COMPARISON OF TOPIC MODELLING TECHNIQUES IN MARKETING – RESULTS FROM AN ANALYSIS OF DISTINCTIVE USE CASES

6-14-2021

COMPARISON OF TOPIC MODELLING TECHNIQUES IN MARKETING - RESULTS FROM AN ANALYSIS OF DISTINCTIVE USE CASES

Janik Wörner
University of Regensburg, janik.woerner@ur.de

Daniel Konadl
University of Regensburg, daniel.konadl@wiwi.uni-regensburg.de

Isabel Maria Schmid
University of Regensburg, isabel.schmid@wiwi.uni-regensburg.de

Susanne Leist
University of Regensburg, susanne.leist@ur.de

COMPARISON OF TOPIC MODELLING TECHNIQUES IN MARKETING - RESULTS FROM AN ANALYSIS OF DISTINCTIVE USE CASES

Research paper

Janik Wörner, University of Regensburg, Regensburg, Germany, Janik.Woerner@ur.de
Daniel Konadl, University of Regensburg, Regensburg, Germany, Daniel.Konadl@ur.de
Isabel Schmid, University of Regensburg, Regensburg, Germany, Isabel.Schmid@ur.de
Susanne Leist, University of Regensburg, Regensburg, Germany, Susanne.Leist@ur.de

Abstract

Currently, topic modelling is an effective analytical tool for the automated investigation of text data. However, identifying the underlying topics is still a challenging task that is dependent on the selection of the proper technique. Moreover, due to the considerable number of topic modelling techniques reported in the literature, uncertainty about the application of the techniques arises for both researchers and practitioners. Therefore, we conducted a comparison of three different topic modelling techniques (LDA, PAM, DMR) to give recommendations for three use cases identified in the literature: content extraction, trend analysis and content structuring. For each of them, we identified several requirements and by conducting the method 'Goal Question Metric', we derived several comparison metrics. We applied these metrics to a real-world Facebook data set (4,155,992 posts) to highlight the differences between the three topic modelling techniques and to give recommendations for our defined use cases.

Keywords: topic modelling, social media analysis, text analysis, marketing use cases

1 Introduction

Topic modelling is a prevalent kind of probabilistic generative model for extracting latent variables from large unstructured data sets (Liu et al., 2016). This can be applied to analyse different data such as bioinformatics data (Coelho et al., 2010), environmental data (Girdhar et al., 2013), and text data (Vayansky and Kumar, 2020). Thus, topic modelling has been studied in different disciplines and is also prevailing in information systems (IS) research, mainly focusing on Latent Dirichlet Allocation (LDA) (Blei et al., 2003) because of its simple applicability and good analysis results (Debortoli et al., 2016, Eickhoff and Neuss, 2017). With the continuous growth of social media and the consequential transformation of the way individuals interact with each other, increasing amounts of written data are created that can be analysed i.a. to support marketing related decision-making (Ghosh and Guha, 2013). Thus, social media data are increasingly used to enrich marketing tasks, such as complaint management (Einwiller and Steilen, 2015, Grégoire et al., 2015), innovation management (Mount and Martinez, 2014, Piller et al., 2012), or sales (Guesalaga, 2016, Marshall et al., 2012). However, the huge amount of written social media contents (Statista, 2020) complicates manual content analysis.
To solve this problem, automated topic discovery techniques and – in particular – topic modelling have been widely investigated (Chinnov et al., 2015, Eickhoff and Neuss, 2017, Hong and Davison, 2010). Topic modelling enables the analysis of a large amount of written social media data to extract embedded topics. Therefore, topic modelling has facilitated addressing marketing related questions and problems that have exceeded the feasibility of in-depth qualitative analysis (Eickhoff and Neuss, 2017). Thus, marketing related problems that refer to (1) content extraction, (2) trend analysis and (3) content structuring have often been discussed in the literature. Companies are required to base their products and services on customer requirements. Therefore, (1) content extraction with topic modelling is an appropriate application to extract customers' praise and criticism for product planning purposes (Irawan et

al., 2020, Rathore et al., 2018). To be aware of evolving trends concerning their own products and services, marketing departments conduct (2) trend analysis. Tracking evolving and changing requirements of customers is imperative to fulfil customers' wishes (Hong et al., 2012, Lozano et al., 2017). Moreover, (3) content structuring can help marketing departments to gain deeper insights into topics and their interrelatedness (Anoop et al., 2015, Srijith et al., 2017). Extracted hierarchical structures can reveal relationships between topics (e.g. price and product quality) and support more coordinated and sounder decision-making. Nevertheless, identifying the underlying topics of these documents is still a challenging task as the reasonable extraction of significant statistics and features from a dataset is dependent on the selection of the proper technique (Vayansky and Kumar, 2020).
As mentioned above, a growing number of IS-related investigations are currently using LDA. However, the basic LDA cannot represent all use cases (e.g. mapping hierarchies) for marketing related tasks so that extensions of LDA are essential. Therefore, not only do the numerous existing techniques for topic modelling hinder practical applications, but also the necessity of advanced technique-related knowledge. Liu et al. (2016) have divided various extensions of LDA into three areas: (I) extension of topic attributes (II) extension of document attributes and (III) extension of word attributes. Although numerous techniques are presented in these three areas, such as the Partially Labelled LDA (PLLDA) (Ramage et al., 2011), the Dirichlet Multinomial Regression (DMR) (Mimno and McCallum, 2008), or the Pachinko Allocation Model (PAM) (Li and McCallum, 2006), these extensions are scarcely applied. Due to the large number of topic modelling techniques in the current research literature, uncertainty about the selection of the right technique can arise. Vakansky and Kumar (2020) addressed this problem by conducting a theoretical comparison of different topic modelling techniques based on a structured literature review. Although this serves as a good overview of various topic modelling approaches and as a starting point for selecting a technique, differences only become obvious when applying them to a real-world data set. Furthermore, the results do not give clear suggestions which problem should be addressed with which technique. We addressed these problems by conducting a comparison between three different topic modelling techniques to give recommendations regarding the three use cases of (1) content extraction, (2) trend analysis, and (3) content structuring. We contribute to close this identified gap by comparing the practical application of the three techniques. This leads to the following research questions:

RQ1: Which criteria can be used to compare the different topic modelling techniques with each other?
RQ2: Which topic modelling technique can be recommended for the marketing related use cases (1) content extraction, (2) trend analysis, and (3) content structuring?

Topic modelling, both in general and especially regarding the analysis of companies´ social media posts, represents an important area for IS research. Accordingly, against the background of marketing we uncover various corporate use cases in the context of social media. By applying and comparing the three topic modelling techniques LDA, DMR, and PAM and by using build time, log-likelihood, coherence, word and topic intrusion as evaluation measures, we want to show differences between these techniques, identify advantages, disadvantages, and various potentials to enhance topic modelling techniques. Hence, we apply LDA, DMR, and PAM to a real-world data set. The remainder of this paper is as follows: section 2 provides a theoretical background. Then, we refer to the derivation of the three use cases and the respective requirements from literature. The transformation of them into topic modelling related metrics is also described here. Next, the procedure of the research 'Goal Question Metric' (cf. Basili, 1994) is described in section 3. Section 4 deals with the selection of the topic modelling techniques that are used for our comparison. The data analysis in section 5 achieves this comparison and further explains the data collection, the preparation of the data and the data analysis. Afterwards, in section 6, we present and discuss our results. Finally, section 7 draws an overall conclusion.

2 Theoretical Background

2.1 Social media

Social media can be defined as '*a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated*

Content (UGC)' (Kaplan and Haenlein, 2010, p.61). Social media connects people with the same interests, activities, backgrounds, or friendships (AlFalahi et al., 2014, Schneider et al., 2009). The active utilisation of social media enjoys particularly great popularity in private use. However, companies have also adopted social media to support value-creation (Hanna et al., 2011, McDonald and Aron, 2012). In particular, many companies apply these Internet-based applications such as content communities (e.g. YouTube), blogs, or social networks (e.g. Facebook or Twitter) to enable communication mainly with external stakeholders (Kietzmann et al., 2011). Thus, companies adopt social media to achieve different business objectives such as branding, including advertising, marketing, and content delivery for creating brand awareness (Culnan et al., 2010, Di Gangi et al., 2010, Kietzmann et al., 2011). Therefore, social media such as Facebook or Twitter serve as an important interface between companies and customers. This interface generates large amounts of data that need to be analysed and interpreted, as a company can strongly benefit from these data. In addition to structured social media data (e.g. timestamps, like counts, etc.), especially unstructured text data contain interesting contents for companies. Posts and comments often include an user´s major wishes, ideas, and expectations towards products, services, or a company in general (Hienerth et al., 2011, Sigala, 2012a, Sigala, 2012b). The so called 'Voice of the Customers' can be used to adjust marketing campaigns, to identify and support the position in the market and to adjust product features on customers´ favourability. However, to uncover this useful information from the large amount of data requires considerable effort (Dahal et al., 2019, Kumar and George, 2007, Womack and Jones, 1996). To avoid this problem, automated analysis of social media data such as social network analysis, sentiment analysis, and topic modelling can be conducted. Especially through the latter one, valuable information for companies can be extracted, as this technique is able to identify various (discussion) topics, perceptions, and opinions (Dahal et al., 2019, Lozano et al., 2017). However, automated analysis such as topic modelling are often complicated, as many companies are not familiar with the applied techniques, its implementation, and its purpose (Dai et al., 2011).

2.2 Topic modelling

Topic modelling aims to determine content structures in underlying document collections. Hereby, topic modelling refers to the use of generative probability models for determining latent relationships within a corpus of text data. The dataset under investigation is to be seen as a mixture of individual documents, where each document affects several corpus-wide topics, that in turn consist of frequently occurring words within the dataset (Blei, 2012). LDA can be considered as one of the most fundamental works in the topic discovery research area, wherefore a growing number of investigations currently uses this technique proposed by Blei et al. (2003). The authors describe their probabilistic model as *'a three-level hierarchical Bayesian model, in which each item of a collection is modeled as a finite mixture over an underlying set of topics'* (p. 993). This means that LDA is a generative model that is based on the assumption that documents are represented by a collection of different, latent topics. Each topic will be represented as a probability distribution over all words of a corpus (Krestel et al., 2009). LDA is frequently used in marketing to identify important issues for the adaptation of marketing campaigns or to identify product and service features currently being discussed (Chae and Park, 2018, Gao et al., 2012, Jeong et al., 2019, Ko et al., 2017, Luo et al., 2015, Xu and Xiong, 2020, Yu et al., 2019). However, the LDA as proposed by Blei et al. (2003) cannot represent all use cases (e.g. mapping hierarchies).

Thus, there are adaptations and extensions that are based on the probabilistic model of Blei et al. (2003). Generally, it can be differentiated between (I) topic-based extensions, (II) document-based extensions and (III) word-based extensions (cf. Liu et al., 2016). (I) Topic-based extensions derive structures and dependencies within the latent topics of a document (cf. Rathore et al., 2018, Rathore and Ilavarasan, 2017, Tuarob and Tucker, 2015). (II) Document-based extensions have the ability to incorporate an additional parameter into the model building (cf. Cheng et al., 2020, Lozano et al., 2017, Zhang et al., 2016). (III) The word-based extensions compute n-grams instead of Bag of Words (BoW) to incorporate the order of words in a document within the model generation procedure (cf. Wallach, 2006).

As can be concluded from previous research, the LDA approach is the predominantly applied technique (cf. Eickhoff and Neuss, 2017). Vakansky and Kumar (2020) provide a good overview of existing topic modelling techniques and accordingly develop a decision tree model supporting the selection of a technique. These authors also include adaptations and extensions of the basic LDA for recommending an appropriate technique. However, their recommendations are theoretical in nature and not derived from practical applications of the techniques. There are numerous topic modelling techniques in the literature, but the number of papers comparing them is scare. Further, it is not clear if the theoretical overview and the decision tree of Vakansky and Kumar (2020) can withstand empirical investigations. The authors only conditionally contribute to the operational application of topic modelling as recommendations should ideally be deduced from the analysis of real-world social media data.

2.3 Corporate use cases and requirements of marketing

Within the prevailing literature, we identified papers that deal with utilising topic modelling techniques in marketing. These applications of topic modelling described in the identified papers cover a wide range of marketing tasks. In particular, three main use cases could be identified: (1) content extraction is concerned about consolidating insights of topics discussed in written social media data. By investigating brand-related content from social media, (1) content extraction enables marketing representatives to develop an understanding of topics and themes, e.g. sustainability or product feature favourability, that are discussed by customers and parties of interest. Thereby, companies can improve their external presentation by, e.g. emphasising the ecological superiority of the own products in brand communications (Chae and Park, 2018), and guide future product planning initiatives by putting more focus on product features that are appreciated by customers (Cirqueira et al., 2017, Irawan et al., 2020, Ko et al., 2017, Rathore et al., 2018). Providing a current overview of customers' major wishes, ideas, and thoughts is therefore a central requirement for these techniques as the ever-increasing amount of social media data along with the breadth of the user base may hinder marketing departments to focus on the essential aspects (cf. Cirqueira et al., 2017, Ko et al., 2017, Lee et al., 2016, Liu et al., 2017, Rathore et al., 2018, Rathore and Ilavarasan, 2017). Hereby, (1) content extraction is of a retrospective nature and is less concerned with speed. Because of the abundance of information provided by social media and the necessity to cover as much useful information as possible, the main aim is to cover the relevant and most frequent topics embedded in social media texts (cf. Gao et al., 2012, Ibrahim and Wang, 2019, Irawan et al., 2020, Wang et al., 2016, Yang et al., 2016). Therefore, we assume training time of the topic modelling techniques as secondary, because there are the two central quality dimensions, relevance and dominance of topics, as the basis for well-funded decisions. Furthermore, techniques for (1) content extraction need to support comparisons with competitors, assessments of a company´s position in the market (Aiello et al., 2013) and effectively support product and service opportunities generation (Ko et al., 2017, Rathore et al., 2018, Rathore and Ilavarasan, 2017).

The second use case we could identify is (2) trend analysis. It has in common with (1) content extraction that it deals with extracting topics from large amounts of written social media data. However, (2) trend analysis focuses on keeping track of emerging trends and their development, while the results of (1) content extraction, are instead point-in-time snapshots of the contents that do not illustrate the dynamic courses of the topics. Marketing representatives applying (2) trend analysis related techniques strive to track how certain topics (e.g., product favourability or customer satisfaction) evolve geographically and temporally (Hong et al., 2012, Jeong et al., 2019, Lozano et al., 2017, Zhang et al., 2017). Thereby, marketing departments can enhance the effectiveness of brand message placement and the allocation of appropriate resources to marketing campaigns depending on geographical and temporal developments. Furthermore, linking topics and contents to groups of interested parties and customers enables companies to adapt brand messages to meet the respective target groups´ expectations and attitudes (Zhang et al., 2016). Therefore, topic modelling techniques for (2) trend analysis need to flexibly incorporate different parameters like authors (Zhang et al., 2016), locations (Cheng et al., 2020, Hong et al., 2012, Lozano et al., 2017, Wang et al., 2007), or time (Cheng et al., 2020, Wang et al., 2012, Zhang et al., 2017) into the model building procedure. Ideally, it should even be possible to include implicit parameters such as places or times mentioned in the texts (Lozano et al., 2017). Since trends describe current issues that influence customers' decision-making, companies want to respond to the resulting customer demands to adapt e.g. marketing campaigns (cf. Luo et al., 2015, Rathore et al., 2018, Zhang et al., 2015, Zhong and Schweidel, 2020). Compared to (1) content extraction, especially with fast sequenced social media data and topics of interests changing quickly, it is necessary to keep track of trends and topic transitions (cf. Wang et al., 2012, Zhang et al., 2017). The requirement to be able to react as quickly as possible is further intensified by reduced product lifecycles and globalised business environments that have made customer needs more dynamic (Jeong et al., 2019). For this reason, the applied topic modelling techniques must provide short training times (cf. Cheng et al., 2020, Wang et al., 2012, Zhang et al., 2017) and as well support quick comprehension of the extracted topics (cf. Jeong et al., 2019, Lozano et al., 2017) in order to keep up with the speed of the trends.

Compared to (3) content structuring, (2) trend analysis captures the dynamic courses of topics and do not collect hierarchical relationships and correlations. Topic modelling techniques for (3) content structuring enable deeper insights from textual social media by extracting not only the topics (cf. (1) content extraction) but also relationships and connections between them. In this way, (3) content structuring supports decisions that need to connect different aspects with each other (e.g. identifying the influence of different product features on customers´ favourability) (Rathore et al., 2018). Mining the inter-relatedness of individual topics can help to detect subtopics (Anoop et al., 2015, Nolasco and Oliveira, 2019, Park et al., 2015, Rathore and Ilavarasan, 2017, Shahbaznezhad, 2016, Srijith et al., 2017) that can be investigated more closely as driving or inhibiting factors. Hereby, it is also possible to identify niche topics at a finer level of granularity of the topical structure. In line with that, recognising the properties of subevents can enrich the understanding of the main event and to create a powerful knowledge about the scenario (Nolasco and Oliveira, 2019, Srijith et al., 2017). In general, corresponding techniques need to extract topics and at the same time establish connections between them. In line with that, the relationships identified between the topics should be understandable and applicable.

In the next step, the identified requirements (cf. tab. 1) are transformed into corresponding metrics that are necessary for evaluating and comparing the topic modelling techniques in section 5.3.

	Requirements	Sources	Metrics
(1) Content Extraction	(a) Cover all relevant and the most frequent topics embedded in textual social media data	Gao et al. 2012, Ibrahim and Wang, 2019, Irawan et al. 2020, Lee et al., 2016, Liu et al., 2017, Wang et al. 2016, Yang et al. 2016	log-likelihood
	(b) Provide a current overview of events and insights about customers' wishes and complaints	Aiello et al., 2013, Chae and Park, 2018, Cirqueira et al., 2017, Ibrahim and Wang, 2019, Ko et al., 2017, Lee et al., 2016, Liu et al., 2017, Rathore et al., 2018, Rathore and Ilavarasan, 2017	
	(c) Support comparisons with competitors and assessments of one's position in the market	Aiello et al., 2013, Ko et al., 2017, Rathore et al., 2018	
(2) Trend Analysis	(d) Contextualise the extracted topics with additional parameters	Cheng et al., 2020, Hong et al., 2012, Lozano et al., 2017, Luo et al., 2015, Wang et al., 2007, Wang et al., 2012, Zhang et al., 2016, Zhang et al., 2017, Zhong and Schweidel, 2020	build time, coherence, word intrusion
	(e) Support a flexible inclusion of different parameters (e.g. authors, locations or time)	Cheng et al., 2020, Lozano et al., 2017	
	(f) Support quick information provision	Cheng et al., 2020, Wang et al., 2012, Zhang et al., 2017	
	(g) Support quick comprehension of contents	Jeong et al., 2019, Lozano et al., 2017	
	(h) Support continuous tracking of trends and developments	Wang et al., 2012, Zhang et al., 2016, Zhang et al., 2017, Zhong and Schweidel, 2020	
(3) Content Structuring	(i) Identify niche topics at a finer level of granularity of the topical structure	Anoop et al., 2015, Nolasco and Oliveira, 2019, Park et al., 2015, Rathore et al., 2018, Rathore and Ilavarasan, 2017, Shahbaznezhad, 2016, Srijith et al., 2017	topic intrusion
	(j) Identify meaningful relationships and the inter-relatedness of topics		
	(k) Cover all aspects that are semantically related to the extracted topics		

Table 1. Identified corporate use cases and their requirements

When evaluating a model with respect to (1) content extraction, the ability of the respective technique to (a) cover all relevant and the most frequent topics embedded in textual social media data is required. An excessive number of topics leads to the generation of not only relevant but also irrelevant topics. If the number of topics is too small, however, the given overview of topics will lack relevant content (cf. Liu et al., 2017, Yang et al., 2016). Therefore, the researcher must ensure that all relevant topics within an underlying dataset are considered within the analysis and thus integrated within the resulting extraction, which (b) provides a comprehensive and current overview of customers' wishes and complaints. This results in a comprehensive, decision driven base of information which supports (c) comparisons with competitors and assessments of one's position in the market. In order to evaluate this descriptive ability, the metric of log-likelihood is used. Using this, it is possible to quantify how accurately a model can represent the underlying data and thus models all relevant information (cf. Daud et al., 2010, Wallach et al., 2009). Furthermore, the evaluation has to consider different circumstances regarding the number of topics to be identified. Therefore, the evaluation of each technique takes place multiple times, with a continuously increasing number of topics. Thus, the strengths and weaknesses in modelling low (high) numbers of topics and thus lowly (highly) differentiated contextual information can be assessed.
When evaluating a model with regard to the described use case of (2) trend analysis, (d) the ability to contextualise the extracted topics and (e) hereby flexibly include different additional parameters (e.g. authors, location or time) is mentioned. Furthermore, corresponding techniques need to enable time-critical reactions to emerging circumstances so that (f) the provision of the topics should be as quick as possible. Therefore, the analysis of the different techniques is twofold. On the one hand, authors refer to the coherence measure (Dahal et al., 2019, Paul and Girju, 2009, Wang et al., 2007), which describes the property of the respective technique to generate topics that correlate well with the human understanding of semantically coherent topics. This results in a semantically meaningful and sound analysis output that (g) supports time-critical decisions and does not need further investigations to be applicable. On the other hand, to validate the calculated reasonability of the respective analysis output, we incorporate word intrusion (Chang et al., 2009). Thus, multiple subjects evaluate the consistency of the extracted topics by analysing the associated words. The objective is to identify the so-called 'intruder' within the topic, which is represented by a single word without contextual relevancy regarding the intruded topic. If the subjects are able to identify the respective intruder, the evaluated topic is consistent with the human understanding of a meaningful and sound topic. Besides the unhindered applicability of the analysis results, we also classified the build time as a time-critical evaluation metric. So, a short build time results in a timely output supporting faster decision-making. Beyond the aforementioned time-critical requirements, the ability to contextualise the extracted topics and (h) to continuously track their development is also required. To assess the possibility of accounting for further contextual information such as geological or time-based data, we qualify the ability to incorporate external information.
Concerning an evaluation of a model with regard to the described use case of (3) content structuring, the ability of the techniques to reveal the underlying structure within the data is focused. Thus, (i) to be able to extract hidden niche topics to identify relationships between the discussed topics, the ability of the respective technique to identify relationships within the data is qualified. Besides the pure ability to identify relationships, the assessment of the (j) meaningfulness of the identified relations is also required. In this regard, we opted to use the topic intrusion approach, which measures how well a topic model's decomposition of a document as a mixture of topics agrees with human associations of topics related within a document (Chang et al., 2009). Using these techniques, the (k) coverage of all aspects within the extracted topics as well as their interrelationships in the relevant document are analysed. The procedure of the analysis is similar to that of word intrusion. Specifically, the subjects are presented the document title alongside a short extract thereof. In addition to the document information, the subject receives four topics, of which three are the most probable topics assigned to the document and the remaining topic embodies the intruder topic to be identified. If the subjects are able to identify the erroneously listed topic, the topics and their contextual relationship to one another are meaningful and sound. A more detailed insight into the tasks of word and topic intrusion can be found in section 5.2.
As our review shows, different use cases and corresponding requirements have been reported in line with the literature that provide insights into the application of topic modelling techniques in marketing. However, the used data sets and the applied topic modelling techniques vary across the different papers,

so that recommendations made on this basis may not be sufficiently reliable. Therefore, within the research at hand, three different topic modelling techniques are applied to the same data set to give recommendations that are not only theory-driven but also based on the results of a data analysis.

3 Procedure of the Research

Our investigation follows the '*Goal Question Metric*' (GQM) approach outlined by Basili et al. (1994) for a systematic development of metrics for conducting a comparison of topic modelling techniques. The GQM approach is based on the idea that measurements in an organisational context depend on a thoroughly defined goal firstly operationalised by relevant enterprise data, which are then interpreted regarding the goal (Basili et al., 1994). Thus, this approach focuses on which informational needs a company exhibit in order to quantify them and consequently examine if the quantified information meets the goals or not. Especially for our investigation, the GQM approach is well qualified as it assures a systematic research procedure in which reproducible results are achieved. According to Basili et al. (1994) the GQM approach is divided into three different levels:

1. Conceptual level (**GOAL**): We set our goal as the development of means to compare different topic modelling techniques with each other regarding different use cases. Therefore, with this investigation we seek to highlight differences in the organisational application between the three different topic modelling techniques LDA, DMR, and PAM.
2. Operational level (**QUESTION**): In order to characterise how the assessment of our goal is performed, we formulate a question. In our study, the requirements of the different use cases for the techniques identified in the literature form the basis for our questions of the GQM approach: 'To what extent do the three topic modelling techniques (LDA, DMR, PAM) meet the requirements of the different use cases (a-k) (cf. tab. 1)?'
3. Quantitative level (**METRIC**): To answer this question, we conducted a quantitative analysis of the topic modelling techniques, that helps us to evaluate to what extent a selected topic modelling technique can meet the requirements. Subsequently, in order to meet these requirements, that we have already identified in a previous step, we now need proper metrics. Therefore, we consulted the research literature and established specific metrics for the requirements to enable a comprehensible comparison. Consequentially, we will be able to formulate recommendations for the allocation of the topic modelling techniques to the use cases.

4 Selection of the Topic Modelling Techniques

As mentioned above, the field of topic modelling has many different techniques, which all try to identify specific topics within large sets of text data by reducing the dimensionality and attaching different weights to the specific data set (Crain et al., 2012). In order to optimally meet the identified use cases, the selection of the techniques to be used is critical to success. Besides the aforementioned LDA with its extensions, a variety of different categories of topic modelling techniques like Latent Semantic Analysis, Probabilistic Latent Semantic Analysis, Correlated Topic Models, Dynamic Topic Models, or Topic Evolution Model exist (cf. Alghamdi and Alfalqi, 2015). Although they potentially could offer benefits in terms of different applications, most approaches lack a ready-to-use implementation or require an advanced technique-related knowledge and therefore suitable applicability for companies is not given. Differently LDA, where a clear dominance in the use has become apparent, as it offers simple applicability and good analysis results (Eickhoff and Neuss, 2017). Due to the multifaceted challenges to be mastered in the analysis of text data, different extensions of the basic LDA procedure have been developed over time which are suitable for the solution of different scenarios depending on their extending characteristics. Generally, a distinction is made between three expanding properties: (I) extension of topic attributes, (II) extension of document attributes, and (III) extension of word attributes (cf. Liu et al., 2016). In order to answer **RQ2**, the identification of topic modelling approaches for processing certain corporate use cases, we decided to compare LDA as well as selected extensions with respect to the identified use cases. Therefore, to achieve optimal coverage of different techniques, we selected one specific technique for each extension class.

The (I) topic-based extensions deal with the mapping of relations within the latent topics of a corpus. In this context, a variety of techniques is highlighted in literature (cf. Griffiths et al., 2004, Liu et al., 2016). The focus is on the identification of relationships between the inferred topics allowing a hierarchical representation of them. Mimno et al. (2007) compared the ability of their PAM algorithm to represent a hierarchical data structure and to predict a topic distribution for new data not included in the training set with a variety of techniques of the same extension class (cf. Mimno et al., 2007). Since PAM was characterised by better evaluation results, we decided to choose PAM as the topic-related extension. PAM represents the relationships among the topics as directed graphs, which allows the representation of a hierarchical structure within the topics.
The extension based on documents (II) enables the consideration of document-specific meta-information such as authors, document titles, points in time, or geographical information (cf. Liu et al., 2016). In this context, approaches such as the author-topic model (cf. Rosen-Zvi et al., 2012), Topics over Time (cf. Wang and McCallum, 2006), embedded topic model (Dieng et al., 2020) as well as DMR (cf. Mimno and McCallum, 2008) are highlighted in the current research literature. Because the technique presented by Mimno and McCallum (2008) is more flexible with respect to the incorporation of additional information as well as performs remarkable in terms of information quality, DMR is chosen for the implementation of (II) extension of document attributes. DMR is an upstream topic model with a particularly attractive technique for integrating any document features. Instead of defining specific random variables in the graphical model for each new document feature, DMR treats the document annotations as features in a log-linear model. The log-linear model parameterises the Dirichlet before the document's topic distribution, making the Dirichlet's hyperparameter document-specific. Since no assumptions are made about the model structure of new random variables, DMR is flexible to include various types of features, resulting in a flexible use of DMR (Benton and Dredze, 2018).
The above-mentioned topic modelling techniques, which are all based on the BoW approach, do not consider the order of words within a document. This resulted in extension (III), attempting to eliminate the interchangeability of words. Therefore, Wallach (2006) argued that the consideration of word orders in the form of bi-grams can lead to improved results when using a topic modelling approach. Since the consideration of word orders in the form of bi-grams did not show any difference with regard to the generated topics and the underlying topic quality compared to LDA, it will be equated with the use of the basic technique in the following. By choosing these techniques, a selection was made which considers each extension class of the basic approach, whereby a broad spectrum of different techniques is compared with regard to their applicability against diverse use cases.

5 Data Analysis

5.1 Data collection

To identify the potential of the different techniques with regard to the applicability to different use cases, an existing data set of Facebook posts was used (cf. Martinchek, 2017). This comprises 4,155,992 documents from the 15 most popular news services in the United States of America for the period from 2012 to 2016. The raw text dataset contains information such as the respective picture URL or the like count, which are not relevant for the application of the implemented topic modelling approaches. Therefore, to reduce the dimensions of the data, a custom converter was developed and applied to the data. The resulting data set contains three parameters after conversion: the ID of the respective document, the respective year – which serves as feature input to determine the topic relevance at different points in time within the analysis via DMR – and the description – which reflects the actual text of the contribution. The following excerpt from the training data set gives an insight into the data (cf. tab. 2).

ID	Year	Description
52921	2016	Dow Drops More Than 300 Points Following Market Rout...
21049	2014	How to Greet People During Flu Season: Handshake, ...

Table 2. Structure of training data

5.2 Data cleansing and analysing

In order to compare the topic modelling techniques empirically on the basis of their analysis results, the data must first be prepared. With respect to this, we applied tokenisation, stopwords removal and case folding as proposed by Boyd-Graber et al. (2014). As the use of stemming procedures does not improve the interpretability of the results, but can potentially even deteriorate the topic stability (Schofield and Mimno, 2016), we did not incorporate stemming.

By conducting this comparison between LDA, DMR, and PAM based on the mentioned evaluation measures, we aim to reveal the strengths and weaknesses of the different techniques in terms of identifying embedded topics within written social media data. As it is necessary to provide similar conditions for a comparison to be valid, all techniques were configured with their default parameters and trained with iteratively increasing numbers of topics. Accordingly, the selected topic range includes 10, 30, 50, 100, and 300 topics (*k*). Furthermore, all evaluation metrics were validated by cross-validation to eliminate the choice of a potentially non-representative test dataset (Bramer, 2007). The evaluation approach is further distinguishing between intrinsic and extrinsic measures. Intrinsic evaluations measure the performance of a component on its defined subtask, usually against a defined standard in a reproducible laboratory setting. Extrinsic evaluations focus on the component's contribution to the performance of a complete application, which often involves the participation of a human in the loop (Resnik et al., 2006).

#	log-likelihood measurements ($\cdot 10^6$)			coherence measurements			build time (min)		
k	LDA	DMR	PAM	LDA	DMR	PAM	LDA	DMR	PAM
10	- 4.86	- 5.03	- 4.85	- 197.96	- 198.50	- 208.34	342	363	832
30	- 4.49	- 4.61	- 4.41	- 232.84	- 220.75	- 236.81	571	634	1,264
50	- 4.29	- 4.40	- 4.19	- 244.29	- 217.90	- 251.17	912	1,083	1,992
100	- 4.06	- 4.10	- 3.88	- 241.74	- 213.03	- 254.92	1,407	1,732	3,481
300	- 3.91	- 3.62	- 3.67	- 215.02	- 186.95	- 241.52	3,180	3,821	6,984

Table 3. Intrinsic evaluation measurements

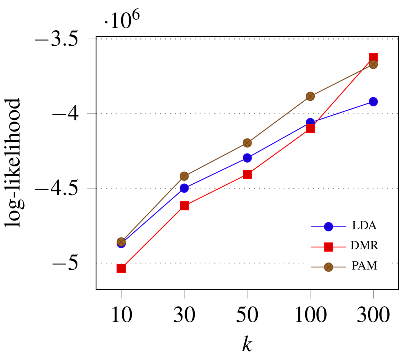

Figure 1. log-likelihood

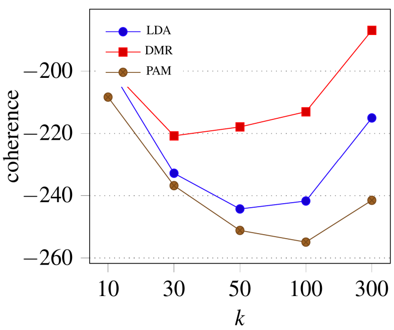

Figure 2. coherence

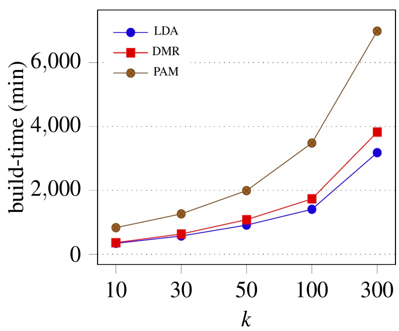

Figure 3. build time

In the first step of the evaluation, the models are compared according to their ability to represent the underlying data. Therefore, the metric of log-likelihood was applied (cf. tab. 3, fig. 1). A higher value represents a better model with regard to the ability to adequately represent the underlying data. The best model here is DMR, with k = 300 (-3.62). LDA performs better than DMR (e.g. k = 30 ≙ -4.86 > -5.03) and worse than PAM (e.g. k = 30 ≙ -4.49 < -4.41) until the number of topics exceeds 100. PAM and LDA have in common that, as the number of topics increases, the log-likelihood improves iteratively (cf. fig. 1). The increase in the number of topics for DMR results in a continuous improvement of the log-likelihood as well. Nevertheless, DMR always outperforms PAM and LDA when k >~ 300. This result implies that DMR is the technique that can be recommended when tasks require many different topics to be identified. However, within tasks that require a lower number of topics to be generated, our results indicate that PAM and LDA may be superior to DMR.

Besides the general ability of the models to adequately represent the underlying data, the semantic quality of the generated topics was evaluated. For that reason, the coherence measure was applied (cf. fig.

2). Therefore, we conclude that DMR generates topics, that are more coherent in general. LDA has a slightly better coherence for k = 10. However, any modelling of k > 10 is outperformed by DMR. Further, PAM performed worst in terms of coherence. Here, all different circumstances of topics are dominated by the other techniques. Nevertheless, PAM generated its most relatable topics for k = 10, which leads to the conclusion, that PAM, similar to LDA, exhibits its strength in modelling low topic dimensions.
To react in a time-critical manner to changing circumstances, it is also necessary to acquire supporting information as quickly as possible. Therefore, to measure the time an approach needs to extract the required information (cf. fig. 3), the build time of each technique is tracked. The respective build time includes data preprocessing, the actual model building and cross-validation. A lower build time indicates a quicker model training, which results in faster information provision. In this regard, LDA outperforms the other two techniques in terms of information extraction time. The difference between LDA and DMR in the lower range of topics is negligible (21 min), but the higher the value of topics, the larger the difference. The difference in the context of k = 300 amounts to 641 minutes. Because of the hierarchies to be modelled by PAM, it generally takes much longer to extract the respective information in comparison to LDA and DMR. Here, PAM needs at least twice as much time under almost all circumstances.
Besides the intrinsic measurements of log-likelihood, coherence and build time, the topic modelling techniques were further assessed by humans to evaluate the semantic quality of the generated topics as well as their interrelationships. Therefore, word and topic intrusion procedures were performed. The respective survey was undertaken by two researchers and administered to 18 participants, all of whom evaluated the semantic coherence of three randomly selected topics (word intrusion) and the decomposition of a single document into its topics and the corresponding relationships (topic intrusion). Each survey ranged between 38 and 51 minutes. The topics as well as the respective excerpts of a document were extracted randomly for each trained model. Further, to account the inter-rater reliability of the results, all participants evaluated the same set of topics or documents respectively for each trained model. Regarding the word intrusion task, the subjects had to identify the intruder within the topics that did not cohere to the semantics of the other presented words. The corresponding results for word intrusion are calculated as the sum of the correctly classified intruders by the test subjects in relation to the total number of tests per model. By analysing three topics per trained model, a total of 810 individual observations were carried out. Regarding the topic intrusion task, the participants had to identify the intruding topic by reading a document title alongside a short extract thereof. The respective document was randomly extracted for each trained model. All participants evaluated the decomposition of the same documents. By this, a total of 270 individual observations are accomplished. The respective results of the topic intrusion task are calculated as the amount of correctly classified intruder topics in relation to the total number of observations per model.

	Word intrusion (%)			Topic intrusion (%)		
k	LDA	DMR	PAM	LDA	DMR	PAM
10	74.0	85.1	53.7	83.3	72.2	83.3
30	72.2	79.6	57.4	66.6	61.1	72.2
50	62.9	75.9	51.8	72.2	72.2	72.2
100	64.8	70.3	48.1	55.5	66.6	55.5
300	55.5	77.7	46.2	55.5	72.2	44.4

Table 4. Extrinsic evaluation measurements

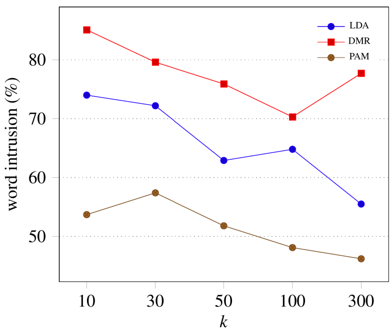

Figure 4. Word intrusion

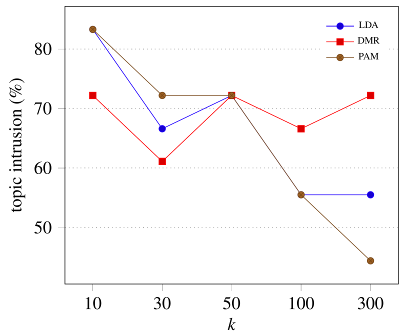

Figure 5. Topic intrusion

Regarding the word intrusion task, it becomes apparent that the previously determined values of coherence are in line with human understanding so DMR generates the most semantically coherent topics. Specifically, the best topics are generated by DMR with k = 10 (cf. fig. 4). The achieved model precision is 85.1% for k = 300. The evaluated minimum precision of DMR is 70.3%. As the results for word intrusion do not show any distinct extremes for different values of k, every model of DMR seems to produce a coherent word-topic distribution. Besides that, LDA also proves to generate the most coherent topics for k = 10 (74.0%). The higher the number of topics to be generated, the more difficult it becomes for the subjects to determine the intruder. This leads to the assumption, that the topics generated by LDA will be increasingly difficult to interpret with a rising number of k, thereby losing their semantic coherence and therefore their meaningfulness. The minimal precision of LDA is recorded for k = 300. Here, only 30 of 54 measurements are classified correctly, resulting in a model precision of 55.5%. There is a clear discrepancy in the quality of topics generated by PAM. The semantically most coherent topics are determined by the model with k = 30. The model precision achieved is 57.4%. The minimum semantic coherence of the topics is with k = 300, with a precision of 46.2%. Thus, it can be concluded that PAM, similar to LDA, has its strength in modelling a smaller number of topics but is clearly inferior to DMR for larger numbers. Besides the evaluation of the topic quality, their decomposition and interrelationships were evaluated. Therefore, topic intrusion tasks were conducted, where the best results were obtained with 83.3% for LDA and PAM with k = 10. Further, it is remarkable, that all techniques achieve the same score with k = 50 (cf. fig. 5), which leads to the conclusion, that all techniques generate consistent topic decompositions for a medium number of topics. The higher the number of topics to be generated, the worse LDA and PAM perform. The worst result is achieved by PAM with k = 300. That decrease of PAM for larger numbers of topics to be generated can be traced back to the formation of the many hierarchical levels, since a high number of k also means that a correspondingly large number of relationships between the different topics must be inferred.

6 Discussion

By evaluating the different techniques, the strengths and weaknesses (cf. tab. 5) of them were identified, that provide information about the applicability of the techniques related to the identified corporate use cases (1) content extraction, (2) trend analysis as well as (3) content structuring settled in marketing.

	Requirements	**LDA**	**DMR**	**PAM**
(1) Content Extraction	(a) Cover all relevant and the most frequent topics embedded in textual social media data	●	●	◗
	(b) Provide a current overview of events and insights about customers' wishes and complaints	●	●	◗
	(c) Support comparisons with competitors and assessments of one's position in the market	●	●	◗
(2) Trend Analysis	(d) Contextualise the extracted topics with additional parameters	○	●	○
	(e) Support a flexible inclusion of different parameters (e.g. authors, locations or time)	○	●	○
	(f) Support quick information provision	●	●	○
	(g) Support quick comprehension of contents	●	●	○
	(h) Support continuous tracking of trends and developments	◗	●	○
(3) Content Structuring	(i) Identify niche topics at a finer level of granularity of the topical structure	○	◗	●
	(j) Identify meaningful relationships and the inter-relatedness of topics	○	○	●
	(k) Cover all aspects that are semantically related to the extracted topics	○	◗	●

Table 5. Results of the comparison ●: applies fully ◗: applies partly ○: applies not

For the first use case – (1) content extraction – the models′ ability to provide an adequate overview of customers' major wishes, ideas, and thoughts to cover all relevant topics embedded in a collection of written social media data was evaluated. Therefore, we applied the metric of log-likelihood, which describes the ability of a model to represent the underlying data as appropriately as possible. This guarantees that the generated output contains all necessary information and that no relevant topics are missing. The standard procedure of LDA as well as that of PAM present their strength compared to DMR in the extraction of topics in the lower range of generated topics. DMR, in contrast, has its strength in the representation of large numbers of topics. This leads to the conclusion that the selection of the best fitting technique depends on the needs to be fulfilled within the extraction task:

- If the content must be very specific, e.g. to support a comparison with competitors and assessments of one's position in the market, a high number of topics is required. So, DMR should be considered.
- If the task requires an extraction on a more abstract level, a low number of topics will mostly be satisfying. If so, the usage of LDA and PAM could be considered. Due to the higher semantic coherence within the topics generated by LDA (cf. tab. 3, tab. 4, fig. 2, fig. 4), the use of LDA is recommended with respect to (1) content extraction for small dimensions.
- The use of PAM is considered as partly applicable, but not recommended, as it is outperformed by LDA and DMR for the criteria being evaluated.

Regarding the second use case (2) trend analysis, the ability of the techniques to generate immediately meaningful and sound output was analysed, which results e.g., in a quick decision supporting information base. This is indispensable regarding the need of time-critical actions within the volatile characteristic of trends. Hereby, DMR shows an advantage leading to the conclusion, that it generates the most reasonable topics (cf. tab. 3, fig. 2). To validate the collected intrinsic evaluation results, the techniques were further evaluated by humans regarding their semantic quality and soundness in an extrinsic way. Here, DMR could be confirmed to generate the most comprehensible output (cf. tab. 4, fig. 4). Further, LDA and PAM show their strength for modelling low numbers of topics as both techniques achieve their best results in the range of 10 (LDA) and 30 (PAM) topics. In addition to direct applicability, the time required for a technique to provide the necessary information was also accounted for. In this respect, LDA provides the fastest output, followed by DMR. The measured discrepancy between these two techniques is negligible for a small number of topics ($k = 10 \triangleq 21$ min.), but the larger the number of topics, the larger the gap ($k = 300 \triangleq 641$ min.). Since the focus of trend analysis is on identifying individual trends and tracking their development, the number of topics will not be that high. Therefore, LDA and DMR are considered capable of reacting to rapidly changing circumstances. PAM, however, requires at least twice as much time for each condition and is therefore not suitable. The strengths and weaknesses of the techniques in the context of (2) trend analysis are represented as follows:

- DMR provides the most reasonable and meaningful topics and is further able to provide them quickly. Additionally, DMR has the advantage of taking external parameters into account. This allows, e.g., tracking the development of topics over a certain period of time or based on geolocation data. Therefore, DMR should be considered regarding (2) trend analysis.
- If the contextualisation of topics does not apply, LDA can also be used for tracking trends and their development, as it can be used to quickly identify meaningful and sound analysis results. Therefore, LDA can be seen as partly applicable regarding trend analysis.
- PAM is not suitable due to the amount of time required and the lack of ability to contextualise topics.

Regarding the last identified use case – (3) content structuring - the ability of the techniques to reveal meaningful relationships and the inter-relatedness of topics was evaluated. Therefore, the extrinsic evaluation metric of topic intrusion was applied. By doing so, it could be guaranteed that all semantically related aspects were extracted. Here, PAM shows a slightly better result than the two remaining techniques for a small number of topics to be generated (cf. tab. 4, fig. 5). Further, DMR underlined its strength in the representation of a high number of topics. A large advantage of using PAM is represented by its ability to model hierarchical structures within the topics themselves. Thus, it is possible to extract general topics as well as their respective subtopics, whereby the topics can be divided into different, thematically consistent groupings that can support the identification of niche topics at a finer level of granularity. The elicited strengths and weaknesses of the techniques are listed in the following:

- As the best evaluation results regarding the topic intrusion task were achieved by PAM, we recommend this technique as an approach to support (3) content structuring.
- Besides that, DMR can also be applied if the number of topics to be generated reaches a comparatively high number. Thus, the applicability is considered partial.
- Since LDA is surpassed by the two techniques here, the application is not considered suitable.

In summary, all the investigated techniques have different strengths and weaknesses in their applicability to the identified use cases. Nevertheless, LDA showed its strength in modelling low-dimensional topics. In comparison, DMR showed to be superior in representing high-dimensional topics. Regarding trend analysis, DMR showed its strength within the generation of semantically meaningful results. For content structuring tasks, PAM showed superior results in extracting meaningful relationships compared to LDA and DMR.

7 Conclusion and Outlook

Analysing written social media data with automated techniques has massively gained in importance as being aware of customers' wishes is no longer sustainable with manual analysis due to the sheer volume of available posts. Topic modelling has shown to be an adequate instrument to support these tasks by extracting the topics discussed within documents (e.g. Eickhoff and Neuss, 2017, Vayansky and Kumar, 2020). However, it can be observed that especially LDA has been given particular attention for marketing-specific applications. Furthermore, since LDA cannot cover all fields of the identified use cases, marketing tasks may only be supported to a limited extent by automated approaches.

Within this work, the use cases (1) - (3) were identified from the literature due to the frequency that the identified papers related to them. For each of these use cases, corresponding requirements were identified and assigned to different metrics (log-likelihood, coherence, build time, word and topic intrusion) for evaluating these topic modelling techniques (cf. **RQ1**). Thus, LDA, DMR and PAM were applied to a real-world data set, evaluated and compared with each other. Thereafter, this work gives recommendations regarding which topic modelling technique could be applied for which use case (cf. **RQ2**).

Through our comparison of topic modelling techniques, practitioners are given means to select a technique that can best support their daily business activities. Decision makers in marketing can classify their concrete task into one of the three identified use cases and derive a recommendation for a suitable technique. Tasks in marketing, which can be enriched by topic modelling, can thus be supported more optimally and thus the performance of this division, which is so important for companies, can be increased. Beyond creating value for practitioners, theoretical contributions in the research area of IS are also provided. First, based on the use cases we derived several requirements for topic modelling techniques and assigned several evaluation criteria to each of them. Second, in order to provide recommendations, we compared LDA, DMR, and PAM with each other regarding five different evaluation metrics by analysing a real-world data set. This comparison is based on the GQM approach and assures therefore a systematic research procedure in which reproducible results are achieved. Based on this comparison potentials for further enhancements (e.g. considering a faster build time for PAM when $k > 50$) could be evolved. Third, we further deduced strengths (such as DMR should be considered regarding trend analysis, cf. section 6) and weaknesses (such as PAM is not suitable for trend analysis due to the amount of time required and the lack of ability to contextualise topics) of the three topic modelling techniques which is valuable for both researchers and practitioners.

There are some limitations to this study: first, the number of papers we incorporated in identifying use cases and related requirements for topic modelling within marketing is limited. Nevertheless, these requirements enabled the central metrics of topic modelling techniques to be assigned and generally valid recommendations for appropriate procedures to be derived. Second, the number of topic modelling techniques being compared is limited to three. Although the extensions proposed by Liu et al. (2016) could thus be covered to a large extent, we plan to include further techniques in future analysis for each extension and thereby further refine our recommendations.

References

Aiello, L. M., Petkos, G., Martin, C., Corney, D., Papadopoulos, S., Skraba, R., Göker, A., Kompatsiaris, I. and Jaimes, A. (2013). Sensing trending topics in Twitter. *IEEE Transactions on Multimedia,* 15 (6), 1268-1282.

AlFalahi, K., Atif, Y. and Abraham, A. (2014). Models of influence in online social networks. *International journal of intelligent systems,* 29 (2), 161-183.

Alghamdi, R. and Alfalqi, K. (2015). A survey of topic modeling in text mining. *Int. J. Adv. Comput. Sci. Appl.(IJACSA)*, 6 (1).

Anoop, V., Asharaf, S. and Alessandro, Z. (2015). Generating and visualizing topic hierarchies from microblogs: An iterative latent dirichlet allocation approach. *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 824-828.

Basili, G., Caldiera, V. and Rombach, H. D. (1994). The goal question metric approach. *Encyclopedia of software engineerin*g, 528-532.

Benton, A. and Dredze, M. (2018). Deep Dirichlet multinomial regression. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1, 365-374.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM,* 55 (4), 77-84.

Blei, D. M., Ng, A. Y. and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 993-1022.

Boyd-Graber, J., Mimno, D. and Newman, D. (2014). Care and feeding of topic models: Problems, diagnostics, and improvements. *Handbook of mixed membership models and their applications*, 225255.

Bramer, M. (2007). Principles of data mining. *Springer*, ISBN: 978-1-84628-765-7.

Chae, B. K. and Park, E. O. (2018). Corporate social responsibility (CSR): A survey of topics and trends using Twitter data and topic modeling. *Sustainability,* 10 (7), 2231.

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L. and Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. *Advances in neural information processing systems*, 288-296.

Cheng, L., Li, J., Candan, K. S. and Liu, H. (2020). Tracking Disaster Footprints with Social Streaming Data. *Proceedings of the AAAI Conference on Artificial Intelligence*, 370-377.

Chinnov, A., Kerschke, P., Meske, C., Stieglitz, S. and Trautmann, H. (2015). An Overview of Topic Discovery in Twitter Communication through Social Media Analytics. *Americas Conference On Information Systems (AMCIS) 2015.*

Cirqueira, D., Pinheiro, M., Braga, T., Jacob Jr, A., Reinhold, O., Alt, R. and Santana, Á. (2017). Improving relationship management in universities with sentiment analysis and topic modeling of social media channels: learnings from ufpa. *Proceedings of the International Conference on Web Intelligence*, 998-1005.

Coelho, L. P., Peng, T. and Murphy, R. F. (2010). Quantifying the distribution of probes between subcellular locations using unsupervised pattern unmixing. *Bioinformatics,* 26 (12), 7-12.

Crain, S. P., Zhou, K., Yang, S.-H. and Zha, H. (2012). Dimensionality reduction and topic modeling: From latent semantic indexing to latent dirichlet allocation and beyond. *Mining text data*, Springer, 129-161.

Culnan, M. J., McHugh, P. J. and Zubillaga, J. I. (2010). How large US companies can use Twitter and other social media to gain business value. *MIS Quarterly Executive,* 9 (4).

Dahal, B., Kumar, S. A. and Li, Z. (2019). Topic modeling and sentiment analysis of global climate change tweets. *Social Network Analysis and Minin*g, 9 (1), 1-20.

Dai, Y., Kakkonen, T. and Sutinen, E. (2011). MinEDec: a decision-support model that combines text-mining technologies with two competitive intelligence analysis methods. *International Journal of Computer Information Systems and Industrial Management Applications,* 3 (3), 165-173.

Daud, A., Li, J., Zhou, L. and Muhammad, F. (2010). Knowledge discovery through directed probabilistic topic models: a survey. *Frontiers of computer science in China,* 4 (2), 280-301.

Debortoli, S., Müller, O., Junglas, I. and vom Brocke, J. (2016). Text mining for information systems researchers: An annotated topic modeling tutorial. *Communications of the Association for Information Systems,* 39 (1), 7.

Di Gangi, P. M., Wasko, M. M. and Hooker, R. E. (2010). GETTING CUSTOMERS' IDEAS TO WORK FOR YOU: LEARNING FROM DELL HOW TO SUCCEED WITH ONLINE USER INNOVATION COMMUNITIES. *MIS Quarterly Executive,* 9 (4).

Dieng, A. B., Ruiz, F. J. and Blei, D. M. (2020). Topic modeling in embedding spaces. *Transactions of the Association for Computational Linguistics*, 8, 439-453.

Eickhoff, M. and Neuss, N. (2017). Topic modelling methodology: its use in information systems and other managerial disciplines. *European Conference On Information Systems (ECIS) 2017.*

Einwiller, S. A. and Steilen, S. (2015). Handling complaints on social network sites–An analysis of complaints and complaint responses on Facebook and Twitter pages of large US companies. *Public Relations Review,* 41 (2), 195-204.

Gao, W., Li, P. and Darwish, K. (2012). Joint topic modeling for event summarization across news and social media streams. *Proceedings of the 21st ACM international conference on Information and knowledge management*, 1173-1182.

Ghosh, D. and Guha, R. (2013). What are we 'tweeting'about obesity? Mapping tweets with topic modeling and Geographic Information System. *Cartography and geographic information science,* 40 (2), 90-102.

Girdhar, Y., Giguere, P. and Dudek, G. (2013). Autonomous adaptive underwater exploration using online topic modeling. *Experimental Robotics*, 789-802.

Grégoire, Y., Salle, A. and Tripp, T. M. (2015). Managing social media crises with your customers: The good, the bad, and the ugly. *Business Horizons,* 58 (2), 173-182.

Griffiths, T. L., Jordan, M. I., Tenenbaum, J. B. and Blei, D. M. (2004). Hierarchical topic models and the nested chinese restaurant process. *Advances in neural information processing systems*, 17-24.

Guesalaga, R. (2016). The use of social media in sales: Individual and organizational antecedents, and the role of customer engagement in social media. *Industrial Marketing Management,* 54, 71-79.

Hanna, R., Rohm, A. and Crittenden, V. L. (2011). We're all connected: The power of the social media ecosystem. *Business Horizons,* 54 (3), 265-273.

Hienerth, C., Keinz, P. and Lettl, C. (2011). Exploring the nature and implementation process of user-centric business models. *Long Range Planning,* 44 (5-6), 344-374.

Hong, L., Ahmed, A., Gurumurthy, S., Smola, A. J. and Tsioutsiouliklis, K. (2012). Discovering geographical topics in the twitter stream. *Proceedings of the 21st international conference on World Wide Web*, 769-778.

Hong, L. and Davison, B. D. (2010). Empirical study of topic modeling in twitter. *Proceedings of the first workshop on social media analytics*, 80-88.

Ibrahim, N. F. and Wang, X. (2019). A text analytics approach for online retailing service improvement: Evidence from Twitter. *Decision Support Systems,* 121, 37-50.

Irawan, M., Wijayanto, R., Shahab, M., Hidayat, N. and Rukmi, A. (2020). Implementation of social media mining for decision making in product planning based on topic modeling and sentiment analysis. *Journal of Physics: Conference Series*, 012068.

Jeong, B., Yoon, J. and Lee, J.-M. (2019). Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment analysis. *International Journal of Information Management,* 48, 280-290.

Kaplan, A. M. and Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons,* 53 (1), 59-68.

Kietzmann, J. H., Hermkens, K., McCarthy, I. P. and Silvestre, B. S. (2011). Social media? Get serious! Understanding the functional building blocks of social media. *Business Horizons,* 54 (3), 241-251.

Ko, N., Jeong, B., Choi, S. and Yoon, J. (2017). Identifying product opportunities using social media mining: application of topic modeling and chance discovery theory. *IEEE Access,* 6, 1680-1693.

Krestel, R., Fankhauser, P. and Nejdl, W. (2009). Latent dirichlet allocation for tag recommendation. *Proceedings of the third ACM conference on Recommender systems*, 61-68.

Kumar, V. and George, M. (2007). Measuring and maximizing customer equity: a critical analysis. *Journal of the Academy of Marketing Science,* 35 (2), 157-171.

Lee, T.-H., Sung, W.-K. and Kim, H.-W. (2016). A text mining approach to the analysis of information security awareness: Korea, United States, and China. *Pacific Asia Conference On Information Systems (PACIS) 2016*.

Li, W. and McCallum, A. (2006). Pachinko allocation: DAG-structured mixture models of topic correlations. *Proceedings of the 23rd international conference on Machine learning*, 577-584.

Liu, L., Tang, L., Dong, W., Yao, S. and Zhou, W. (2016). An overview of topic modeling and its current applications in bioinformatics. *SpringerPlus,* 5 (1), 1608.

Liu, X., Burns, A. C. and Hou, Y. (2017). An investigation of brand-related user-generated content on Twitter. *Journal of Advertising,* 46 (2), 236-247.

Lozano, M. G., Schreiber, J. and Brynielsson, J. (2017). Tracking geographical locations using a geo-aware topic model for analyzing social media data. *Decision Support Systems,* 99, 18-29.

Luo, J., Pan, X. and Zhu, X. (2015). Identifying digital traces for business marketing through topic probabilistic model. *Technology Analysis & Strategic Management,* 27 (10), 1176-1192.

Marshall, G. W., Moncrief, W. C., Rudd, J. M. and Lee, N. (2012). Revolution in sales: The impact of social media and related technology on the selling environment. *Journal of Personal Selling & Sales Management,* 32 (3), 349-363.

Martinchek, P. (2016). 2012-2016 Facebook Posts. URL: *https://data.world/martinchek/2012-2016-facebook posts*.

McDonald, M. and Aron, D. (2012). Amplifying the enterprise: the 2012 CIO agenda. *Gartner Executive Programs*, Gartner Inc.

Mimno, D., Li, W. and McCallum, A. (2007). Mixtures of hierarchical topics with pachinko allocation. *Proceedings of the 24th international conference on Machine learning*, 633-640.

Mimno, D. M. and McCallum, A. (2008). Topic models conditioned on arbitrary features with Dirichlet-multinomial regression. *UAI*, 411-418.

Mount, M. and Martinez, M. G. (2014). Social media: A tool for open innovation. *California Management Review,* 56 (4), 124-143.

Nolasco, D. and Oliveira, J. (2019). Subevents detection through topic modeling in social media posts. *Future Generation Computer Systems*, 93, 290-303.

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., Ungar, L. H. and Seligman, M. E. (2015). Automatic personality assessment through social media language. *Journal of personality and social psychology,* 108 (6), 934-952.

Paul, M. and Girju, R. (2009). Cross-cultural analysis of blogs and forums with mixed-collection topic models. *Proceedings of the 2009 conference on empirical methods in natural language processing*, 1408-1417.

Piller, F. T., Vossen, A. and Ihl, C. (2012). From social media to social product development: the impact of social media on co-creation of innovation. *Die Unternehmung,* 65 (1).

Ramage, D., Manning, C. D. and Dumais, S. (2011). Partially labeled topic models for interpretable text mining. *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 457-465.

Rathore, A. K., Das, S. and Ilavarasan, P. V. (2018). Social media data inputs in product design: Case of a smartphone. *Global Journal of Flexible Systems Management,* 19 (3), 255-272.

Rathore, A. K. and Ilavarasan, P. V. (2017). Social media analytics for new product development: case of a pizza. *International Conference on Advances in Mechanical, Industrial, Automation and Management Systems (AMIAMS)*, 213-219.

Resnik, P., Niv, M., Nossal, M., Schnitzer, G., Stoner, J., Kapit, A. and Toren, R. (2006). Using intrinsic and extrinsic metrics to evaluate accuracy and facilitation in computer-assisted coding. *Perspectives in Health Information Management Computer Assisted Coding Conference Proceedings*, 2006.

Rosen-Zvi, M., Griffiths, T., Steyvers, M. and Smyth, P. (2012). The author-topic model for authors and documents. *arXiv preprint,* 4169.

Schneider, F., Feldmann, A., Krishnamurthy, B. and Willinger, W. (2009). Understanding online social network usage from a network perspective. *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement*, 35-48.

Schofield, A. and Mimno, D. (2016). Comparing apples to apple: The effects of stemmers on topic models. *Transactions of the Association for Computational Linguistics,* 4, 287-300.

Shahbaznezhad, H. (2016). Fan page management via content generation and feedback strategies. *SSRN*, 2769428.

Sigala, M. (2012a). Exploiting web 2.0 for new service development: Findings and implications from the Greek tourism industry. *International Journal of Tourism Research*, 14 (6), 551-566.

Sigala, M. (2012b). Social networks and customer involvement in new service development (NSD). *International Journal of Contemporary Hospitality Management*.

Srijith, P., Hepple, M., Bontcheva, K. and Preotiuc-Pietro, D. (2017). Sub-story detection in Twitter with hierarchical Dirichlet processes. *Information Processing & Management,* 53 (4), 989-1003.

Statista (2020b). Number of social network users worldwide from 2017 to 2025. URL: *https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/* (visted on 25.10.2020).

Tuarob, S. and Tucker, C. S. (2015). Quantifying product favorability and extracting notable product features using large scale social media data. *Journal of Computing and Information Science in Engineering*, 15 (3).

Vayansky, I. and Kumar, S. A. (2020). A review of topic modeling methods. *Information Systems*, 94, 101582.

Wallach, H. M. (2006). Topic modeling: beyond bag-of-words. *Proceedings of the 23rd international conference on Machine learning*, 977-984.

Wallach, H. M., Murray, I., Salakhutdinov, R. and Mimno, D. (2009). Evaluation methods for topic models. *Proceedings of the 26th annual international conference on machine learning*, 1105-1112.

Wang, C., Wang, J., Xie, X. and Ma, W.-Y. (2007). Mining geographic knowledge using location aware topic model. *Proceedings of the 4th ACM workshop on Geographical information retrieval*, 65-70.

Wang, S., Chen, Z., Fei, G., Liu, B. and Emery, S. (2016). Targeted topic modeling for focused analysis. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 1235-1244.

Wang, X. and McCallum, A. (2006). Topics over time: a non-Markov continuous-time model of topical trends. *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 424-433.

Wang, Y., Agichtein, E. and Benzi, M. (2012). TM-LDA: efficient online modeling of latent topic transitions in social media. *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 123-131.

Womack, J. P. and Jones, D. T. (1996). Beyond Toyota: how to root out waste and pursue perfection. *Harvard business review,* 74 (5), 140-172.

Xu, S. and Xiong, Y. (2020). Setting socially mediated engagement parameters: A topic modeling and text analytic approach to examining polarized discourses on Gillette's campaign. *Public Relations Review,* 46 (5), 101959.

Yang, L., Lin, H., Lin, Y. and Liu, S. (2016). Detection and extraction of hot topics on chinese microblogs. *Cognitive Computation,* 8 (4), 577-586.

Yu, D., Xu, D., Wang, D. and Ni, Z. (2019). Hierarchical topic modeling of Twitter data for online analytical processing. *IEEE Access*, 7, 12373-12385.

Zhang, H., Kim, G. and Xing, E. P. (2015). Dynamic topic modeling for monitoring market competition from online text and image data. *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, 1425-1434.

Zhang, P., Gu, H., Gartrell, M., Lu, T., Yang, D., Ding, X. and Gu, N. (2016). Group-based Latent Dirichlet Allocation (Group-LDA): Effective audience detection for books in online social media. *Knowledge-Based Systems,* 105, 134-146.

Zhang, X., Zhao, L., Chen, Z., Boedihardjo, A. P., Dai, J. and Lu, C.-T. (2017). Trendi: Tracking stories in news and microblogs via emerging, evolving and fading topics. *2017 IEEE International Conference on Big Data*, 1590-1599.

Zhong, N. and Schweidel, D. A. (2020). Capturing changes in social media content: a multiple latent changepoint topic model. *Marketing Science,* 39 (4), 827-846.