Quality Management in Social Business Intelligence Projects

María José Aramburu[1], Rafael Berlanga[2] and Indira Lanza-Cruz[2]
[1]Department de'Enginyeria i Ciència dels Computadors, Universitat Jaume I, E-12071 Castelló de la Plana, Spain
[2]Department de Llenguatges i Sistemes Informàtics, Universitat Jaume I, E-12071 Castelló de la Plana, Spain

Keywords: Quality Management, Social Data, Business Intelligence.

Abstract: Social networks have become a new source of useful information for companies. Increasing the value of social data requires, first, assessing and improving the quality of the relevant data and, subsequently, developing practical solutions that apply them in business intelligence tasks. This paper focuses on the Twitter social network and the processing of social data for business intelligence projects. With this purpose, the paper starts by defining the special requirements of the analysis cubes of a Social Business Intelligence (SoBI) project and by reviewing previous work to demonstrate the lack of valid approaches to this problem. Afterwards, we present a new data processing method for SoBI projects whose main contribution is a phase of data exploration and profiling that serves to build a quality data collection with respect to the analysis objectives of the project.

1 INTRODUCTION

Social networks have become a new source of useful information for companies, helping them, among others, to know the opinions of their customers, to analyse the trends of the market, and to discover new business opportunities (Ruhi, 2014). Although social media data are highly heterogeneous and difficult to manage, they can produce meaningful information for decision-makers. The research here presented focuses on how to build quality social media data collections for Social Business Intelligence (SoBI) projects.

SoBI is defined in (Gallinucci et al., 2015) as the discipline that aims at combining corporate with social media data to let decision-makers analyse and improve their business needs based in the trends and moods perceived from de environment. Until now, companies have used social networks mainly for marketing purposes. SoBI tools are often applied by marketing departments to monitor the results of their activities by means of a group of social media metrics (e.g. number of likes, followers, or replies) (Keegan & Rowley, 2017) and methods (Lee, 2018). However, social media metrics are rarely combined with other business measures to calculate key performance indicators of different purpose (Ruhi, 2014). The integration of social media metrics with corporate data to produce new strategic indicators are some of the SoBI related applications that can bring new opportunities to companies (Agostino et al., 2018).

In current SoBI projects, filtering Twitter by means of a set of keywords (i.e., topics and hashtags) generates a flow of potentially relevant tweets. By processing them, a group of measures and attributes is extracted which are used to calculate metrics and indicators. However, most times, in the large volume of data retrieved, there are many tweets apparently related to the subject of analysis but that, because of their origin, intention or specific contents, are not useful. Therefore, we consider that before processing the tweets, some filtering tasks are necessary. However, previous frameworks for social media analytics have not given the required importance to data quality in their data preparation phases (Stieglitz et al., 2018).

In this paper, we highlight the importance of building a quality data collection in the data preparation phase of SoBI projects. After defining in Section 2 the special requirements of the analysis cubes of a SoBI project and reviewing, in Section 3, previous work to demonstrate the lack of valid approaches to this problem, we present, in Section 4, a new data processing method for SoBI projects. Its main contribution is a new phase of exploration and profiling of the retrieved data that serves to build a quality data collection with respect to the analysis objectives of the project.

Aramburu, M., Berlanga, R. and Lanza-Cruz, I.
Quality Management in Social Business Intelligence Projects.
DOI: 10.5220/0010495703200327
In *Proceedings of the 23rd International Conference on Enterprise Information Systems (ICEIS 2021) - Volume 1*, pages 320-327
ISBN: 978-989-758-509-8
Copyright © 2021 by SCITEPRESS – Science and Technology Publications, Lda. All rights reserved

2 BUSINESS INTELLIGENCE WITH SOCIAL MEDIA DATA

Traditional Business Intelligence (BI) projects apply several types of tools to exploit the facts of a subject of analysis, like for example, the sales of a company. Usually, these facts are represented into the analysis cubes of a multidimensional data model. A subject of analysis is a business context dependent description that clearly determines the analysis objectives of the project, and consequently, the measures and dimensions of its analysis cubes.

In the case of a SoBI project, the facts that serve to feed the analysis cubes come from a collection of user-generated contents (e.g. tweets). These contents are external to the company and consists of unstructured data with a high level of heterogeneity, which makes much more difficult to obtain valid facts. Here, the main issue is that by simply translating the subject of analysis into a group of keywords, the retrieved tweets will include many posts generated with very different purposes and out of the scope of the SoBI project. Consequently, many of the retrieved posts do not add any value to the analysis tasks and may even be counterproductive, due to the misinformation and the noise that they can introduce in the intended analysis. For example, in our experiments, when retrieving tweets with opinions from the users of a Ford car model, it is impossible to avoid the retrieval of many memes about the actor Harrison Ford and the words "fiesta", "escort" or "focus", which also refers to different Ford car models.

2.1 Defining Analysis Cubes

The subject of analysis of a SoBI project must be contextualised in terms of the organization's strategic objectives (Ruhi, 2014) (Lee, 2018). In our approach, this task consists in defining the analysis objectives of the project by means of an analysis cube, that is, of a group of measures which will be analysed from different points of view and levels of detail. Among these measures, there can be metrics extracted by processing tweet contents together with other coming from tweets metadata (e.g. number of followers, replies, likes, …). The relevant data will depend on the analysis objectives to be achieved. These are some examples:

- If the objective is brand awareness, metrics such as share of voice, are an indicator of the visibility of the brand with respect to its competitors.
- To gain engagement from customers or prospects, audience metrics such as number of comments, shares and trackbacks serve to define good indicators.
- To improve customer service, monitor the comments and negative sentiment of customers to generate the right responses and solutions, as well as indicators such as resolution time and satisfaction scores after giving support.

Additionally, the dimensions of the analysis cubes should represent all the attributes, categories, and hierarchies that the analysis objectives of the project require. For example, when the managers of a car brand want to analyse the sentiments and opinions about their models, they can apply several measures and points of view. By processing tweet contents, it can be possible to estimate measures such as the general polarity of a tweet from opinions about both the brand or the car model, as well as, the specific aspects assessed by the user (i.e. engine, ecology, design, …). The temporal dimension can be applied to follow up the evolution of these measures, and the spatial dimension would allow studies by considering the geographical distribution of the sources of opinion. However, to discover new valuable insights, it would be interesting to have additional dimensions for the categories of car models of the brand as well as for the different car aspects. Furthermore, to have the measures mapped against a dimension of types of users would allow the execution of analysis operations adapted to the different actors of this application (e.g. journalists, professional drivers, general drivers, …). Modelling all these elements requires understanding the brand business model and its strategic objectives.

When building an analysis cube, a main issue is how to obtain a social media data collection from where to extract the measures of the analysis cubes along with all the dimensional attributes. To produce reliable insights, the analysis cubes must be complete and uniformly filled, with no gaps nor biased dimensions. For this reason, it will be necessary to retrieve enough tweets by means of a full range of possible keywords and hashtags but, at the same time, to avoid overloading the collection with redundant and non-relevant tweets. Therefore, it is necessary to execute filtering operations on the collection. Determining the feasibility of the project and the range of retrieval and cleansing operations needed to build a quality social media data collection is a complex task that current methodologies do not give enough importance to.

3 RELATED WORK

Defining a methodology to build quality collections of social media data involves several things. By one hand, it is important including in the methodology some data cleaning operations. On the other hand, it is necessary to determine the best quality metrics to be applied by cleaning operations as well as to assess the quality of the overall collection. In this section, we review previous methodologies from the point of view of quality management, and then, we summarise the different approaches to measuring credibility, the most important quality attribute for social media.

3.1 Previous Methodologies

Modern methodologies for processing Big Data consider that quality management should take part of all the phases of the pipeline (Taleb et al., 2018) (Pääkkönen & Jokitulppo, 2017). Considering the data acquisition phase, the Big Data quality framework presented by (Taleb et al., 2015) highlights the need for cleansing, integration, filtering and normalization operations to improve the quality of the collection and, in this way, to save on costs and perform accurate data analysis.

The uniform data management approach of (Goonetilleke et al., 2014) reviews three main groups of research challenges to address when building a Twitter data analytics platform. For data collection, the main issue is the specification of the best set of retrieval keywords and hashtags; for data pre-processing, they demand for specific data processing and extraction strategies for Twitter data; and finally, for data management, they explain that quality management is a major issue that requires declarative languages to query social networks. However, as noted by (Stieglitz et al., 2018), in the papers that already document the data tracking and preparation steps of their social media projects, these steps are often dealt with superficiality, never with as much extension as data analysis tasks. The authors conclude that data discovery, collection and preparation phases for social media analytics require more research.

In the quality management architecture for social media data of (Pääkkönen, P. & Jokitulppo, J., 2017), the data acquisition, data processing & analysis, and decision-making phases can include functionalities for quality control and monitoring. In this approach, data quality management consists in assigning values to quality attributes which can be applied at extraction, processing, and analysis time from the point of view of the data source, the data and the user respectively. The quality, organizational and decision-making policies of the organization define the criteria to filter the quality data. Although, the proposed architecture can represent all these data quality elements, the authors do not propose a methodology for defining and applying them.

The methodology for SoBI of (Francia et al., 2016) recognizes that crawling design can be one of the most complex and time-consuming activities and aims at retrieving in-topic clips by filtering off-topic clips. They also explain that filtering off-topic clips at crawling time could be difficult due to the limitations of the crawling languages and propose to filter them at a later stage by using the search features of a documents database. The authors note that manually labelling a sample of the retrieved clips enables the team to trigger a new iteration where the crawling queries are redefined to cut off-topic clips out more effectively. However, the proposed methodology does not include any cleaning operations in its functional design nor determines whether to process them before or after retrieving the social data. This work does not consider the quality of data as a main objective, and it does not explain how to obtain a good quality collection.

A second methodology for SoBI, proposed by (Abu Salih, 2015), consists of five stages that process social media data and integrate them in the data warehouse files. They propose to execute cleaning operations to remove dirty data at the data acquisition stage prior to data storage. Afterwards, during data analysis, the collected data is processed to infer a value of trust for the relevant data. In the last stage of this methodology, a data structuring process serves to integrate traditional and social media data in order to produce new insights. In this way, the exploited social media data has a minimum level of trust with respect to its domain, although, the analysis objectives of the SoBI project are not considered as additional criteria to validate the source data.

3.2 Quality Metrics for Social Media

Measuring social media data quality can be performed using different metrics and techniques. The literature review clearly reveals that credibility is the most important quality attribute for social media, and many different approaches have been proposed to measure it. It is important to clarify that for these authors, credibility is a broad concept that intersects with other semantically related quality attributes such as trust, reliability, believability, veracity, relevance, validity and, in some cases, even understandability and reputation.

For measuring the credibility of social media data, statistical or machine learning techniques are usually adopted. Among the many metrics used to feed these algorithms, there are those obtained by processing tweets contents, mainly looking for textual properties, writing styles, linguistic expressions, sentiments, and additional elements like URL´s or pictures. Social parameters extracted from tweets metadata about the post and its poster are a second source of metrics. Finally, there are a last group of metrics with information about the behaviour and actions of the users in the social network. Table 1 shows metrics applied in the literature to measure credibility (Sikdar et al., 2013) (Gupta et al., 2014) (Viviani & Pasi, 2016) (Păvăloaia et al., 2020) and which, in many cases, could also be applied to assess other quality attributes. The wide range of metrics applied shows that, in each approach, credibility can be understood in a different way, and that it is up to the user to choose the best metrics taking into account the context of each project, its domain and the technologies applied.

In social media, a quality attribute of utmost importance is reputation which can be also defined as the authority of the poster. Previous work frequently considers simple measures such as the number of followers to calculate indicators of good reputation, i.e. when a user is in many Twitter lists and has many followers is because the contents that generates satisfy many users. This approach ignores that user's interest can be diverse and evolve and change over time. However, recently, more realistic approaches have proposed to consider this quality attribute as a time and domain-dependent parameter (Abu-Salih et al., 2020).

Finally, in (Pasi et al., 2019), the adequacy and the potentialities to describe the issue of the assessment of the credibility of user-generated content in social media as a multi-criteria decision-making problem have been discussed. Their approach to determining the credibility of online reviews considers features connected to the contents, the information sources and the relationships established in social media platforms. These features are evaluated by the users in terms of their impact on veracity. By considering different aggregation schema for the partial performance scores and their impact, the authors calculate an overall score of veracity. With respect to data-driven approaches based on Machine Learning techniques, their approach makes the user more aware of the choices that led to the proposed decision and can make the considered problem less data-dependent.

In this section, we have reviewed relevant methodologies and techniques to build collections of social media data for SoBI from the point of view of quality management. The conclusion is that while most approaches to social media analysis for decision support apply different quality criteria during data collection and preparation, it is not clear at this point how to define a general-purpose quality conceptual model for social media data. Previous work has provided us with many different metrics with multiple purposes that depend primarily on the application. The experience demonstrates that, whatever technology applied to decide the quality of data, a good combination of different types of metrics use to be part of the solution. However, there is no systematic methodology for defining and applying these metrics in order to build a reliable collection of social data. In this case, the main question is how to find the best metrics that can be applied to both prepare the collection and measure its quality with respect to the analysis objectives of a SoBI project, i.e. the construction of an analysis cube with a set of social measures that can be analysed from different points of view and levels of detail.

Table 1: Sample of metrics to measure credibility in social media data found in the literature.

Tweet contents	Post and poster metadata	User behaviour
# Chars/words	Account age	# Retweets
# Punctuation symbols	Listed count	# Tweets
# Pronouns	Status count	# Tweets favorited
# Swear words	Favourites count	# Mentions
# Uppercases	# Friends	# Tweets are a reply/retweet
# Emoticons	# Followers	Mean time between tweets
#URLs/images	# Followings	# Likes received
# Hashtags	Ratio of followers to friends	# Directed tweets
# Misspelled words	Mean text length in tweets	# Users that propagate the user
# Sentences	Mean hashtags in tweets	# Users the user propagates from
Average length of sentences	Mean # URLs/ mentions in tweets	# Tweets propagated by other users
# Product mentions	Verified user	# Users that converse with the user
# Product features mentioned	User image in user profile	Mean number of conversations
# Opinion sentences	Tweet geographical coordinates	Average length of chain-like behaviour

4 QUALITY MANAGEMENT IN A SoBI PROJECT

Evaluating the quality of social media data for a SoBI project requires the definition of the best quality indicators for the source data (Immonen et al., 2015). Tweets present many different aspects that would serve to filter them; being Tweet contents and users' attributes and actions, the main sources of quality metrics. However, the selection of the best quality metrics for a SoBI project is a complex task that requires a deep understanding of the business context and objectives of analysis, as well as the social media data to be managed (Berlanga et al., 2019).

In data intensive applications, quality conditions should serve as criteria to program the cleaning operations and to measure the quality of the overall collection. As explained in (Sadiq & Indulska, 2017), traditional methods for managing data quality follow a top-down approach: the analysis of user requirements produces some quality rules that serve to govern data, to assess data quality, and to execute cleaning operations. This approach is suitable for managing the quality of data generated internally by an organization. However, when the organization does not control the external processes that generate the available data, as in the case of social media data, quality assessment requires prior knowledge about the data. In these cases, the data quality management follows a bottom-up approach that starts with the execution of some exploratory analysis and data profiling tasks. These tasks help to find data quality rules and requirements that will drive the data collection process. In order to execute the preliminary exploration of the available data, interactive, statistical and data mining techniques can be applied over collections of data. Some of the data mining techniques that can help in these tasks are clustering, classification, data modelling and data summarization (Stieglitz et al., 2018).

Following this approach, in this section we present a new method for Social Media data processing for SoBI applications whose main contribution is a first phase of exploration and profiling of the retrieved data that serves to build a quality data collection with respect to the analysis objectives of the project.

4.1 A Data Processing Methodology

The analysis objectives of a SoBI project are to analyse from various perspectives (i.e. dimensions and categories of analysis) the social measures drawn from a complete collection of relevant tweets. Often these will be simple measures such as the number of likes or followers, especially if the aim is to analyse the success of a marketing campaign. For example, if the objective is brand awareness, social measures such as voice share are an indicator of the brand's visibility in relation to its competitors that can be calculated by counting hashtags and mentions. However, there are other more difficult measures to analyse, such as when analysing feelings about different aspects of a product or service (e.g. a holiday package or a car).

To deal with the completeness of the collection and the relevance of the tweets in it, in this paper we propose a new data processing methodology that consists of three main phases: Collection Construction, Data Preparation and Data Exploitation (see Figure 1). The first phase is the construction of a collection of tweets through an exploratory process executed by the user and directed by the quality of the recovered data. When a quality data collection is ready, in the data preparation phase, the facts of the analytical cubes are extracted from the posts and then exploited in the last phase of the process.

As Figure 1 illustrates, during the Collection Construction phase, the user executes some data exploratory and profiling tasks to assess and improve data coverage and data quality until obtaining a quality collection that meets the project's analysis objectives. More specifically, this phase consists of two complementary and iterative tasks:

a) Evaluating the subject coverage of the collection with respect to its topics and users. Analysing the vocabulary of the collection will help to know when to redefine the keywords applied to program the Twitter API to obtain either a more complete or precise collection. Furthermore, profiling the range of users that post the retrieved tweets along with their metadata is important in determining the measures and dimension attributes available to be part of the analysis cubes, as for example, the users' demographic data present in their descriptions.
b) Analysing and improving the quality of the collection by filtering the posts of low quality or out of the scope. Finding the best quality metrics that help to clean a collection requires exploring and profiling it in order to discover its main characteristics and the sources of noise. Then, cleaning operation can apply different types of quality metrics extracted by processing the tweets contents and metadata as well as the user descriptions. For example,

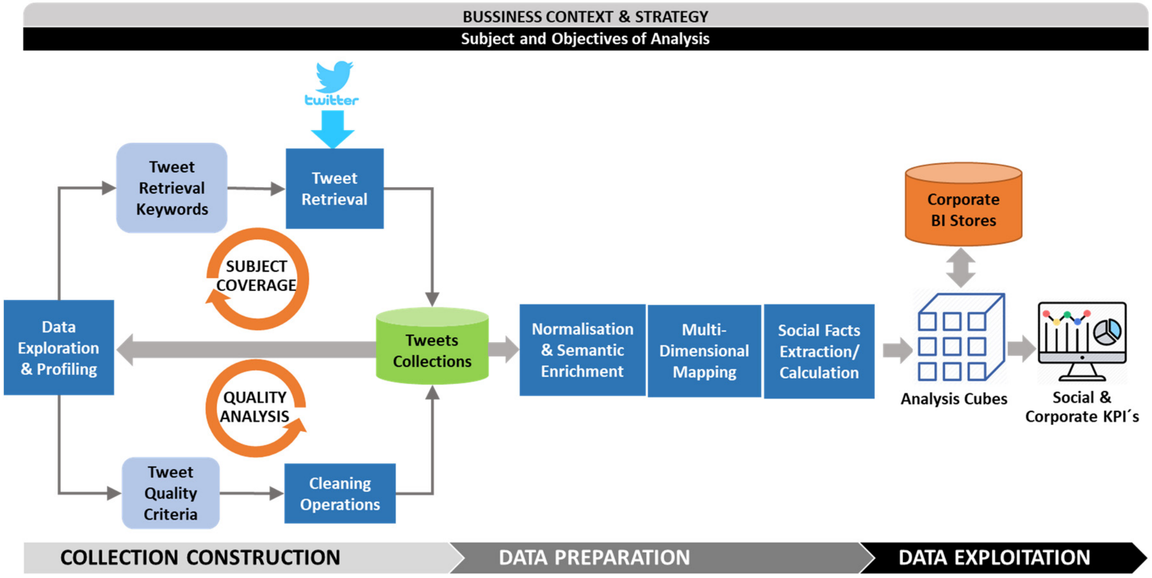

Figure 1: The proposed social media data processing method for Social Business Intelligence.

from a collection of user opinions about cars, memes should be removed because their low added value as well as the ad posts sent by marketing departments. A global analysis enhances the detection of long-term quality problems, such as redundancy, bias and noise, which are often difficult to detect from local analysis (i.e., directly over the streamed data). Analysing a long-term data stream also allows building robust language models for a given domain, smoothing the effects of punctual viral events (Lanza-Cruz et al., 2018).

In the Collections Construction and Data Preparation phases, the processing of tweets to extract the measures and values that serve both to clean the collection and to feed the analysis cubes, can be made in different ways. Some values are directly available in the tweets metadata, such as post-date and number of followers of the user. Other values can be calculated with a simple processing like counting tweets over a period. Evaluating the grammatical richness of a post is executed by a process that calculates some textual measures (Gupta et al., 2014). However, more sophisticated Natural Language Processing techniques are required when analysing the vocabulary of the whole collection (Berlanga et al., 2019) or extracting the facets and polarity of opinion expressions (García-Moya et al., 2011). Other intelligent techniques like Entity Resolution and Ontology Mappings, mainly applied for the Semantic Enrichment and Dimensional Mapping tasks of Figure 1, have shown useful to extract dimensional attributes from social media data (Berlanga et al., 2015) (Pereira et al., 2018). Furthermore, the tweets in the collection can be applied to train Machine Learning algorithms that discover some useful properties of the tweets, as for example their credibility and intention, and to classify them into analysis categories. The main objective of all these tasks is to produce a complete analysis cube whose measures and dimension attributes can enable reliable multidimensional studies.

Finally, in the Data Exploitation phase, the analysis cubes constructed by processing the tweets collection can be stored into the Corporate Data Warehouse for future uses. OLAP applications, or any other Business Intelligence or Data Mining tools, can be applied to analyse and extract new insights from these cubes. If necessary, the stored social media facts can be combined with corporate data to design new key performance indicators tied to the strategic business objectives such as ROI and profit margins. In this way, users can associate social media actions with sales volume, revenue increases or decreases and other relevant metrics and reveal new insights.

The focus on integrating social media metrics and internal business measures is important in our approach. However, combining qualitative data from social networks with the quantitative data hosted in traditional BI systems may seem a difficult task. Note that as pointed out by (Ruhi, 2014) this will require the semantic integration of the data elements in both kinds of external and internal data sources. From a practical point of view, this integration can be possible if the dimensional parameters applied to construct the analysis cubes of social media and corporate data are compatible (i.e. mappable).

At this point, it is important to note that the success of a SoBI project will mainly depend on the completeness of the analysis cube that can be constructed. This means that sometimes, it will be difficult or even impossible to build the collection of tweets to fill the analysis cube in all its dimensions. To this end, the data exploration and profiling tasks of the Collection Construction phase of our method will serve to assess and improve the coverage of the data until a quality collection is obtained that will allow the construction of the analysis cube that meets all the analysis objectives of the project. For example, if the objective of the analysis is to study the evolution of opinions about our products of different types of market participants (e.g. customers, sellers, users journalists, ...), it will be necessary to retrieve a representative sample of opinion tweets about each product in our catalogue, for all the dates of the analysis period, and for each type of user profile.

In addition, when the analysis cube must be complete, the availability of metadata becomes an important issue. For example, some analysis tasks, such as segmentation of market opinions by gender, age, location, or profession, require extraction of metadata from tweets and user descriptions. However, Twitter users do not always provide these parameters and some tweets will not be valid because they lack some key dimensional values. This issue may mean that the SoBI project is not feasible and that its analysis objectives should be modified by redefining or eliminating some attributes or dimensions of the analysis cube. For example, in most cases, the analysis of the opinions of consumers must be executed at country level due to the lack of precise metadata with the geo-location of the users.

5 CONCLUSIONS

Up to our knowledge, this is the first general approach to the construction of social media data collections oriented towards corporate BI tasks. With respect to previous work, the main contribution of this methodology is that it considers collection construction as an iterative exploration process in which the user analyses the current collection from the point of view of the analysis objectives and discovers clues about how to improve it. This data exploration & profiling task has not received enough attention in previous methodologies and projects and, however, in our experiments we have validated that it helps users to identify the retrieval conditions and cleaning operations that the construction of each specific collection requires, as well as, to assess the feasibility of the analysis objectives of the SoBI project with respect to the availability of reliable social media data (Lanza-Cruz et al., 2018) (Berlanga et al., 2019) (Aramburu et al., 2020).

For future work, an open issue is how to complement the here presented processing method with intelligent tools that guide the user in the selection of the best tweet quality metrics and criteria for each specific data domain and application. How to combine them into global quality indicators for a collection with respect to the analysis objectives of a SoBI project would be the next step (Berlanga et al., 2019). New SoBI application scenarios are also interesting for future research (Aramburu et al., 2020).

ACKNOWLEDGEMENTS

This research has been funded by the Spanish Ministry of Industry and Commerce grant number TIN2017-88805-R and by the pre-doctoral grant of the Universitat Jaume I with reference PREDOC/2017/28.

REFERENCES

Abu-Salih, B., Wongthongtham, P., Beheshti,S., Beheshti, B. (2015). Towards A Methodology for Social Business Intelligence in the era of Big Social Data incorporating Trust and Semantic Analysis". In Second International Conference on Advanced Data and Information Engineering (DaEng-2015), ed. Bali, Indonesia: Springer, 2015.

Abu-Salih, B., Chan, K.Y., Al-Kadi, O. et al. Time-aware domain-based social influence prediction. J Big Data 7, 10 (2020). https://doi.org/10.1186/s40537-020-0283-3.

Agostino, D., Arnaboldi, M., & Azzone, G. (2018). Social Media Data into Performance Measurement Systems: Methodologies, Opportunities, and Risks. Azevedo, E. et al. (Eds.), Handbook of Research on Modernization and Accountability in Public Sector Management. IGI Global. 254-275.

Aramburu, M.J.; Berlanga, R.; Lanza, I. (2020) “Social Media Multidimensional Analysis for Intelligent Health Surveillance”. Int. J. Environ. Res. Public Health, 17, 2289, https://doi.org/10.3390/ijerph17072289.

Berlanga, R., García-Moya, L., Nebot, V., Aramburu, M. J., Sanz, I., & Llidó, D. M. (2015). SLOD-BI: An Open Data Infrastructure for Enabling Social Business Intelligence. International Journal of Data Warehousing and Mining (IJDWM), 11(4), 1-28. doi:10.4018/ijdwm.2015100101

Berlanga, R., Lanza-Cruz I. & Aramburu, M. J. (2019) "Quality Indicators for Social Business Intelligence," 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS), Granada, Spain, 2019, pp. 229-236, doi: 10.1109/SNAMS.2019.8931862.

Francia M., Gallinucci E., Golfarelli M., Rizzi S. (2016) Social Business Intelligence in Action. In: Nurcan S., Soffer P., Bajec M., Eder J. (eds) Advanced Information Systems Engineering. CAiSE 2016. Lecture Notes in Computer Science, vol 9694. Springer, Cham.

Gallinucci, Golfarelli & Rizzi, (2015). Advanced topic modeling for social business intelligence. Information Systems., 53, pp. 87-106.

García-Moya, L., Anaya-Sánchez, H., Berlanga, R., Aramburu, M.J. (2011) Probabilistic Ranking of Product Features from Customer Reviews. In: Vitrià J., Sanches J.M., Hernández M. (eds) Pattern Recognition and Image Analysis. IbPRIA 2011. Lecture Notes in Computer Science, vol 6669. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-21257-4_26.

Goonetilleke, O., Sellis, T., Zhang, X., & Sathe, S. (2014). Twitter analytics: a big data management perspective. ACM SIGKDD Explorations Newsletter, 16(1), 11-20.

Gupta, A., Kumaraguru, P., Castillo, C. & Meier, P. (2014). TweetCred: Real-Time Credibility Assessment of Content on Twitter. Proceedings of the 6th International Conference on Social Informatics. 228-243. 10.1007/978-3-319-13734-6_16.

Immonen, A., Pääkkönen, P. & Ovaska, E. (2015). Evaluating the Quality of Social Media Data in Big Data Architecture. IEEE Access. 3. 1-1. 10.1109/ACCESS.2015.2490723.

Keegan, B. & Rowley, J. (2017). Evaluation and decision-making in social media marketing. Management Decision. 55. 15-31. 10.1108/MD-10-2015-0450.

Lanza-Cruz, I.; Berlanga, R.; Aramburu, M.J. (2018). Modeling Analytical Streams for Social Business Intelligence. Informatics. 5. 33. https://doi.org/10.3390/ informatics5030033

Lee, I. (2018). Social media analytics for enterprises: Typology, methods, and processes. Business Horizons. 61.2. 199-210, https://doi.org/10.1016/j.bushor.2017.11.002.

Pääkkönen, P. & Jokitulppo, J. (2017). Quality management architecture for social media data. J Big Data 4, 6 (2017). https://doi.org/10.1186/s40537-017-0066-7.

Pasi, G., Viviani,M. Carton,A. (2019). A Multi-Criteria Decision Making approach based on the Choquet integral for assessing the credibility of User-Generated Content. Information Sciences, vol 503. Pages 574-588. https://doi.org/10.1016/j.ins.2019.07.037.

Păvăloaia, V.-D., Anastasiei, I.-D., & Fotache, D. (2020). Social Media and E-mail Marketing Campaigns: Symmetry versus Convergence. Symmetry, 12(12), 1940. doi:10.3390/sym12121940.

Pereira,V., Fileto, R., Santos,W., Wittwer,M., Reinhold, O., Alt, Rainer (2018). A Semantic BI Process for Detecting and Analyzing Mentions of Interest for a Domain in Tweets. WebMedia '18: Proceedings of the 24th Brazilian Symposium on Multimedia and the Web. 197–204, https://doi.org/10.1145/3243082.3243100.

Ruhi, U. (2014). Social Media Analytics as a BI Practice: Current Landscape & Future Prospects. Journal of Internet Social Networking & Virtual Communities. 1-12. 10.5171/2014.920553.

Sadiq, S. & Indulska, M. (2017). Open data: Quality over quantity. International Journal of Information Management. 37. 150-154. 10.1016/j.ijinfomgt.2017.01.003.

Sikdar, S., Kang, B., ODonovan, J., Höllerer, T. and Adah, S. (2013). Understanding Information Credibility on Twitter, International Conference on Social Computing, Alexandria, VA, pp. 19-24, doi: 10.1109/SocialCom.2013.9.

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. 156-168.

Taleb,I., Serhani, M. A. & Dssouli, R. (2018), Big Data Quality: A Survey, IEEE International Congress on Big Data (BigData Congress), San Francisco, CA, pp. 166-173, doi: 10.1109/BigDataCongress.2018.00029.

Taleb,I. Dssouli, R. & Serhani, M.A. (2015), Big data pre-processing: A quality framework. 2015 IEEE international congress on big data. 191-198.

Viviani, Marco & Pasi, Gabriella. (2016). Quantifier Guided Aggregation for the Veracity Assessment of Online Reviews. International Journal of Intelligent Systems. 32. 10.1002/int.21844.