Full Length Article

Automated content analysis and crisis communication research

Toni G.L.A. van der Meer (PhD.)

Amsterdam School of Communication Research, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands

A R T I C L E   I N F O

A B S T R A C T

Communication plays a central role in how crisis events evolve. The huge collection of today's digital available content from actors such as organizations, news media, and the public provides scholars with the opportunity to analyze large-sized collections of crisis-related communication and provide supplementary evidence for previous findings from smaller scaled research. However, the massive costs and complexity of analyzing these large-scaled data sets have hindered their use within the field of crisis research. This paper aims to provide an overview of how automated content analysis can potentially simplify and complement the analysis of these large collections of texts. Computational methods have long been used in the field of computer science and are currently gaining momentum within the field of crisis communication. This paper discusses the dictionary method, supervised method, and the unsupervised method as potential useful tools for analyzing crisis communication.

© 2016 Published by Elsevier Inc.

1. Introduction

Organizational crisis situations[1] and their societal consequences repeatedly occupy our news screens. Correspondingly, organizations report that they frequently face a crisis (Verhoeven, Tench, Zerfass, Moreno, & Verčič, 2014). This omnipresence of crisis situations and the potential negative societal effects of these critical situations have increased the scholarly attention for crisis management. Crisis management has become a key element of crisis research, mainly because multiple stakeholders, and the organization itself, will suffer when the management regarding a crisis fails (Coombs, 2007). Within crisis management, communication and the interaction with multiple involved actors is acknowledged to be a fundamental factor; when the communication is inefficient, so will be the crisis management efforts (Coombs, 2015). It can even be stated

---

E-mail address: G.L.A.vanderMeer@UvA.nl

[1] Crisis is a broad term, frequently used by both academics and practitioners to refer to a wide variety of events and issues. In general, it refers to a breakdown in a system, creating shared stress (Perry, 2007). In the context of crisis management, Coombs (2015) divides crisis in disaster and organizational crisis. Disasters refer more to disrupting events that pose great societal danger, while organizational crisis mainly refers to the threating effects of an unpredictable event on important expectations of stakeholders and the negative consequences for the organization. This study makes no explicit difference between crisis and disaster. It needs to be acknowledged that there are significant differences in communication and how social media is used between different crisis types. These differences might have implications for the applications of automated content analysis. Nevertheless, it is assumed that the different computational methods are useful for studying communication in the context of all types of crises and disasters. The methods can be used to categorize all types of communication based on different starting points or research questions related to crisis research. The final results of the analysis, and the extent to which they are theoretically interesting, depend on the cases that are studied, however, the practical application of the methods remains comparable.

that, although crises have real origins, they are constituted in the communicative interplay between several actors, whose perceptions produce real consequences (Kleinnijenhuis, Schultz, & Oegema, 2015).

Due to the far-reaching consequences of communicative efforts during a crisis, a significant body of research and numerous cases about crisis communication exist today. Scholars and practitioners aim to understand the flow of communication in times of crisis and how the communicative interplay can affect outcomes such as public panic (e.g., Liu & Kim, 2011; Van der Meer & Verhoeven, 2013), crisis escalation (e.g., Seeger, 2002), post-crisis organizational reputation (e.g., Coombs & Holladay, 2008), or financial markets (Kleinnijenhuis, Schultz, Utz, & Oegema, 2013).

The increasing body of crisis research in the field of communication has adopted multiple methodological approaches to unravel the dynamics of crisis communication. So far, crisis literature is dominated by studies applying experimental designs to understand public responses towards organizational crisis communication (Coombs, 2007; Kim & Cameron, 2011). Additionally, case studies are still the majority of the extant crisis research. For example, multiple scholars have analyzed, under different conditions, the effectiveness of the crisis-response strategies for specific cases as a way to minimize or avoid post-crisis damage. Based on Benoit's (1997) speculative image restoration strategies, Coombs (2007) categorized several response strategies as denial, diminish, and rebuild. Extensive empirical research has demonstrated how these strategies, for various crisis situations, differently affect several outcome variables such as the organization's post-crisis reputation (e.g., Coombs & Holladay, 2008) and secondary crisis communication (Schultz, Utz, & Göritz, 2011).

The digital age has brought substantial changes to the field of crisis research. In general, crisis situations set in motion a large amount of messages from various actors (Thelwall & Stuart, 2007). The shift towards online publication and archiving of different news outlets – e.g., online news websites and online archiving of newspapers – and organizations – e.g., online press releases and organizational statements on corporate websites – provides crisis researchers with new opportunities to study large amounts of crisis communication data. Furthermore, social media has become an integral part of crisis situations (Madden, Jansoke, & Briones, 2016; Ott & Theunissen, 2015; Van der Meer & Verhoeven, 2013), increasing the accessibility of public crisis communication via online platforms such as blogs, Facebook, or Twitter. Analysing the huge collection of content and understanding the complex dynamics of this contemporary media landscape in the context of organizational crisis situations requires a larger scale of analysis. Therefore, an emerging research avenue in the field of crisis communication applies forms of automated content analysis to study the communicative processes and effects using large amounts of crisis data. As scholars have recognized that much of the crisis is constructed and formed within the discourse of communication among different domains or actors (Kleinnijenhuis et al., 2015; Van der Meer, Verhoeven, Beentjes, & Vliegenthart, 2014), the use of automated content analysis provides opportunities to enrich the body of crisis literature using large data sets. In other words, this automated approach can help to provide supplementary evidence for what crisis scholars so far have suspected based on qualitative or small-scale quantitative research.

Academics in crisis research, just like academics from other social sciences, have just started to recognize the opportunities of (newly) available automated content analysis. The general aim of these computational methods, which commonly find their origin in computer science, is to automatically identify or classify certain patterns within large amounts of texts with reduced costs and time (Flaounas et al., 2013). With the use of computer-assisted methods, this classification becomes more replicable and is (likely) to be without bias due to subjective interferes of the researcher (Riff, Lacy, & Fico, 2014).

This paper aims to map the available and applicable automated content approaches for the field of crisis research. An overview of such techniques is provided to gain practical understanding of what computational methods can be used for within crisis research, guided by the overview paper of Grimmer and Stewart (2013) in the context of political communication. Both deductive and inductive computational approaches will be discussed. Deductive approaches are mainly used to analyze content based on a priori defined categories or taxonomies while inductive approaches can be applied to explore (new) patterns in text. Therefore, these different methods can serve different purposes, for example confirmatory analysis of expectations regarding content or consequences of communication based on existing theory and smaller scale analyses or more exploratory objectives aiming to further build theory based on population samples. Additionally, for each method an example will be provided of a study that applied this approach to gain insights in how these techniques can be used to answer questions related to crisis research.

2. Principles of automated content analysis

Before discussing the potential useful automated content methods, some principles need to be addressed (Grimmer & Stewart, 2013). First, automated content analyses are, of course, not free from drawbacks. Automated methods are not equivalent to manual methods. The computer-aided part makes these methods more systematically reliable and therefore more replicable, however, it cannot replace human augment. Due to the complexity of language, automated content analysis might only amplify careful reading of text. Most automated content analysis rely on the *bag of words* approach where word frequencies are used as features of text and word order does not inform the analysis (e.g., Hellsten, Dawson, & Leydesdorff, 2010; Miller, 1997). All these automated methods might fail to provide an accurate account to actually process texts. Therefore, it is argued that automated content analysis should be solely used to help researchers to content analyze large amounts of text where careful thought, reading, and interpreting the output is still essential and should be guided by the researcher.

Within the analysis of text, a wide range of different research questions and designs should lead to the use of different methodological and statistical approaches. Therefore, there is no guarantee that certain methods will be applicable to each study design. In some cases, the output of such methods may be simply wrong or misleading. Even if a specific automated