A Framework for Understanding the Impact of Perspective on Classification of Fake News

Andy Novobilski, Ph.D.
(678) 717-3698
andy.novobilski@ung.edu

Introduction

A question has long given pause to those trying to plan future activities. The question, "What is truth?" has been answered by a plethora of philosophers (DIffen, 2017), social planners (Perote-Pena & Piggins, 2009), and marketing execs (Mascarenhas, 2007) over the years in a manner best suited to the context the question was asked in. With current interest focused on the impact of Fake News as a social engineering tool to affect change of major events (Hadnagy, 2011), a parallel need has arisen to provide measurements that reflect both quantitative and qualitative aspects of the data content comprising a news story (Bureau, 1973).

While many classification techniques provide a repeatable process for labeling a news story as "Fake" or "Not-fake", they do so from a perceptual lens that is located within a context that is specified relative to a set of social beliefs (Dastani & Van Der Torre, 2002). This renders the concept of Fake or Not-fake irrelevant from a factual basis and requires the classification be given relative to the social context the classification is being made from. While this does not negate the value of the classification, it does require additional information to convert the data contained in the news item to actionable information that can be used by the reader. In this sense, the question of whether a news story is fake or not is no longer as important as quantifying the value of the story to provide actionable information to the reader from her perspective.

The overall project introduces a conceptual framework for converting a news story to actionable information by introducing a spectrum based classification scale that gives the story a value that varies from 0.0 (totally "Not-fake") to 1.0 (totally "Fake") using a combination of two Naive Bayesian Networks, one focused on verifiable facts held within the story and the second focused on items contained within the story which are presented as fact but are one or more pieces of data combined via the editorial perspective of the source providing the story. In addition to the classification scale, the framework includes a belief function (Community, n.d.) that provides an independent indication of the validity of the classification scale.

This paper explores the potential to use the classification scale to enable a reader holding an individual perspective to use news stories from sources with perspectives that might, or might not be the same as the reader. Specifically, a simulation was conducted in which two readers, one whose perspective aligns with CNN and a second whose perspective aligns with Fox News, where shown news stories provided by CNN and Fox News that covered the same actual politically-oriented event in time. The classification scale value, and the accompanying belief value, were generated on a per reader basis for each story with the belief value used as a measure of usefulness in converting the data within the news story to actionable information. Data for the simulation, gathered from the GDELT Project (GDELT, n.d.), was organized into three sets of data - information regarding the actual event, CNN's news coverage of the event, and Fox News news coverage of the event.

Based on the outcome of the initial simulation, near term further study is indicated in two areas. First, would the substitution of editorial content for news content increase the sensitivity of the belief value to the classification scale value for a particular editorial when read from a particular perspective. Second, would the results hold in areas of social interest outside of politics.

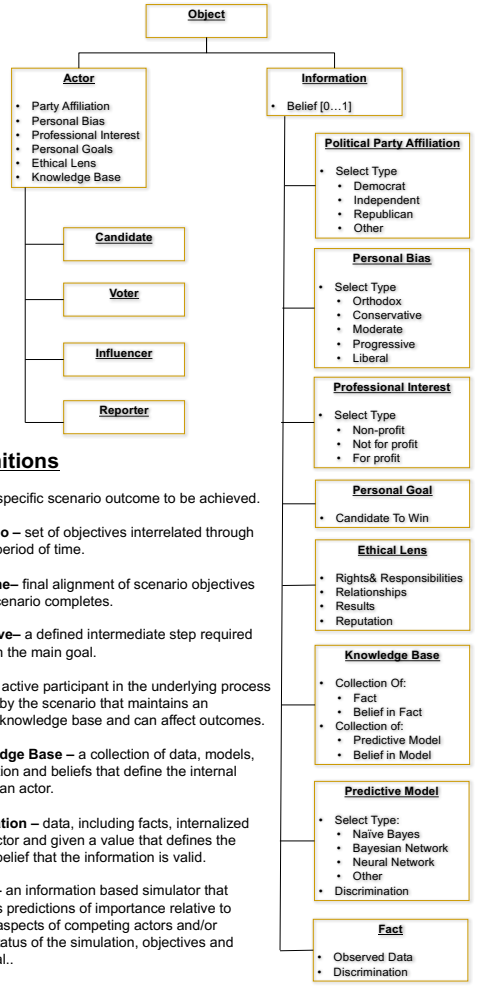
Works Cited

- Bureau, A. &. (1973). Quantitative and Qualitative Aspects of Daily Newspaper Reading. Magazine Media 360.
- Community. (n.d.). Dempster-Shafer Theory. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Dempster–Shafer_theory
- Dastani, M., & Van Der Torre, L. (2002). Specifying the MErging of Desires into Goals in the Context of Beliefs. EurAsia-ICT 2002: Information and Communication Technology. Lecture Notes in Computer Science. 2510, pp. 824-831. Berlin: Springer.
- DIffen. (2017, 06 19). Aristotle vs. Plato. Retrieved from Diffen: http://www.diffen.com/difference/Aristotle_vs_Plato
- GDELT. (n.d.). The GDELT Project. Retrieved from The GDELT Project: http://www.gdeltproject.org
- Hadnagy, C. (2011). Social Engineering: The Art of Human Hacking. Indianapolis, IN: Wiley Publishing, Inc.
- Mascarenhas, O. A. (2007). Responsible Marketing: Concepts, Theories, Models, Strategies and Cases. North Richland Hills, TX: Roval Publishing Co.
- Perote-Pena, J., & Piggins, A. (2009). Social Choice, Fuzzy Preferences and Manipulations. In R. G. Thomas A. Boylan, Economics, Rational Choice and Normative Philosphy (pp. 95-110). New York, NY: Routledge.

Notes

- Ethical Lens (n.d). Based on work pioneered by Catharyn A. Baird, J.D., the CEO and Founder of EthicsGame® and Professor of Business, Emerita, Regis University, Denver, Colorado. https://www.ethicsgame.com/exec/site/about_us.html

Framework Class Design



Definitions

Goal – specific scenario outcome to be achieved.

Scenario – set of objectives interrelated through a fixed period of time.

Outcome– final alignment of scenario objectives when scenario completes.

Objective– a defined intermediate step required to obtain the main goal.

Actor – active participant in the underlying process defined by the scenario that maintains an internal knowledge base and can affect outcomes.

Knowledge Base – a collection of data, models, information and beliefs that define the internal state of an actor.

Information – data, including facts, internalized by an actor and given a value that defines the actor's belief that the information is valid.

Model – an information based simulator that provides predictions of importance relative to hidden aspects of competing actors and/or future status of the simulation, objectives and final goal..

Data – raw input available from sources, possibly unverified, available to actors.

Belief – weight given to the accuracy of information based on the actor's internal Knowledge Base.

Fact – process specific data that is True.

Scenario from the 2016 USA Presidential Campaign

Goals – There are three competing goals in this scenario:
- Democratic Hilary Clinton elected as President
- Republican Donald Trump elected as President
- Other party candidate elected.

Scenario – using the first national presidential debate on September 26th, 2016, each candidate will strive to show the voting public they are the absolute best choice to lead the nation forward and their opponent is the absolute worse.

Outcome– each candidate, representing themselves and their party, seeks to be seen as achieving a decisive victory over their opponent that is beyond doubt.

Objective– to answer each question in a manner consistent in achieving the stated outcome with the minimum amount of potential issues arising from a misinformed answer or rebuttal. A secondary objective is to provide (and provoke from the opposing candidate) sound bites that will advance the candidates progress towards achieving the end goal.

Actor – The actors are Candidate Hilary Clinton, Candidate Donald Trump, Moderator Lester Holt, CNN reporter, Fox News reporter, Republican Voter, Democratic Voter, Independent Voter, and non-partisan influencer fact-checking.

Information – actionable items derived from data and filtered by the actors core makeup, starting Knowledge Base and set of predictive models in use by the actor. Information includes analyzed statements by other actors in the scenario. Information related to the scenario continues to be processed past the end time of the actual event.

Data – transcription of debate to include moderator and both candidates.

Fact – statements taken from the debate transcript that were verifiable by first person awareness of the truth of the statements made.

Future Study

Moving forward, the next steps are to:

1. Fully implement the framework using discrete values and simulate the outcome of the scenario from both candidates perspectives, and correlate the results to the actual outcome as it relates to the goal.
2. Model the second and third presidential debates relative to both the influence of previous debates and the debates ability to affect the actual goal.
3. Repeat steps 2 and 3 using continuous distributions where possible.
4. Configure framework to represent a generic scenario of one philosophical orientation against a second, such that the framework correctly predicts Mr. Trumps victory in the 2016 Presidential Election.
5. Apply framework to mid-term elections held in the Georgia's 6th congressional district by substituting Karen Handel for Donald Trump and John Ossoff for Hilary Clinton.

UNG | UNIVERSITY of NORTH GEORGIA™