

Chapter 37

Discovering News Frames: An Approach for Exploring Text, Content, and Concepts in Online News Sources

Loretta H. Cheeks

Arizona State University, Tempe, USA

Tracy L. Stepien

Arizona State University, Tempe, USA

Dara M. Wald

Iowa State University, Ames, USA

Ashraf Gaffar

Arizona State University, Mesa, USA

ABSTRACT

The Internet is a major source of online news content. Current efforts to evaluate online news content including text, storyline, and sources is limited by the use of small-scale manual techniques that are time consuming and dependent on human judgments. This article explores the use of machine learning algorithms and mathematical techniques for Internet-scale data mining and semantic discovery of news content that will enable researchers to mine, analyze, and visualize large-scale datasets. This research has the potential to inform the integration and application of data mining to address real-world socio-environmental issues, including water insecurity in the Southwestern United States. This paper establishes a formal definition of framing and proposes an approach for the discovery of distinct patterns that characterize prominent frames. The authors' experimental evaluation shows the proposed process is an effective approach for advancing semi-supervised machine learning and may assist in advancing tools for making sense of unstructured text.

DOI: 10.4018/978-1-7998-2460-2.ch037

INTRODUCTION

Approximately 80% of all data today exists in digital form as unstructured text (e.g., news, e-mails, social media feeds, contracts, memos, clinical notes, and legal briefs) (Raghavan et al., 2004). The Internet is the premier digital platform for online news content and unstructured text (Chung, 2008). The ability to unlock hidden structure and latent meanings in unstructured text is an important area of research because text is a fundamental device of communication and human interaction for expressing real-world issues.

“The formation and transmittal of group standards, values, attitudes, and skills are accomplished largely by means of verbal communication” (Cartwright, 1953). As a consequence, efforts to understand social interaction, cooperation and influence require the study of text. Given the widespread use of unstructured text for individual and mass communication, such as email, online news and social media feeds, it is particularly important to understand how social influence and information about complex socio-environmental issues spread through online content.

The use of strategic devices for presenting salient aspects and perspectives about an issue while using certain keywords, as well as stereotyped images and sentences, for the purpose of conveying latent meanings about an issue, it is called framing (Entman, 1993). The framing of news stories can shape public interpretation of social and environmental news (Druckman & Bolsen, 2002; Sheuefele & Tewksbury, 2007; Tewksbury & Scheuefele, 2009). Public opinion, attitudes, beliefs, and behaviors can be influenced by how an issue is framed, particularly when framing comes from elites (Druckman & Bolsen, 2002).

The causal effect of media communication, specifically the influence of the words and frames the media uses to influence public perceptions of social and environmental issues, has been studied extensively. However, it traditionally has been examined using content analysis, which was developed specifically to aid in the interpretation of social discourse or text for communication research (Holsti, 1969; Krippendorff, 1980). Content analysis involves methodical evaluation and categorization of text (Riffe et al., 2005). Current approaches to content analysis, however, require scholars and researchers to thoroughly examine documents in search of patterns in the text. This approach to text analysis is dependent on humans and limits the applicability of the analysis of large-scale unstructured text. Thus, a more effective tool for unlocking latent meanings found in unstructured text could enhance our understanding of online behaviors, responses to online advertising, and media influence on public perceptions.

Text mining, also known as text data mining (Hearst, 1997), is a multidisciplinary field involving information retrieval, text analysis, information extraction, clustering, categorization, visualization, database technology, and machine learning. Text mining, coupled with an interest in understanding social influence and information diffusion for document summarization (i.e., topic modeling, sentiment analysis, and opinion mining), is an active area of research (Bindela et al., 2015; Kempe et al., 2003). Text mining, when combined with machine learning algorithms, techniques and methodologies, offers an added value to data integration tasks: they highlight the similarities between heterogeneous sources and text features, which reduces uncertainty and risk exposure when performing the integration tasks.

The widespread availability of large data repositories, like those found in online news articles, creates an opportunity to develop new methods of text mining. These methods use machine learning algorithms and mathematical techniques to select, organize, and evaluate large quantities of unstructured text. While research on the use of frames and public attitudes in traditional news venues has been widely explored, the identification and analysis of frames in unstructured online news has received minimal attention.

Therefore, the purpose of this research is to test a new approach for discovering distinct frame patterns using a process that advances semi-supervised machine learning for document clustering and clas-

18 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/discovering-news-frames/252053

Related Content

A Taxonomy of Data Mining Problems

Nayem Rahman (2020). *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications* (pp. 512-528).

www.irma-international.org/chapter/a-taxonomy-of-data-mining-problems/252041

R4 Model for Case-Based Reasoning and Its Application for Software Fault Prediction

Ekbal Rashid (2020). *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications* (pp. 1118-1140).

www.irma-international.org/chapter/r4-model-for-case-based-reasoning-and-its-application-for-software-fault-prediction/252074

A Dynamic and Scalable Decision Tree Based Mining of Educational Data

Dineshkumar B. Vaghela, Priyanka Sharma and Kalpdram Passi (2020). *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications* (pp. 841-866).

www.irma-international.org/chapter/a-dynamic-and-scalable-decision-tree-based-mining-of-educational-data/252060

Jung's Collective Unconscious, Integrative (Mind-Body-Spirit) Yoga, and Self-Realization

Manoj Sharma (2019). *Media Models to Foster Collective Human Coherence in the PSYCHecology* (pp. 93-108).

www.irma-international.org/chapter/jungs-collective-unconscious-integrative-mind-body-spirit-yoga-and-self-realization/229330

Evaluating the Impact of ACTA

(2025). *Qualitative Insights Through Applied Cognitive Task Analysis* (pp. 337-352).

www.irma-international.org/chapter/evaluating-the-impact-of-acta/370594