

The organizational management is the main beneficiary since the related technological resources can make new knowledge discoveries, combine them with those already existing, and disseminating the results to the organization to promote continuous process improvement (Caione, Guido, Martella, Paiano, & Pandurino, 2016; Handzic, 2011). Knowledge discovery may apply in different data-structuring contexts, but it is important to emphasize that unstructured data are common in Big Data, and in this case, specific tools are required for web mining, text mining, and natural language processing (NLP) (Carvalho & Costa, 2019; Usai, Pironti, Mital, & Aouina Mejri, 2018).

Organizational knowledge management, decision support systems, big data, and data science have all become more integrated, and organizations, in turn, make use of this integration to drive strategic change. The main link among these elements is the need for information technology in the organization, promoting suitable means for sharing information and knowledge, even as they enable decision-making aligned with operational needs to ensure efficient troubleshooting (Navarro, Ruiz, & Peña, 2017; Wang & Noe, 2010).

One of the most critical sectors of public services that may benefit from data analysis is public security. The use of data from different sources (including the social web) and the application of analytical tools could lead, for instance, to the concept of the “smart city,” defined by Manjunatha & Annappa (2018) as a milestone of urban planning and development that integrates information technology with people’s routines, promoting sustainability and quality of life. Public security benefits from this integration insofar as it can make use of data such as past criminal registers, historical registered cases, and real-time information, enabling forecasting of events that threaten the public welfare.

Social web mining and sentiment analysis are areas from text mining dedicated to retrieve user-published records (social web mining) and to analyze these records to classify them according to polarity (sentiment analysis) expressed by the user (Kamel et al., 2010; He et al., 2015). Both have interesting toolkits that grant analytical power to be explored by areas such as public security to assess, for example, the level of satisfaction of people in social networks regarding the actions of policing, investigations, tracking of criminal activity, application, and maintenance of security policies in general (Carvalho & Costa, 2019). In this sense, related tools and their applications in public security deserve to be evidenced through a literature review identifying what is being applied and for what purposes, in other words, demonstrating what types of problems they are solving.

This article aims to: (i) present the results of a literature review about social web mining and sentiment analysis in public security; (ii) present the concept of a framework for the analytical process involved in the literature thematic; and (iii) define an agenda to support future research based on the analytical process. The emphasis is on the public-security area as a critical sector with absolute social repercussions, since it directly deals with the protection of people acting as social agents themselves, the main information providers, using several kinds of platforms (notably the social web).

The rest of this article is organized as follows: Section 2 presents the research procedure. Section 3 contains a summary of the literature findings. Section 4 contains the literature review. Section 5 proposes the framework. Section 6 establishes the research agenda. Finally, Section 7 contains the conclusions.

RESEARCH PROCEDURE

The reported research has a descriptive character, involving social web mining and sentiment analysis related to social impressions of public security. Its first part was a literature search to clarify which elements are necessary to formulate a framework for the analytical process involved.

Social web mining exists to extract people’s registers on several social web sources as weblogs, forums, personal websites, news sites, and social networks, revealing a field of computational

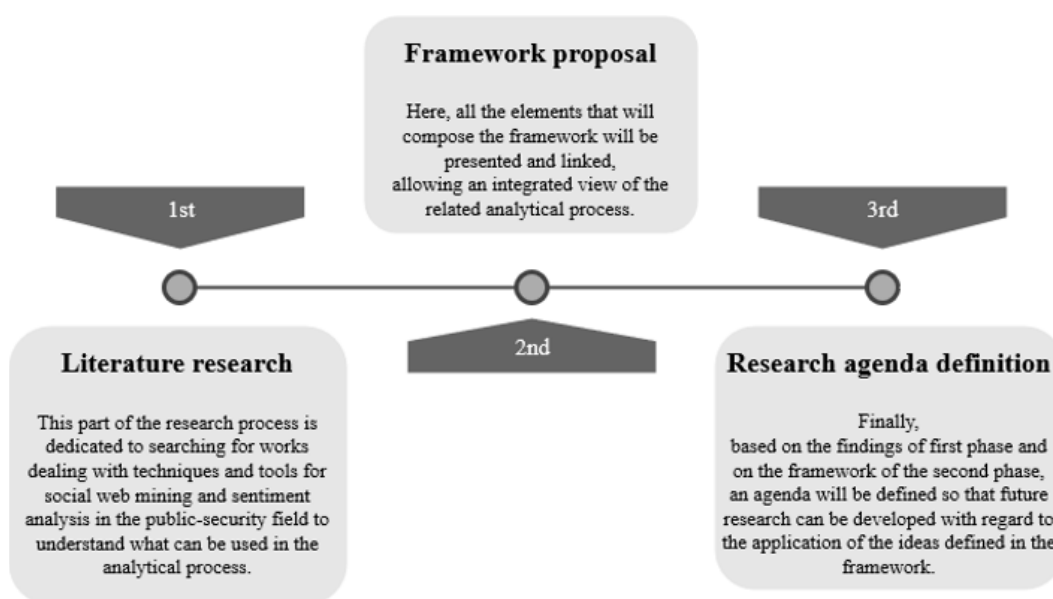
applications with abundant techniques (Ravi & Ravi, 2015; Stieglitz, Mirbabaie, Ross, & Neuberger, 2018). This makes the understanding of social web mining principles and purposes fundamental for using them adequately in the social field of interest (for instance, public security).

Figure 1 presents the flow of the research procedure, in three phases: first, literature research on the themes of interest, identifying available analytical techniques/tools and data sources; second, the proposal of the framework, based on the literature findings; and third, the definition of an agenda to support future research.

The first phase detail follows, still within this methodological section, and its results will be presented subsequently. The literature directly influenced the second and third phases, so the framework proposal and the research agenda will be presented after the literature review section.

Literature Research Procedure

Figure 1. Research flow in three phases



The literature review was developed based on the procedures described by Kitchenham and Charters (2007), Akter et al. (2019), and Sundermann, Domingues, Sinoara, Marcacini, and Rezende (2019). Definitions there should include (i) objectives, (ii) research questions, (iii) inclusion and exclusion criteria, (iv) search and selection strategies, and (v) filtering and evaluation of the findings. Figure 2 presents the related workflow.

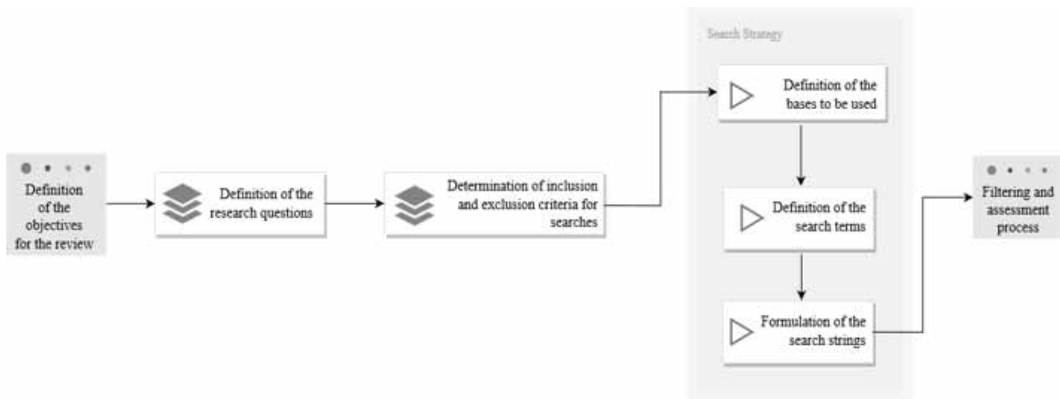
Objectives

To identify and select studies about techniques and tools for social web mining and sentiment analysis in the public-security field is the main objective of the literature review.

Research Questions

Q1: Which tools can be applied to retrieve data from the social web and to analyze the feelings related to public security?

Figure 2. Workflow for literature research and review



Regarding “tools”, the idea here is to identify methods, techniques, software, or any other type of technology that can be applied throughout the analytical process represented in the framework proposal. There are different tools for extracting textual records, preprocessing, identifying topics, and classifying sentiments, for example, so it is important to have a sense of what literature findings are applying for specific purposes.

Q2: Which data sources and informational elements should be considered for collection and analysis in the process?

In this case, the idea is to identify the data sources used by the literature findings and which informational elements were considered for application in the analysis. These elements, for example, can be user data (gender, age), date of publication of texts, and location.

Inclusion/Exclusion Criteria

These criteria were defined to limit the results, filtering them based on the language, kinds of materials, and publication period:

- Studies about web mining and sentiment analysis;
- Studies about public sentiment on security;
- Only studies in English;
- Kind of materials: journal articles, conference articles, books and book chapters;
- Period: all material published until April 2019;
- Access: material available to search through the Research Portal of the Brazilian Coordination for the Improvement of Higher Education Personnel and open-access materials.

Search and Selection Strategy

The following databases were used to ensure the quality of the search results: Scopus, Web of Science, Science Direct, IEEE Xplore, and Springer Link. For each database, using its specific notation, search strings with varying combinations of the terms were applied: “public security,” “social networks,” “social web,” “sentiment analysis,” and “opinion mining.” All the searches were done by “titles, abstracts, and keywords” (or the equivalent composition, depending on the database).

Findings Filtering

The findings filtering was done in a first moment using the initial search results to identify duplicated material, and since the number of findings was small, this was an easy task. Then, the literature findings were evaluated to identify which materials made a relevant contribution to knowledge about the research themes (i.e., web mining and sentiment analysis).

FINDINGS

Before the presentation of the literature review itself, it is interesting to present an overview of the results obtained by using the procedure previously described. Table 1 contains the strings applied with their respective results (i.e., the number of materials per database).

Regarding the kind of material, defined by the inclusion/exclusion criteria, four kinds of publications were searched. Noteworthy is that Springer Link categorizes book conference articles as well as book chapters. Figure 3 contains the percentage found for each kind.

Table 1. Strings and number of findings per database

Strings	Bases				
	Scopus	IEEE Xplore	Web of Science	Science Direct	Springer Link
String 1: “public security” AND “web mining” AND (“sentiment analysis” OR “opinion mining”)	0	4	0	0	2
String 2: “public security” AND (“social networks” OR “social web”) AND (“sentiment analysis” OR “opinion mining”)	1	12	0	1	20
String 3: “public security” AND (“social networks” OR “social web”) AND “web mining” AND (“sentiment analysis” OR “opinion mining”)	0	3	0	0	0

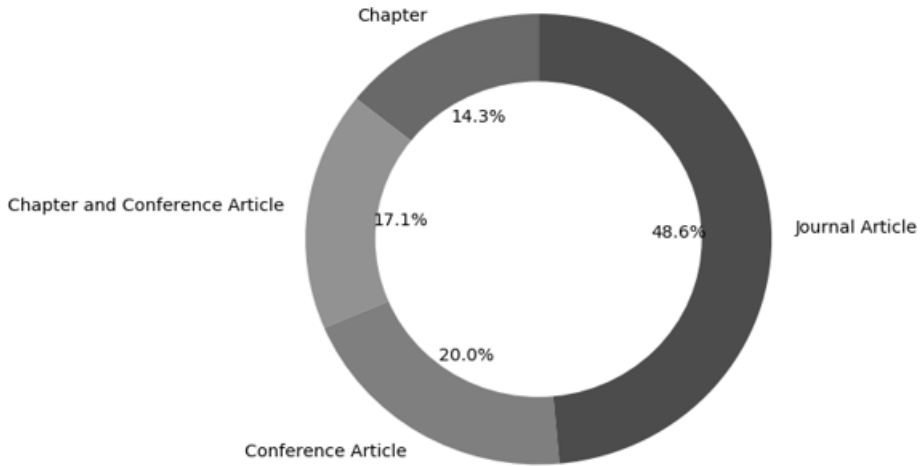
Journal articles predominate, followed by conference papers and book chapters. On the latter two kinds, the graph shows that the hybrid kind “chapter and conference article” is an intermediate element that, added to either “chapter” or “conference paper,” does not change the order of predominance.

The largest number of findings occurred from the Springer Link database, and it was easy to verify that there were no duplicates between the two strings that presented results. The total amount of findings from Springer Link was 22.

IEEE Xplore was the second-ranked database on the number of findings, with a total of 19. For this database, the strategy to eliminate the duplicates was based on a little Python script, using Pandas to handle data frames and Glob libraries to read multiple data files in comma-separated values (csv) format for the case reported here. The filtering resulted in 14 unique findings at the intersection of each string for this database. However, the search on IEEE Xplore returned two lists of content in conference proceedings that were eliminated, resulting in 14 unique findings.

Scopus and ScienceDirect each returned just one result and only for the second string, and it was also easy to detect that both results referred to the same finding. Web of Science did not return any findings for the three strings.

Figure 3. Percentages per kind of findings

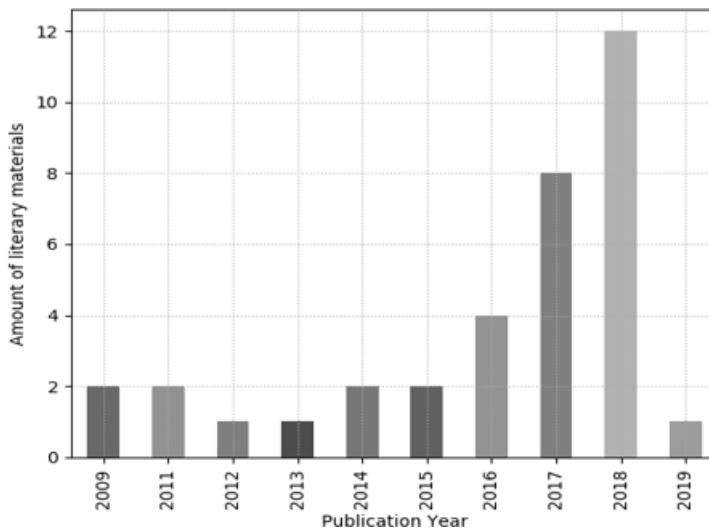


With thirty-five findings (twenty-two from Springer Link, twelve from IEEE Xplore and one from Scopus/ScienceDirect), new analyses were performed. Figure 4 presents the count of findings per year.

The year 2018 presented the largest number of literature items (twelve findings) about the researched themes, followed by 2017, with eight findings. The third was for 2016 (four findings), followed by 2009, 2011, 2014, and 2015 (each with two findings). Finally, the smaller amounts were for 2012, 2013, and 2019, each with only one finding. Since no initial time/period restrictions were applied, 2009 can be taken as the starting point for the publications on the thematic of the present article.

The last analysis filtered the materials for their accessibility. Although the searches had presented a last filtered amount of 35 findings, not all materials were accessible by the means defined in the inclusion/exclusion criteria. So, even though their abstracts could partially contain some findings, it was judged that using only the information contained in these abstracts did not guarantee an analysis

Figure 4. The number of findings per year



that would be sufficiently good to define the alignment of the works with the research. Based on this, the literature review was done using a final count of twenty-four texts. Table 2 presents the classification of these findings, according to their kind.

Table 3 presents the count related to the databases where the materials were found.

Figure 5 presents the count of the ten most frequent keywords related to these 24 final findings.

The final count revealed a total of 111 keywords, and the first six were the most frequent terms since from the seventh until the last one, the frequency is unitary. A word cloud also could be elaborated based on the keywords, as Figure 6 shows.

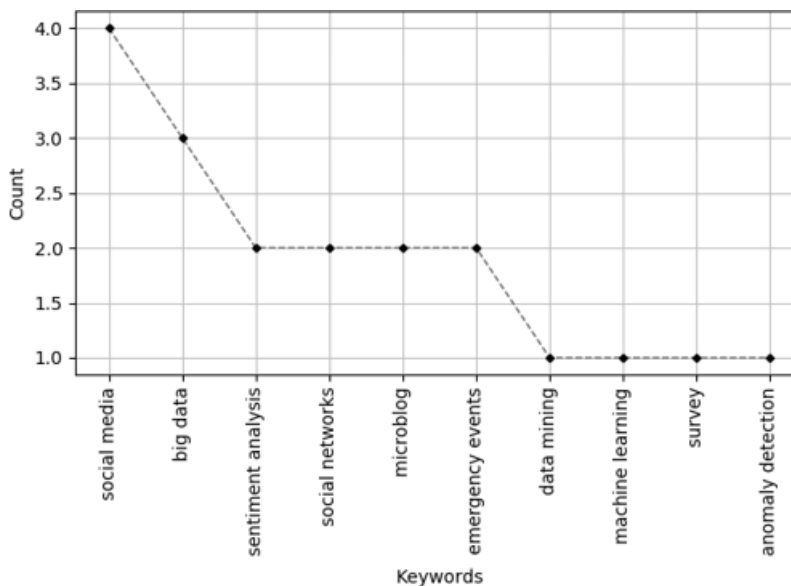
Table 2. The number of accessible findings per kind

Kind	Amount
Journal article	16
Book chapter	1
Conference article	7
Total	24

Table 3. The number of accessible materials per database

Base	Amount
Springer Link	11
IEEE Xplore	12
Scopus/ScienceDirect	1
Total	24

Figure 5. The ten most frequent keywords related to the final twenty-four findings



networks, and web 3.0, responsive to users' actions and able to perform forecasting and make suggestions. These authors suggest the presence of open and connected repositories to be used in the creative process.

Predicting the social-web user's behavior has received attention from public-security authorities aware that digital communication platforms are a means of data extraction, to indicate behavior trends that aid in anticipating actions and avoiding risk situations (Cui et al., 2016).

In this field of behavior prediction, the bibliographic research found interesting studies:

- Lin, Mao, and Zeng (2017) worked on refining the classification of feelings based on personality trends, using data extracted from a Chinese microblog service;
- Bao et al. (2009) proposed a model based on the association of emotion and terms or topics of users to perform socio-affective text mining;
- Jin, Zhang, and Zhang (2018) formulated an index to assess the negativity of users' moods related to events receiving comment on social networks;
- Chen, Cai, Huang, and Jiao (2016) worked with the identification of users hired to make floods of posts on social networks, based on their behavior patterns.

In parallel to public security, the literature aided to detect the transportation area as deserving of the attention of researchers on sentiment analysis. For instance, Cao, Wang, and Lin (2018) developed a study using semantic information extracted from Chinese social networks to support the analysis of vehicle traffic conditions based on users' registers. The authors highlight that threatening situations may be detected according to the conditions and kinds of traffic occurrences reported by users. Lv, Chen, Zhang, Duan, and Li (2017) carried out a bibliographical survey of the transportation area and detected that social web/social networks, web mining, text mining, and sentiment analysis are recurrent topics.

Big Data and Internet of Things are themes of great interest in urban planning and development, mainly related to the idea of "smart cities." According to Manjunatha and Annappa (2018), public security, transportation/traffic control, and monitoring, health, agricultural production, and public governance are where potential users of analytical tools typically associated with data science and analysis should be sought.

Xu et al. (2017) developed an interesting study that presented a method based on spatial-temporal graphs to describe semantic relations among various concepts. Their study provides an analytical methodology for both punctual analyses as well as analytical software development. The authors sought to validate the method's accuracy through an experiment using a movie database. The application of spatial-temporal graphs in the experiment was based on the semantic relation "person co-acted with person" in movies. An accuracy rate of 96.7% was verified using the extracted data.

Other media than texts for use in the analysis were also detected. For instance, techniques such as cluster analysis may be used to extract characteristics of images available on the social web (Wazarkar & Keshavamurthy, 2019). This kind of technique can help in detecting social events or suspect activities registered on users' images and videos. Peng et al. (2017) surveyed state-of-the-art techniques, based on the joint analysis of several kinds of media describing a kind of application (iMonitor) with which some intelligent systems have been designed and implemented to enable media monitoring: PRISM (United States), Tempora (United Kingdom), and Golden Shield (China).

Zheng et al. (2017) comment that there are still many unsolved problems, highlighting those related to machine understanding of the fuzzy elements on human language, especially to detect and avoid risks, mentioning possible applications in criminal justice.

Primary Literature

Since the secondary findings highlighting the wide range of analytical tools and strategies appeared at the opening of the literature review, now the focus shifts to the primary findings and applications

directly associated with public security. Table 4 presents a synthesis of the identified articles sorted as primary literature about social web mining and sentiment analysis applied to public security.

The table notes that the primary findings have their techniques, methods, or technologies distributed according to the typology commented upon by Yue et al. (2018), and the social web mining is present as a feeder for the set of techniques. The typology is divided into the following sets:

- Natural language processing (NLP) set contains techniques for the extraction of useful information from unstructured texts as data sources, and the sentiment analysis is a typical example of a problem that may be solved using these techniques (Lv et al., 2017; Sun et al., 2018).
- Machine learning set contains techniques that enable computers to learn about a specific subject and make inferences analogously to humans. Support vector machines (SVM), naïve Bayes classifiers, maximum entropy method, neural networks, and deep learning, are machine learning techniques that gained notoriety in the literature about sentiment analysis, as they allow better refinement of the data and the process results (Yang & Chen, 2017; Song, Kim, Lee, Kim, & Youn, 2017).

The next subsections will present a synthesis based on the primary findings in Table 4, using this typology.

Natural Language Processing and Computational Linguistics

The article by Zhang et al. (2018) presents applications of these tools in the context of public security for emergencies, aiming to create a comprehensive lexicon of emotions, with data extracted from a Chinese microblog social network. Initially, the authors relied on an existing ontological dictionary from which they extracted words related to emotional situations used on the network. The Word2Vec tool was used to transform the messages/postings into a vector of words, applying the SO-PMI algorithm to calculate the value of pointwise mutual information in a new word, and another defined for comparison to get the emotional tendency of the new word in the process. This procedure was performed several times, enabling the construction of the lexicon. The final value of the text sentiment was left to the Language Technology Platform, which implements automatic routines for processing and treatment of input texts. While NLP is the basis of the initially described procedure, the authors also applied machine-learning concepts to achieve refinements.

Tang (2013) presents applications relating to the concept of “harmonious society,” proposed by Chinese leaders to define a balanced society in terms of the number of people, their quality of life, and the balanced distribution of wealth. Several indexes can be used to measure harmony (authors cite examples such as the Gini coefficient, the index of prosperity related to the Legatum Institute, the happiness index, green GDP, and green GDP quality index). The final proposition of the authors was measuring the perception of online societal risks as a means of detecting societal hazards, according to a codification done by the Institute of Psychology of the Chinese Academy of Science. The use of computational linguistics and NLP is noteworthy, as is the use of users’ registers mined from social-web channels, such as microblogs, to identify and sort potential societal risks through postings on these media.

The study by Jin et al. (2017) was applied to the social networks’ users in emergency events. The objective of the work was to provide references so that the emergency services could monitor public opinion to aid the government in dealing with emergencies. Four periods define an emergency event, using an earthquake as the basis for the analysis.: incubation, development, decline, and calm. These periods reflect how public opinion manifests itself in response to warnings about the emergency events disseminated on the social web. Although the article does not mention the techniques applied, only reporting the application of sentiment analysis/opinion mining, in the theoretical development of the material are listed techniques/methods such as signal analysis theory, intuitionistic fuzzy inference, Bayesian networks, and the fuzzy comprehensive method.

Table 4. Primary findings, with specific application to public security

Authors	Main Application	Techniques/ Technologies	Research Problem	Data Sources	Elements for Analysis
Sun, Zhang, Ding, & Quan (2018)	Detecting abnormal opinions, sentiments, patterns, or aspects of such patterns	Convolutional Neural Network Long and Short-Term Memory for Sentiment Analysis using the multivariate Gaussian model	The main problem in this work is to assess the method applied, demonstrating how accurate it is and validating its applicability for detecting abnormal behaviors on social networks.	Users of a social network	Messages shared on the social network
Leventakis & Kokkinis (2018)	Community policing	<ul style="list-style-type: none"> - Geospatial processing and analysis - Semantic information extraction by multimedia analysis - Sentiment analysis - Case-based reasoning - Data warehousing and Decision Support Systems 	This work presents a set of modules developed within a project dedicated to community policing. The main problem here is to identify what are the technological elements to support the related policing actions.	<ul style="list-style-type: none"> - Users of dedicated mobile app and web portal - Community policing communications channels in the social web 	Messages shared via mentioned channels
Zhang, Zhu, & Wang (2018)	Public emergency events	<ul style="list-style-type: none"> - Word2Vec - SO-PMI algorithm - Social web mining - Language Technology Platform (LTP) - Recurrent Neural Network (RNN) 	This work is concerned with some issues related to the existing lexicons for sentiment classification. Based on these issues, the article's problematic is to propose an alternative procedure using the mentioned technologies, measuring its accuracy.	<ul style="list-style-type: none"> - Users of a social network - NLP&CC2013 dataset 	Messages shared on the social network
Yang, Liu, Liu, & Cui (2018)	Predictions about the occurrences of collective actions	<ul style="list-style-type: none"> - Social web mining - Sentiment analysis - Collective Emotional Contagion Model - Deep neural networks (DNN) - TensorFlow (to implement DNN) 	Here, the main problem is the prediction of the probability of collective actions every day in a next month, ensuring early decision-making about them.	<ul style="list-style-type: none"> - Users of social networks - Arab Spring dataset 	Messages shared on the social network
Tang (2013)	<ul style="list-style-type: none"> - Harmonious society measurement - Societal risks identification 	<p>The work identifies some forms to measure a harmonious society:</p> <ul style="list-style-type: none"> - Gini coefficient - Legatum Prosperity Index - Happiness Indices - GDP Quality Index - Harmony Indices - Tightness Score <p>On the other hand, it identifies computational techniques to measure the risk perception:</p> <ul style="list-style-type: none"> - Computational linguistics/ NLP 	This work defines the measurement of a harmonious society as its main problem, seeking to identify indices for this to be possible.	Social Web in general	Words extracted from people's messages/ texts on the social web
Zuo, Wu, Zhang, Wang, & Xu (2018)	Public opinion analysis, with an application in public security	<ul style="list-style-type: none"> - CAMEL* that uses maximum entropy and latent Dirichlet allocation - Alternatively, a non-parametric alternative to CAMEL, that uses - coupled Dirichlet Processes *Cross-collection Auto-labeled Max Ent-LDA 	This work states cross-media aspect-opinion mining as a challenge to be addressed since there are various types of online media. The main problem here is how to deal with the cross-media aspect-opinion mining.	<ul style="list-style-type: none"> - Social web - News portals 	<ul style="list-style-type: none"> - Online reviews (for experimental purpose) - News texts - Messages shared on microblogs service
Xu et al. (2016)	Detect and describe urban emergency events	Spatial-temporal mining using heuristics associated with the 5W Model (When, Why, What, Where, Who)	The main problem in this work is to provide a good description of an urban emergency event.	Users of a social network	Messages shared on the social network
Jin, Li, Fu, Zhang, & Ding (2017)	Governmental emergency events management	<ul style="list-style-type: none"> - Social web mining - Opinion mining 	The central problem explored in this article is how the government uses opinions on microblogging services to generate alerts, orientations, and promote control in emergencies.	Users of a social network	Messages shared on the social network during the life cycle of emergency events
Koot et al. (2014)	Support for security officers	<p>Web crawling and data extraction using Python's frameworks and libraries, namely:</p> <ul style="list-style-type: none"> - Django for web development - OAuth 2.0 for social networks authentication - BeautifulSoup for HTML parsing - urllib for URL handling - AJAX for the front-end 	The problem explored in this paper is the use of social network analysis tools to partially automate the activities of security agents in extracting information from user profiles, assisting, for example, criminal investigations.	Users of multiple social networks	Personal data in social networks users' profiles

Koot et al. (2014) developed a study to support public-security agents with intelligent analysis using the information provided by the Internet, specifically by social networks. The authors proposed an analytical tool architecture for the public profile of social networks' users to perform web mining. They describe all the technological framework for the tool development: Python language and related technologies, such as OAuth, to manage social networks' API authentications, BeautifulSoup for HTML parsing, urllib to handle the URLs, Ajax (another programming language) for the front-end development, and all this combined through the object-oriented framework Django, widely applied in web development. They also point out the interest in engaging NLP tools for analyzing sentiments, modeling topics, analyzing trends, and providing results visualization.

Leventakis and Kokkinis (2018) deal with applications for communitarian policing identifying technological elements able to support policial decisions and actions for public-security maintenance. In their article, they present the INSPC²T project, developed by the European Union, as the main goal for presenting researched contributions to enabling the collaboration of police and society on several levels. The project includes modules related to mobile applications and web portals for citizens' use; events geospatial processing; multimedia analysis; business intelligence (notably with the use of web mining and NLP for sentiment analysis); the application of case-based reasoning to develop expert systems; measuring data aging; application of data warehousing and decision support systems; use of security portals; and simulating trainings for communitarian policing.

Xu et al. (2016) propose in their article a method for urban emergency events, using social networks as the main data sources, specifically collecting users' messages. The method is based on the "five W" questions—who, where, why, when, and what—to look for spatial-temporal information (i.e., spatial-temporal mining) and semantic information provided by the users' crowd in digital social networks. In this work, the authors do not name any NLP technique, but its use is supposed by the semantic nature mentioned in the method's architecture.

Machine Learning

Sun et al. (2018) sought to detect anomalous behavior using social-network users' messages. In order to verify the behavior pattern, the authors applied analysis involving the Gaussian multivariate distribution, using deep learning based on convolutional neural networks to make the detection. Zhang et al. (2018) used recurrent neural networks to refine the treatment that started with NLP. The objective, in this case, was to improve the sentiments classifications over word vectors extracted from users' registers on social networks.

Yang et al. (2018) present in their work a form to make forecasts about collective actions using deep learning and communications on social media. They define certain kinds of collective actions as possible threats to social welfare, containing several risks, especially when they occur during social conflicts, which can affect economic, social, and public-safety levels. Therefore, using deep neural networks, the authors model a forecasting framework for collective actions whose data sources are social media. In this case, deep neural networks were applied to perform binary classification, based on a training set. The work also uses the models as the species competitions and epidemic propagation, to make predictions about the emotional contagion leading to collective events on social media.

The work of Zuo et al. (2018) is the last of the primary findings that describe the application of analytical tools to public security. This material purposed the Cross-collection Auto-labeled MaxEnt-LDA (CAMEL), described as a novel model that makes use of the maximum entropy method and the Dirichlet latent allocation to do complimentary aspects-based opinion mining in asymmetric collections. The work also presents a nonparametric proposal based on Dirichlet processes for estimating a specific number of topics, which would be extremely difficult using only the CAMEL (parametric).

FRAMEWORK FOR THE ANALYTICAL PROCESS

The literature review made it possible to understand the technological trends associated with the researched themes, helping to obtain answers to the two research questions presented in the research-procedure section. The answers are the basis for the framework construction related to the analytical process.

Which Tools Can Be Applied to Retrieve Data From the Social Web and To Analyze the Feelings Related to Public Security?

The understanding of which analytical tools should be used for the composition of the framework revolves around what can be defined as analytical moments, following initial indications realized by Stieglitz et al. (2018). The first moment within the framework is the choice of the data sources, i.e., which kinds of social media should be used for mining. This determination aims to establish whether the search will focus on (for example) news sites, e-commerce sites, blogs, search engines, social networks, or combined types. Once the basic typology is defined, the specific platforms can be selected—for example, in the case of social networks, platforms such as Twitter, Facebook, YouTube, or LinkedIn.

The second moment is the search for and constant extraction of raw data. Search and extraction strategies will depend on the previous choice since each platform will have a specific API. Consequently, this strategy will define which technological tools to use. Fundamentally, web crawling and scraping are the starting strategies for the mining and analytical process, dealing here with the web-mining process (Sapountzi & Psannis, 2018). The strategy for storing extracted (unstructured) raw data should be established and big data schemas may be chosen if the amount of data demands it (Manjunatha & Annappa, 2018; Xu et al., 2016).

The extracted and stored data are the starting point for the third moment within the framework, where the raw data will be preprocessed to ensure that they receive a suitable format for subsequent processing. The application of NLP-based preprocessing techniques, such as tokenization, stemming, stop-words removal, extraction, and weighting of characteristics, is emphasized (Ravi & Ravi, 2015; Yang & Chen, 2017). The preprocessed data, already in a format adjusted for analysis, should be stored in adequate databases to enable the sequence.

The fourth moment contains the routines for refining the preprocessed data (Zhang et al., 2018), enabling the sentiment analysis in the fifth moment. In this part of the framework, the machine-learning techniques enter to find hidden patterns in the treated and formatted data (Sun et al., 2018) as well as to perform topics modeling and extract only textual registers related to the desired topics (Bao et al., 2009). The refined data also require proper storage, ensuring their availability for the next moments.

On the fifth moment, the sentiment analysis uses the refined data, the main goal of the framework. As mentioned, this analysis is a typical NLP problem; however, the machine-learning techniques may be used here too. The fundamental result of this moment is to obtain measures of polarity and subjectivity related to sentiment about a theme (Zhang et al., 2018). Indicators can also be obtained on the robustness of the analysis of feelings: accuracy, precision, recall, and the F1 score (Hailong, Wenyan, & Bo, 2014).

The sixth and final moment is dedicated to the results visualization, user-friendly interfaces presenting the classification/pattern of the extracted sentiments (Koot et al., 2014; Leventakis & Kokkinis, 2018).

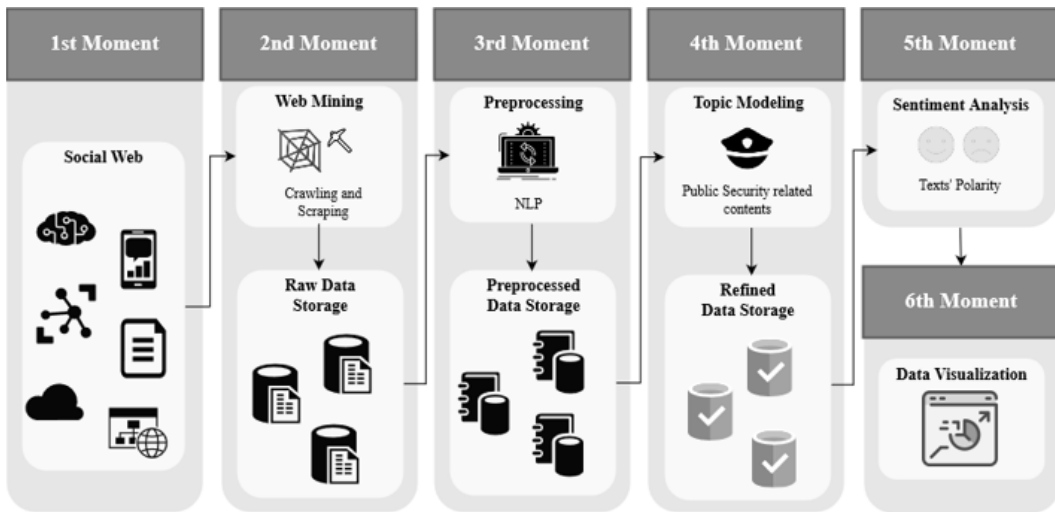
Which Data Sources and Informational Elements Should Be Considered for Collection and Analysis in the Process?

Virtually all the literature findings easily answer this question. First, data sources are the users themselves who feed the social web with their records in a variety of formats, but mainly through social networks' posts. Second, the informational element for sentiment analysis is the user's textual

register—for instance, the shared messages in social networks, expressing in an unstructured way the user’s sentiment about some theme or event that creates repercussions on the social web. The whole process, described in answer to the first question, revolves around users’ registers, so the entire process is about their extraction, treatment, and refinement, as well as the analysis to measure the expression of the feelings.

Figure 7 presents the schema of the framework, following the moments described in answer to the first question.

Figure 7. Framework for the analytical process



Fairly close approaches also appear in works such as those by Charalabidis et al. (2011), Koot et al. (2014), Lin et al. (2017), Manjunatha and Annappa (2018), and Xu et al. (2016), however, it should be clear that the proposed framework is dedicated to sentiment analysis related to public security issues. This specification is evident in its fourth moment, where the main refinement of the process is devoted to topic modeling, helping to identify textual records on public security and related topics.

The proposed framework provides methodological guidelines for future applications, enabling (i) the development of sentiment analysis on public security and (ii) the development of a dashboard for use by the agents of public-security management. With that, it only remains to define a research agenda, presenting topics for future research.

RESEARCH AGENDA

From the literature review and the framework, this section presents ideas for applications related to the analysis of feelings dedicated to public security. Among the possible forms of presentation of research agendas, elements related to Design Science in Information Systems research was chosen (Hevner, March, Park, & Ram, 2004), where guidelines are defined for conducting structured and systemic research with well-defined objectives; and a structure for future proposals, based on the article by Loebbecke and Picot (2015).

Artifacts—elements related to the analytical process—could be identified as referrals within the research agenda, indicating possibilities of advances in the applications treated from the theoretical and conceptual point of view. Three possible artifacts for future developments are described below.

Artifact 1: Analytical Process Addressed to Governmental Agencies of Public Security and Safety Management

Notably, the studies of Charalabidis et al. (2011) and Manjunatha and Annappa (2018) point out that certain nations have been developing for some time projects designed to carry out diverse analyses based on the social web. They aim to verify opinions or evaluate popular sentiment about events that may have social repercussions for the actions of public-management agencies, regardless of the orientation of these projects.

The framework defined above shows an analytical process that comprises seeking technical and technological tools for mining the social web, preprocessing raw data collected and necessary refinements leading to sentiment analysis, generating reports for government agencies in public-security management, and allowing the development of systems with interfaces dedicated to the interactive and configurable presentation of the results by decision-makers.

Artifact 2: Decision Dashboard for Configuration of Sentiment Analysis Parameters and Results Visualization

This artifact will specifically require the combination of web mining, artificial-intelligence tools, software engineering, and interface design to construct a sentiment analysis/opinion-mining decision-support dashboard on social-networking platforms. Some of the findings of the bibliographic research also point out developments in this regard, namely, the works of Manjunatha and Annappa (2018) and Peng et al. (2017).

The dashboard aims to reflect and expand the possibilities associated with the previous artifact (analyses for government agencies in public-security management). This artifact is intended for public managers; however, the provision of information open to the general population should also be considered, as it is the main recipient of government actions, as well as the primary provider of information needed for such analyses.

Artifact 3: Spatial and Temporal Analysis of the Sentiment Trends About Public Security

The third and last artifact is the specific development of spatial and temporal analyses of the general feeling related to public security, following the tendencies that Leventakis and Kokkinis (2018) and Xu et al. (2017) point out. This artifact could be linked to the previous two, but its highlight is the incorporation of interfaces with geographic information systems and the use of georeferenced information combined with sentiment analysis (Reinoso, Farooq, & Forum, 2015; Song & Xia, 2016). Various intersections can emerge, for example, the discovery of patterns of crime occurring in certain regions (Nepomuceno & Costa, 2019), with the expression of the population's feelings from georeferenced records in social networks.

CONCLUSION

This paper aimed to achieve three interconnected goals: (i) review of the literature on analytical tools applied to understand people's feelings/sentiments about public security; (ii) design of an analytical framework for analyzing feelings/sentiments about public security, according to perceived trends in the literature findings; and (iii) definition of an agenda for presenting to the community the developments related to the theme, indicating elements (artifacts) for further research, generated from the concepts searched with specific application to public security.

The literature search results may seem rather restrictive when compared to other researches (mainly systematic reviews). However, there are two interpretations of this perception. First, protocol definitions were quite restrictive. Additions to the inclusion and exclusion criteria could have been made, just as more databases could have been used to perform the searches, and other configurations

could have been devised within the strings to increase the number of findings. Second, there is a gap in the research to be filled in terms of real-world applications of sentiment-analysis tools concerning public security, which signals the positive path that this research initiated and can take from here.

Regarding the first interpretation, although terms referring directly to analytical tools and data sources have been used, the use of the term “public security” may have been restrictive. Other terms associated with the theme could also result in a greater number of literature findings; a systematic review could also be considered as an extension of the research agenda.

Although it does not add new features to the web mining or sentiment analysis fields, the proposed framework is fundamental for the description and fulfillment of the defined agenda, representing the basic concepts to be used, as well as providing guidelines for the selection of specific tools for upcoming developments, among those that were presented in the literature review. This research tends to consider this framework as a divider of the theoretical-conceptual from the practical moments, the latter of which will be the fruit of its unfolding. The research can also determine that its structure is conceptually satisfactory since it is aligned with the tendencies of the associated areas, as well as other works reported in the literature review, pointing to developments with methods and architectures similar to the one used in the framework proposed here.

The agenda presents three interconnected possibilities for future research, ensuring that researchers have a vision of niches to be worked with government agencies in public-security management and providing decision-support tools for the formulation of strategies aligned with social needs in this area.

The framework proposed in this article should be applied to collect text from social network users in order to extract from these texts those related to public security issues such as policing, violence and crime control, and other actions. This application will be made in the metropolitan region of Recife (Pernambuco, Brazil). With the extracted texts, sentiment analysis will be applied to initially provide local public security management agencies with feedback on how people judge security actions taken. Secondly, a dashboard will be delivered so that both the authorities and the interested citizens can hold consultations. This instrument is expected to serve as a support tool for public security decisions, adding public opinion directly to the associated planning.

FUNDING

This work was financed in part by the Coordination for the Improvement of Higher Education Personnel (Brazil) – Finance Code 001, and by the National Council for Scientific and Technological Development (Brazil).

REFERENCES

- Akter, S., Bandara, R., Hani, U., Fosso Wamba, S., Foropon, C., & Papadopoulos, T. (2019). Analytics-based decision-making for service systems: A qualitative study and agenda for future research. *International Journal of Information Management*, 48(January), 85–95. doi:10.1016/j.ijinfomgt.2019.01.020
- Bao, S., Xu, S., Zhang, L., Yan, R., Su, Z., Han, D., & Yu, Y. (2009). Joint Emotion-Topic Modeling for Social Affective Text Mining. In *2009 Ninth IEEE International Conference on Data Mining* (pp. 699–704). IEEE. doi:10.1109/ICDM.2009.94
- Bisio, F., Meda, C., Zunino, R., Surlinelli, R., Scillia, E., & Ottaviano, A. (2015). Real-time monitoring of Twitter traffic by using semantic networks. *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 966–969. doi:10.1145/2808797.2809371
- Bjurstrom, S. (2015). Sentiment Analysis Methodology for Social Web Intelligence. In *Proceedings of the Twenty-first Americas Conference on Information Systems* (pp. 1–12). Puerto Rico: Association for Information Systems.
- Caione, A., Guido, A. L., Martella, A., Paiano, R., & Pandurino, A. (2016). Knowledge base support for dynamic information system management. *Information Systems and e-Business Management*, 14(3), 533–576. doi:10.1007/s10257-015-0294-3
- Cao, D., Wang, S., & Lin, D. (2018). Chinese microblog users' sentiment-based traffic condition analysis. *Soft Computing*, 22(21), 7005–7014. doi:10.1007/s00500-018-3293-8
- de Carvalho, V. D. H., & Costa, A. P. C. S. (2019). Social Web Mining as a Tool to Support Public Security Sentiment Analysis. In P. S. Freitas, F. Dargam, R. Ribeiro, J. M. Moreno, & J. Papathanasiou (Eds.), *5th International Conference on Decision Support System Technology* (pp. 164–169). Funchal: EURO Working Group on Decision Support Systems. Retrieved from https://icdsst2019.files.wordpress.com/2019/06/icdsst2019_proceedings_final.pdf
- Charalabidis, Y., Koussouris, S., & Ramfos, A. (2011). A cloud infrastructure for collaborative digital public services. *Proceedings - 2011 3rd IEEE International Conference on Cloud Computing Technology and Science, CloudCom 2011*, 340–347. doi:10.1109/CloudCom.2011.53
- Chen, G., Cai, W., Huang, J., & Jiao, X. (2016). Uncovering and Characterizing Internet Water Army in Online Forums. In *2016 IEEE First International Conference on Data Science in Cyberspace (DSC)* (pp. 169–178). IEEE. doi:10.1109/DSC.2016.77
- Cui, P., Liu, H., Aggarwal, C., Wang, F., Cao, L., Yu, P. S., & Qiang, Y. et al. (2016). Uncovering and Predicting Human Behaviors. *IEEE Intelligent Systems*, 31(2), 77–88. doi:10.1109/MIS.2016.37
- Dobre, C., & Xhafa, F. (2014). Intelligent services for Big Data science. *Future Generation Computer Systems*, 37, 267–281. doi:10.1016/j.future.2013.07.014
- Hailong, Z., Wenyan, G., & Bo, J. (2014). Machine learning and lexicon based methods for sentiment classification: A survey. *Proceedings - 11th Web Information System and Application Conference, WISA 2014*, 262–265. doi:10.1109/WISA.2014.55
- Handzic, M. (2011). Integrated socio-technical knowledge management model: An empirical evaluation. *Journal of Knowledge Management*, 15(2), 198–211. doi:10.1108/13673271111119655
- He, W., Wu, H., Yan, G., Akula, V., & Shen, J. (2015). A novel social media competitive analytics framework with sentiment benchmarks. *Information & Management*, 52(7), 801–812. doi:10.1016/j.im.2015.04.006
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *Management Information Systems Quarterly*, 28(1), 75. doi:10.2307/2514865
- Jin, R., Zhang, H.-L., & Zhang, Y. (2018). The social negative mood index for social networks. In *2018 IEEE Third International Conference on Data Science in Cyberspace (DSC)* (pp. 1–5). IEEE. doi:10.1109/DSC.2018.8570298
- Jin, X., Li, H., Fu, H., Zhang, J., & Ding, B. (2017). A study on early-warning response of emergency events network public opinion based on analysis of Microbloggers' Behaviors. *PIC 2016 - Proceedings of the 2016 IEEE International Conference on Progress in Informatics and Computing*, 753–756. doi:<ALIGNMENT.qj></ALIGNMENT>10.1109/PIC.2016.7949599

- Kamel, M. N., Sanfilippo, A. P., Corley, C. D., & Wheeler, S. (2010). Social Web mining and exploitation for serious applications: Technosocial Predictive Analytics and related technologies for public health, environmental and national security surveillance. *Computer Methods and Programs in Biomedicine*, 100(1), 16–23. doi:10.1016/j.cmpb.2010.02.007 PMID:20236725
- Kitchenham, B., & Charters, S. (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering*. EBSE Technical Report EBSE-2007-01.
- Koot, G., Veld, M. A. A. H. I. T., Hendricksen, J., Kaptein, R., De Vries, A., & Van Den Broek, E. L. (2014). Foraging online social networks. *Proceedings - 2014 IEEE Joint Intelligence and Security Informatics Conference, JISIC 2014*, 312–315. doi:10.1109/JISIC.2014.62
- Leventakis, G., & Kokkinis, G. (2018). Developing and Assessing Next Generation Community Policing Social Networks with THOR Methodology. In G. Leventakis & M. R. Haberfeld (Eds.), *Community-Oriented Policing and Technological Innovations* (pp. 47–62). Springer International Publishing. doi:10.1007/978-3-319-89294-8_6
- Lin, J., Mao, W., & Zeng, D. D. (2017). Personality-based refinement for sentiment classification in microblog. *Knowledge-Based Systems*, 132, 204–214. doi:10.1016/j.knsys.2017.06.031
- Loebbecke, C., & Picot, A. (2015). Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda. *The Journal of Strategic Information Systems*, 24(3), 149–157. doi:10.1016/j.jsis.2015.08.002
- Lv, Y., Chen, Y., Zhang, X., Duan, Y., & Li, N. L. (2017). Social media based transportation research: The state of the work and the networking. *IEEE/CAA Journal of Automatica Sinica*, 4(1), 19–26. <ALIGNMENT.qj></ALIGNMENT>10.1109/JAS.2017.7510316
- Manjunatha, & Annappa, B. (2018). Real Time Big Data Analytics in Smart City Applications. In *2018 International Conference on Communication, Computing and Internet of Things (IC3IoT)* (pp. 279–284). IEEE. <ALIGNMENT.qj></ALIGNMENT>10.1109/IC3IoT.2018.8668106
- Navarro, J. L. A., Ruiz, V. R. L., & Peña, D. N. (2017). The effect of ICT use and capability on knowledge-based cities. *Cities (London, England)*, 60, 272–280. doi:10.1016/j.cities.2016.09.010
- Nepomuceno, T. C. C., & Costa, A. P. C. S. (2019). Spatial visualization on patterns of disaggregate robberies. *Operations Research*, (0123456789). Advance online publication. doi:10.1007/s12351-019-00479-z
- Peng, Y., Zhu, W., Zhao, Y., Xu, C., Huang, Q., Lu, H., Zheng, Q., Huang, T., & Gao, W. (2017). Cross-media analysis and reasoning: Advances and directions. *Frontiers of Information Technology & Electronic Engineering*, 18(1), 44–57. doi:10.1631/FITEE.1601787
- Poleto, T., de Carvalho, V. D. H., & Costa, A. P. C. S. (2017). The Full Knowledge of Big Data in the Integration of Inter-Organizational Information. *International Journal of Decision Support System Technology*, 9(1), 16–31. doi:10.4018/IJDSST.2017010102
- Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowledge-Based Systems*, 89, 14–46. doi:10.1016/j.knsys.2015.06.015
- Reinoso, G., Farooq, B., & Forum, C. T. R. (2015). Urban Pulse Analysis Using Big Data. In *Canadian Transportation Research Forum 50th Annual Conference* (p. 16). Montreal: Transportation Association of Canada (TAC). Retrieved from <https://trid.trb.org/view/1417784>
- Sapountzi, A., & Psannis, K. E. (2018). Social networking data analysis tools & challenges. *Future Generation Computer Systems*, 86, 893–913. doi:10.1016/j.future.2016.10.019
- Song, J., Kim, K. T., Lee, B., Kim, S., & Youn, H. Y. (2017). A novel classification approach based on Naïve Bayes for Twitter sentiment analysis. *Transactions on Internet and Information Systems (Seoul)*, 11(6), 2996–3011. doi:10.3837/tiis.2017.06.011
- Song, Z., & Xia, J. (2016). Spatial and Temporal Sentiment Analysis of Twitter data. In *European Handbook of Crowdsourced Geographic Information* (Vol. 25, pp. 205–221). Ubiquity Press. doi:10.5334/bax.p
- Stieglitz, S., Mirbabaie, M., Ross, B., & Neuberger, C. (2018). Social media analytics – Challenges in topic discovery, data collection, and data preparation. *International Journal of Information Management*, 39(December), 156–168. <ALIGNMENT.qj></ALIGNMENT>10.1016/j.ijinfomgt.2017.12.002

- Sun, X., Zhang, C., Ding, S., & Quan, C. (2018). Detecting anomalous emotion through big data from social networks based on a deep learning method. *Multimedia Tools and Applications*, (420), 1–22. doi:10.1007/s11042-018-5665-6
- Sundermann, C. V., Domingues, M. A., Sinoara, R. A., Marcacini, R. M., & Rezende, S. O. (2019). Using opinion mining in context-aware recommender systems: A systematic review. *Information (Switzerland)*, 10(2), 1–45. doi:10.3390/info10020042
- Tang, X. (2013). Exploring on-line societal risk perception for harmonious society measurement. *Journal of Systems Science and Systems Engineering*, 22(4), 469–486. doi:10.1007/s11518-013-5238-1
- Tien, J. M. (2013). Big Data: Unleashing information. *Journal of Systems Science and Systems Engineering*, 22(2), 127–151. doi:10.1007/s11518-013-5219-4
- Usai, A., Pironti, M., Mital, M., & Aouina Mejri, C. (2018). Knowledge discovery out of text data: A systematic review via text mining. *Journal of Knowledge Management*, 22(7), 1471–1488. doi:10.1108/JKM-11-2017-0517
- Wang, S., & Noe, R. A. (2010). Knowledge sharing: A review and directions for future research. *Human Resource Management Review*, 20(2), 115–131. doi:10.1016/j.hrmr.2009.10.001
- Wazarkar, S., & Keshavamurthy, B. N. (2019). A soft clustering technique with layered feature extraction for social image mining. *Multimedia Tools and Applications*, 78(14), 20333–20360. Advance online publication. doi:10.1007/s11042-018-6881-9
- Xu, Z., Liu, Y., Yen, N., Mei, L., Luo, X., Wei, X., & Hu, C. (2016). Crowdsourcing based Description of Urban Emergency Events using Social Media Big Data. *IEEE Transactions on Cloud Computing*, 7161(c), 1–1. doi:10.1109/TCC.2016.2517638
- Xu, Z., Xuan, J., Liu, Y., Choo, K. K. R., Mei, L., & Hu, C. (2017). Building spatial temporal relation graph of concepts pair using web repository. *Information Systems Frontiers*, 19(5), 1029–1038. doi:10.1007/s10796-016-9676-4
- Yang, P., & Chen, Y. (2017). A survey on sentiment analysis by using machine learning methods. In *2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)* (pp. 117–121). IEEE. doi:10.1109/ITNEC.2017.8284920
- Yang, W., Liu, X., Liu, J., & Cui, X. (2018). Prediction of collective actions using deep neural network and species competition model on social media. *World Wide Web (Bussum)*. Advance online publication. doi:10.1007/s11280-018-0655-1
- Yue, L., Chen, W., Li, X., Zuo, W., & Yin, M. (2018). A survey of sentiment analysis in social media. *Knowledge and Information Systems*, 1–47. doi:10.1007/s10115-018-1236-4
- Zhang, W., & Zhu, Y. Chun, & Wang, J. Peng. (2018). An intelligent textual corpus big data computing approach for lexicons construction and sentiment classification of public emergency events. *Multimedia Tools and Applications*. Advance online publication. doi:10.1007/s11042-018-7018-x
- Zheng, N., Liu, Z., Ren, P., Ma, Y., Chen, S., Yu, S., Xue, J., Chen, B., & Wang, F. (2017). Hybrid-augmented intelligence: Collaboration and cognition. *Frontiers of Information Technology & Electronic Engineering*, 18(2), 153–179. doi:10.1631/FITEE.1700053
- Zuo, Y., Wu, J., Zhang, H., Wang, D., & Xu, K. (2018). Complementary Aspect-Based Opinion Mining. *IEEE Transactions on Knowledge and Data Engineering*, 30(2), 249–262. doi:10.1109/TKDE.2017.2764084

Victor Diogho Heuer de Carvalho is Assistant Professor in the Backland Campus at Universidade Federal de Alagoas (Delmiro Gouveia), working in the undergraduate courses of Management and Civil Engineering. He is also Doctoral Candidate in the Graduate Program in Management Engineering at Universidade Federal de Pernambuco (Recife), and member of the research group in Information and Decision Systems, currently working with research on applications of knowledge discovery tools in Public Security area.

Ana Paula Cabral Seixas Costa is Associate Professor of Management Engineering at the Federal University of Pernambuco, member of the CDSID (Center for Decision Systems and Development). Her main research interests are in Information Systems, Decision Support Systems and Multiple Criteria Decision Making (MCDM). Her works have been published in journals such as IJIM, EJOR, IJPE, IJPM and Expert Systems. Actually She is Head of the Management Engineering Department at UFPE. She is a member of the EURO Working Group on Decision Support Systems. She has actively participated in international events in the area of Information Systems and Decision Support, such as EWGDSS International Conference, GDN and AMCIS, as well as streaming the area in events such as IFORS and EURO.