Winter January 5, 2015

Towards News Verification: Deception Detection Methods for News Discourse

Victoria L. Rubin
Niall J. Conroy
Yimin Chen

Towards News Verification: Deception Detection Methods for News Discourse

Victoria L. Rubin, Niall J. Conroy, and Yimin Chen
Language and Information Technology Research Lab (LIT.RL)
Faculty of Information and Media Studies
University of Western Ontario, London, Ontario, CANADA
vrubin@uwo.ca, nconroy1@uwo.ca, ychen582@uwo.ca

Abstract

News verification is a process of determining whether a particular news report is truthful or deceptive. Deliberately deceptive (fabricated) news creates false conclusions in the readers' minds. Truthful (authentic) news matches the writer's knowledge. How do you tell the difference between the two in an automated way? To investigate this question, we analyzed rhetorical structures, discourse constituent parts and their coherence relations in deceptive and truthful news sample from NPR's "Bluff the Listener". Subsequently, we applied a vector space model to cluster the news by discourse feature similarity, achieving 63% accuracy. Our predictive model is not significantly better than chance (56% accuracy), though comparable to average human lie detection abilities (54%). Methodological limitations and future improvements are discussed. The long-term goal is to uncover systematic language differences and inform the core methodology of the news verification system.

1. Introduction

Mistaking fake news for authentic reports can have costly consequences, as being misled or misinformed negatively impacts our decision-making and its consequent outcomes. Fake, fabricated, falsified, disingenuous, or misleading news reports constitute instances of digital deception or deliberate misinformation. "Digital deception", a term signifying deception in the context of information and communication technology, is defined here as an intentional control of information in a technologically mediated environment to create a false belief or false conclusion [1]. Few news verification mechanisms currently exist in the context of online news, disseminated via either institutional or non-institutional channels, or provided by news aggregators or news archives. The sheer volume of the information requires novel automated approaches. Automatic analytical methods can complement and enhance the notoriously poor human ability to discern truth from deception.

A substantial body of the automated deception detection literature seeks to compile, test, and cluster predictive cues for deceptive messages but discourse and pragmatics (the use of language to accomplish communication) has rarely been considered thus far.

The online news context has received surprisingly little attention in deception detection compared to other digital contexts such as deceptive interpersonal e-mail, fake social network profiles, dating profiles, product reviews or fudged online resumes. It is, however, important to automatically identify and flag fake, fabricated, phony press releases, and hoaxes. Such automated news verification systems offer a promise of minimizing deliberate misinformation in the news flow. Here we take a first step towards such news verification system.

1.1. Research Objectives

This research aims to enable the identification of deliberately deceptive information in text-based online news. Our immediate target is the ability to make predictions about each previously unseen news piece: is it likely to belong to the truthful or deceptive category? A news verification system based on the methodology can alert users to potentially deceptive news in the incoming news stream and prompt users to further fact-check suspicious instances. It is an information system support for critical news analysis in everyday or professional information-seeking and use.

1.2. Problem Statement Elaboration

1.2.1. News Context. Daily news constitutes an important source of information for our everyday and professional lives. News can affect our personal decisions on matters such as investments, health, online purchasing, legal matters, travel or recreation. Professionals analysts (for instance, in finances, stock market, business, or government intelligence) sift through vast amounts of news to discover facts, reveal patterns, and make future forecasts. Digital news – electronically delivered online articles – is easily

accessible nowadays either via news source websites, or by keyword searching in search engines, or via news feed aggregation sites and services that pull together users' subscription feeds and deliver them to personal computers or mobile devices (e.g., drudgereport.com, newsblur.com, huffingtonpost.com, bloglines.com). Online news sources, however, range in credibility – from well-established institutional mainstream media websites (e.g., npr.org, bbc.com, cbc.ca) to the non-institutional websites of amateur reporters or citizen journalists (e.g., the CNN's iReport.com, thirdreport.com, allvoices.com, and other social media channels and their archives).

1.2.2. Citizen Journalism Context. The mis-information problem [2] is exacerbated in the current environment of user-contributed news. "An increasing number of media distributors relies on contributions from amateur reporters producing authentic materials on the spot, e.g., in cases of natural disasters or political disturbances. With mobile devices it is easy to forge media on the spot of capturing and publishing them. Thus, it is increasingly harder to determine the originality and quality of delivered media, especially under the constant pressure to be first on the news market" [3]. Citizen journalists are not obliged to follow the guidelines of source-checking and fact-checking cultivated in professional journalism, now dubbed as "News 1.0" or "the discipline of strict verification". Non-institutional news media, including "citizen journalism" [4] or "News 2.0", allow unverified posts to pass for bona-fide reporting. In many cases, the news produced by citizen journalists is reliable and verified, but there have been cases in which news has been intentionally faked, both within institutional and amateur reporting. The speed and ease by which information can be created and disseminated, coupled with new mechanisms for news production and consumption, require new verification tools applicable on a large scale.

1.2.3. Examples of Fabricated News. In October 2008, three years prior to Steve Jobs' death, a citizen journalist posted a report falsely stating that Jobs had suffered a heart attack and had been rushed to a hospital. The original deliberate misinformation was quickly "re-tweeted" disregarding the fact that it originated from the CNN's iReport.com which allows "unedited, unfiltered" posts. Although the erroneous information was later corrected, the "news" of Jobs' alleged health crisis spread fast, causing confusion and uncertainty, and resulting in a rapid fluctuation of his company's stock on that day (CBC Radio [5]). This is just one example of deceptive information being mistaken for authentic report, and it demonstrates the very significant negative consequences such errors can create. More recently, the 2013 Boston Marathon terrorist attack "evoked an outpouring of citizen journalism" with charity scams and false rumors about who the killers were [6]. Other examples of companies "struck by phony press releases" include the fiber optic manufacturer, Emulex, and Aastrom Biosciences [7].

1.2.4. Motivations to Deceive and Misinform. Why would anyone bother falsifying information in the news? Several driving forces are apparent: a) to maximize one's gains, reputation, or expertise; or b) to minimize the reputation of others (people or organizations) by decreasing their ratings or trustworthiness. One of the more legitimate reasons is c) to set up copyright traps for detecting plagiarism or copyright infringement. For instance, the ANP in the Netherlands once deliberately included a false story about a fire in their radio newscast to verify if Radio Veronica really had stolen its news from the ANP. Several hours later, Radio Veronica also aired the story [8]. Reputable sources may declare their intentions to fabricate news, but the news may still be misconstrued as genuine. The Chicago youth magazine, Muse, for instance, regularly includes a two-page spread of science and technology news, with one false story for readers to guess [8]. Such deliberately fake news is not immediately identifiable, especially when taken out of context (in digital archives or aggregator sites).

2. Literature Review

2.1. Human Abilities to Discern Lies

What is known about human abilities to spot deception? Interpersonal Psychology and Communication studies have shown that people are generally not that successful in distinguishing lies even when they are alerted to the possibility [9], [10], [11]. On average, when scored for accuracy of the lie-truth discrimination task people succeed only about half of the time [12]. A meta-analytical review of over 100 experiments with over 1,000 participants, [13] determined an unimpressive mean accuracy rate of 54%, slightly above chance [14].

Nonetheless, recent studies that examine communicative behaviors suggest that deceivers communicate in qualitatively different ways from truth-tellers. In other words, the current theory suggests that there may be stable differences in behaviors of liars versus truth-tellers, and that the differences should be especially evident in the verbal aspects of behavior [15]. Liars can perhaps be identified by their words –

not by what they say but by how they say it [16]. There is a substantial body of research that seeks to compile, test, and cluster predictive cues for deceptive messages. However, there is no general agreement on an overall reliable invariant set of predictors that replicate with statistical significance across situations [15], genres of communication, communicators and cultures [17].

2.2. Automated Deception Detection

Automated deception detection as a field within Natural Language Processing and Information Science develops methods to separate truth from deception in textual data by identifying verbal predictors of deception with text processing and machine learning techniques. The task of automated deception detection is methodologically challenging [13] and has only been recently proven feasible [18], [19], [20], [21], [22].

Previously suggested techniques for detecting deception in text reach modest accuracy rates at the level of lexico-semantic analysis. Certain lexical items are considered to be predictive linguistic cues, and could be derived, for example, from the Statement Validity Analysis (as in [23]). Though there is no clear consensus on reliable predictors of deception, deceptive cues are identified in texts, extracted and clustered conceptually, for instance, to represent diversity, complexity, specificity, and non-immediacy of the analyzed texts (e.g., [22]). When implemented with standard classification algorithms (such as neural nets, decision trees, and logistic regression), such methods achieve 74% accuracy [19]. Existing psycholinguistic lexicons (e.g., LWIC by [24]) have been adapted to perform binary text classifications for truthful versus deceptive opinions, with an average classifier demonstrating 70% accuracy rate [25]. These modest results, though usually achieved on restricted topics, are promising since they surpass notoriously unreliable human abilities in lie-detection.

What most studies have in common is the focus on lexics and semantics (the use of words and their meaning), and some syntax (the use of phrasal and sentence structures). Discourse and pragmatics (the use of language to accomplish communication) has rarely been considered thus far [26], [27],[28].

2.3. Deception Detection for News Verification

In spite of the enormous difficulty of the automated detection task, several digital contexts have been examined: fake product reviews [29 & Glance, 2013], opinion spamming [30], deceptive interpersonal e-mail [31], fake social network profiles [32], fake dating profiles [33], and fudged online resumes [34]. There has been, however, surprisingly little, if any, well-known effort in this field to analyze digital news and automatically identify and flag phony press releases, hoaxes, or other varieties of digital deception in news environments. Academic scholarship in journalism is an appropriate source for an interdisciplinary exploration and preliminary suggestions for automation. For instance, an analysis of ten major cases of fabricated news in American mainstream media [35] suggests that news editors watch out for recognizable patterns to prevent journalistic deception: "Deceptive news stories are more likely than authentic news stories to be filed from a remote location, to be on a story topic conducive to source secrecy, to be on the front page (or magazine cover), to contain more sources, more ''diverse'' sources and more hard-to-trace sources" (p. 159). This study [36] found deceptive news "portrayed a simpler world" (p. 1).

Like other artifacts of deliberate, disruptive, or malevolent acts (such as fraud or spam), instances of digital deception are not as readily available or accessible for comparative analyses with authentic news. Scarce data availability requires a careful corpus construction methodology for a reliable "gold-standard", so that positive and negative instances of digital deception in the news context can be systematically compared and modeled. News reports exhibit fewer certainty markers (softened, solidified, or hedged statements, e.g., "perhaps" ,"I believe", "surely") compared to editorials [37], [21], [38], [39] but it is unknown whether deceptive news exhibit more or less certainty as well as factuality [40, 41] as compared to authentic news and editorials. News is to some extent predictable in its discourse structure (e.g., headline, lead, main events etc., per [42, 43]) but it is less regulated than some of the other previously scrutinized discourse types (such as reviews or resumes). Fabrication requires heightened creativity and extra rhetorical persuasion in achieving believability.

Since news verification is an overall discourse level decision – is the news fabricated or not? – it is reasonable to consider discourse / pragmatic features of each news piece.

3. Theoretical Approaches

Rhetorical Structure Theory (RST) and Vector Space Modeling (VSM) are the two theoretical components we use in our analysis of deceptive and truthful news. The RST-VSM methodology has been previously applied to free-form computer-mediated

communication (CMC) of personal stories [28], [27]. In this work we test the applicability of the RST-VSM to the news discourse, given news structural peculiarities and differences from CMC. RST is used to analyze news discourse and VSM is used to interpret discourse features into an abstract mathematical space. Each component is discussed in turn per [28], [27].

3.1. Rhetorical Structure Theory (RST)

RST analysis captures the coherence of a story in terms of functional relations among different meaningful text units, and describes a hierarchical structure for each story [44]. The result is that each analyzed text is converted to a set of rhetorical relations connected in a hierarchical manner with more salient text units heading this hierarchical tree. The Theory differentiates between rhetorically stand-alone parts of a text, some of which are more salient (nucleus) than the others (satellite). In the past couple of decades, empirical observations and previous empirical research confirmed that writers tend to emphasize certain parts of a text in order to express their most essential idea. These parts can be systematically identified through the analysis of the rhetorical connections among more and less essential parts of a text. RST relations (e.g., purpose, elaboration, non-volitional result) describe how connected text segments cohere within a hierarchical tree structure, which is an RST quantified representation of a coded text [27], [28].

3.2. Vector Space Modeling (VSM)

We use a vector space model for the identification of these sets of rhetorical structure relations. Mathematically speaking, news must be modeled in a way suitable for the application of various computational algorithms based on linear algebra. Using a vector space model, each news text can be represented as vectors in a high dimensional space [45], [46]. Then, each dimension of the vector space is equal to the number of rhetorical relations in a set of all news reports under consideration. Such representation of news text makes the vector space model very attractive in terms of its simplicity and applicability [47], [28], [27].

The news reports are represented as vectors in an n-dimensional space. The main subsets of the news space are two clusters, deceptive news and truthful news. The element of a cluster is a news story, and a cluster is a set of elements that share enough similarity to be grouped together, as deceptive news or truthful news [48]. That is, each news can be described by a number of distinctive features (rhetorical relations, their co-occurrences and positions in a hierarchical structure); together, these features make each news story unique and identify the story as a member of a particular cluster, per [28], [27]. In our analysis, the distinctive features of the news are compared, and when a similarity threshold is met, they are placed in one of two groups, deceptive or truthful.

Similarity cluster analysis is based on distances between the samples in the original vector space [49]. Modifying the similarity-based clustering framework [50] and adapting RST-VSM methodology [28], [27] to the news context, we test how well RST-VSM can be applied to news verification.

4. Methods

4. 1. Research Question

We hypothesized that if the relations between discourse constituent parts in deceptive (fabricated) news reports differed from the ones in truthful (authentic) reports, then a systematic analysis of such relations could help to detect and filter deceptive news, in essence verifying the veracity of the news.

Our investigation was guided by the overall research questions: How do the rhetorical relations among discourse constituent parts differ between truthful and deceptive? The question was investigated in the NPR "Bluff the Listener" news report data through three sub-questions:

A. Are there significant differences in the frequency of assignments of the RST relations to the news that belong to the truthful group, as opposed to those in the deceptive group?
B. Can news reports be clustered based on the RST relation assignments per RST-VSM methodology (per [28], [27])? If so, how accurately?
C. Is there a subset of RST relations that can be used as a predictor of truth or deception of the news; and if so, how accurately?

4.2. Data Collection and Data Source

Obtaining reliable positive and negative data samples is one of the challenges in automated deception detection research and requires careful selection of training and test data. The difficulty is in ground truth verification: finding suitable data "in the wild" and conducting the fact checks to obtain ground truth is costly, time-consuming, and labor intensive [26], [51].

We used a source that clearly marked fake news and the ground truth was established a priori. Starting with professional journalists faking news appeared

reasonable since fake narratives are difficult to write well, except by highly skilled experts [52].

The US National Public Radio (NPR) website contains transcripts of a weekly radio show "Wait, Wait, Don't Tell Me" with its "Bluff the Listener" segment, dating back to the spring of 2010. (Mass media portrayal of lie-detection has been previously considered by the deception detection community. For instance, a recent study [53] found that the "Lie to Me" TV show increased its viewers' suspicion (by reducing their truth bias) and, in fact, negatively impacted their deception detection ability while instilling a false sense of confidence in their abilities.) "Bluff the Listener" does not claim or attempt to educate their listeners in the skills of lie-detection. It is a simple test of intuition and perhaps a guessing game.

We collected all "Bluff the Listener" show transcripts available from March 2010 to May 2014 (with NPR's explicit permissions). Each "Bluff the Listener" show contains three thematically-linked news reports (triplets), one of which is authentic (truthful) and the other two are fake (deceptive). The news triplets are written to be read aloud to the listeners who call to participate in the truth-identification game, but the format of the transcripts is in the radio announcement style, which is similar to written news. Most news reports are typically humorous and some are highly unlikely or unbelievable (e.g., a ship captain plotting his ship's course across land or a swim instructor not knowing how to swim). The news triples are intended to bluff the listeners by persuading them to misidentify one of the two fake news as truthful, for entertainment value.

Methodologically speaking, we were interested in testing the applicability of the RST-VSM methodology in the news context, as well as the suitability of the specific show as the data for deception detection.

4.3. Data Analysis

4.3.1. Data. Our random sample originally consisted of 144 news transcripts which yielded 168 associated RST analyses for these texts. Coder Y analyzed 60 news reports (30 randomly selected from 2011 "Bluff the Listener shows", and 30 from 2013). Coder N analyzed another 60 news sample (30 from 2010, and 30 from 2012), with 120 news reports in total between the two coders. In addition, both coders analyzed 24 news reports for intercoder consistency (one set of 12 news reports, consisting of one triplet taken from each year between 2010 and 2013; and additional set of 12 most reports from year 2014). As a result, our overall sample dataset amounts to 144 randomly selected news reports making up 168 sets of RST relation analysis (including 24 duplicated sets of analysis).

4.3.2. RST analysis procedures. RST analysis was conducted by two analysts (Coders N and Y) applying the extended relation set (ExtMT.rel) in the RST Tool version 3.45 software.

The shortest distance between two points may be a straight line, ¶ but navigators on the high seas know they have to account for wind patterns ¶ and currents ¶ and something that experts call land. ¶ Some should have offered this helpful tip to Polish freighter captain Thaddeus Dudek ¶

Figure 1. An RST segmentation sample (11/07/2013 "Bluff the Listener" truthful story)

Each news report was first segmented into RST elemental discourse units (Fig. 1).Using the *Structurer* tab of the RST Tool, relations were applied to the segments (Figure 2), starting from the main topic (top-level unit), labeling the most obvious relations first and assigning other potential candidate relations top-to-bottom, left-to- right.

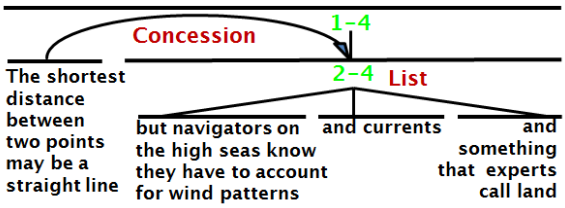

Figure 2. Sample of RST relations assignments to four discourse segments.

Each annotator re-read each news report several times to verify the logic of the analysis. On subsequent passes, more complex or ambiguous relations were assigned, while consulting the inventory of relation definitions (www.sfu.ca/rst/01intro/definitions.html) and example analyses (www.sfu.ca/rst/02analyses/index.html). At times, certain segments required modification from the original partitioning, and certain previous relation assignments were reconsidered in order to uncover the hierarchy of the coherent and logically nested discourse structure. Great care was taken to ensure that the analysis points to a single segment or span as the central news message. This is a time-consuming manual step that is necessary for now. There are several attempts to move RST analysis from manual tool-aided work to full automation [54], [55], [56], [57], [58], but none are available as of yet.

4.3.3. Coder consistency procedures. For the purpose of improving agreement between the two analysts in this manual step, several texts were segmented and RST relations were assigned collaboratively (per

procedures in 4.3.2). Coder practices were compared carefully and discussed on three different occasions (lasting 1.5-2 hours each). Segmentation practices were deeded to not be substantially different and were consequently disregarded in the inter-coder reliability tests.

The formal RST website relation descriptions and examples were used as a pseudo-codebook in the relation assignments, with an addition of one extra relation a rhetorical Question, used to mark the connection between rhetorical questions and answers, with the question as Nucleus (Figure 3). Several guiding principles of relation assignment were also adopted in an attempt to increase consistency.

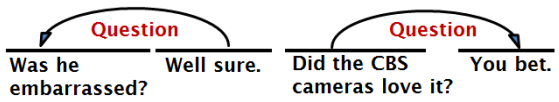

Figure 3. Examples of Questions.

4.3.4. Inter-coder reliability test methods and calculations with consequent data manipulations. Realizing that subjectivity of applying RST relations is a known critique, we conducted two inter-coder reliability tests in which we were looking to improve our RST analysis procedures for consistency and further formalize the principles for RST relation assignment, with an eye on potential automation of the steps and decisions made.

Two intercoder test sets were used, 12 news reports each, analyzed by each analyst (Coder Y and N) independently. Intercoder Test Set 1 (coded in May 2014) consisted of 12 news reports (or 4 triplets), selected one per year from 2010-2013 shows. Each triplet contained 3 news reports, 2 of which were deceptive and one truthful. Intercoder Test Set II (coded in June 2014) contained 4 triplets from 2014, randomly selected out of the 22 shows aired up to date in 2014, resulting in 12 news in total.

Each coder assigned the same 24 news reports a total of 447 RST relations (231 relations between the news constituent segments in Intercoder Set I, and 216 – in Set II). The segmentation (into elementary discourse unites per RST) was kept constant by a preliminary agreed-upon segmentation procedure with mutual verification and renegotiation of disagreed upon segments, if any. The hierarchical structures (assembling of the relation into the discourse trees) were individual coder decisions.

Coder N's and Coder Y's assignments for each of the 447 RST relation were compared pair-wise at the level of these discourse segments. For instance, in Figure 2, the span of segments 2-4 is assigned List as a relation, and the span 1 → 2-4 is Concession. If both annotators assigned List to the 2-4 span, it was counted as agreement (1). If one of the coders assigned Background instead of Concession to the 1→ 2-4 span, it was counted as disagreement (0). A confusion matrix was used to reflect counts of matching and mismatching assignments. Coders' percent agreement and Cohen's kappa [59] were calculated.

Intercoder Test I (performed in May 2014 on 12 news reports) yielded 216 relations between discourse segments in their discourse structures. With an inventory of 33 categories in the coding scheme (using the classic RST set, plus an additional Question relation appropriate for the radio show, Figure 3), a 50% inter-coder agreement was reached on assigning RST relations correctly (107 out of 216).

After an iterative error analysis and adoption of several principles for consistency on relation assignments, the test was repeated with the Intercoder Set II) which improved the agreement by 10%, yielding 139 agreed upon assignments out of 231 (60%).

The average agreement between coders Y and N in two Intercoder Tests (performed one month apart with some consistency negotiation procedures) was 55% (or 246 agreements on 447 relations among discourse segments The Cohen's kappa was 0.51, interpreted as mid-range moderate agreement (0.61–0.80) [59].

After the second attempt to reach better intercoder agreement, we noted that certain relations were consistently confused or used inconsistently by both coders. Those relations were deemed indistinguishable (at least in practice, if not in theory), given the cognitive difficulty of keeping 33 relations in mind during the analysis. Certain vagueness in the original RST relation definitions may also be at fault (e.g., in List and Sequence).

We continued to remedy the situation by constructing 3 abstract relational categories that lumped some relations that carry similar rhetorical meaning. Even though the RST theorists may object to this move, such technique is consistent with accepted practices of joining predictive cues in deception detection into more abstract concepts (e.g., [22]). Below are the three lumped categories under their generic name (preceded by a GR notation): Elaboration + Evaluation + Evidence + Interpretation = GR1_Elaboration; 22Antithesis + Background + Circumstance + Preparation = GR2_Background; and Conjunction + List + Sequence = GR3_Lists. In addition, we removed the following 7 relations that were never or extremely rarely used by the analysts:

Enablement, Justify, Multi-nuclear restatement, Otherwise, Summary, Unconditional, and Unless.

As a result of these data manipulations, the number of relations was reduced to 18 (from 33) and the RST assignments across both Sets I, II to 430 (from 447).

The resulting intercoder agreement on the lumped data (with rare data point removed) then reached 69% agreement (296 out of 430) and the achieved 0.64 Cohen's kappa statistic can now be interpreted in the lower range of substantial agreement (0.61–0.80), per [59]. The lumped dataset consisting of 132 news reports and 430 RST assignments resulted in improved intercoder reliability and was further used for clustering and predictive modeling.

4.3.5. Statistical Procedures for Predictive Modeling. To perform logistical regression, 100 randomly selected news reports (76% of 132) were used as a training set for the logistic regression, with the other 32 (24%) retained as a test bed.

R (version 3.1.1; [60]) package {bestglm} was used to select the best subset of predictor variables for a logistic regression according to Akaike information criterion (AIC). {bestglm} uses complete enumeration process (described by [61]) which tests efficiently all possible subsets of predictor variable variations (using the training dataset). The selected model equation was used to predict truth or deception for the test dataset. The chi-square test of independence was used to compare predictions for the test data to chance results.

5. Results

5.1. Modeling Deceptive / Truthful Centers

An RST-VSM process of clustering deceptive versus truthful texts was performed using the dataset of 132 news reports, made up of an equal amount of deceptive and non-deceptive texts. To reiterate, these news reports were analyzed in terms of RST structure (with a set of 18 RST relations) and examined around whether this structure related to deceptive value. A VSM was used to asses each news report's position in a multi-dimensional RST space. Clustering of truthful and deceptive data points in this space was evaluated based on distances to hypothetical cluster centers.

The coding process of assigning RST relations produced a statistics file for each news report which was translated to a multi-dimensional vector representing RST frequencies, normalized by its Euclidean vector length so that they may be represented in Euclidean similarity space.

Batch clustering was performed on a set of reports via the vector space description and subsequent transformation to a similarity space description. Similarity is judged to be the non-zero distance between vertices; in this case, we chose the metric of the Euclidean distance between a news report vector and cluster center.

The construction of a deceptive model used 100 news reports (chosen at random out of 132, or 76%). The remaining 32 reports (24%) were set aside for the purpose of model evaluation. We computed deceptive and truthful cluster centers by finding the normalized frequency means from each relation, from the deceptive and truthful groups respectfully.

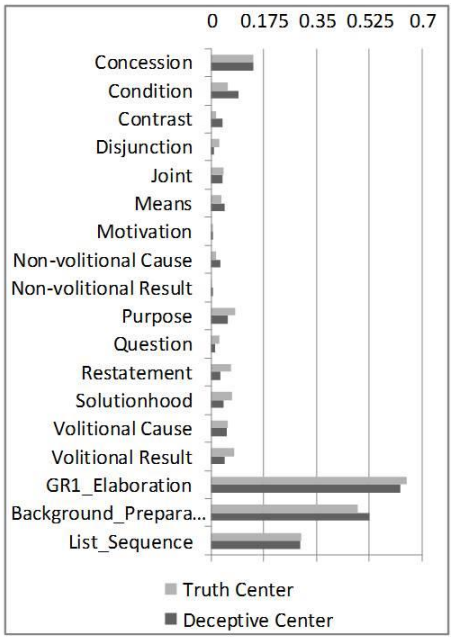

Figure 4. Truthful and deceptive centers (n=100).

Performing an independent samples t-test indicated statistical differences between truthful and deceptive centers for certain relations, pointing to the possibility that deceptive and truthful reports could be discriminated by the presence or absence of these relations. The distribution of deceptive and truth centers for each relation is provided in Figure 4, Truthful vs Deceptive Centers, for n=100 stories. Disjunction (p=0.053) and Restatement (p=0.037) relations show significant differences between truthful and deceptive stories, with these relations more likely occurring in truthful stories.

5.2. Clustering

A clustering visualization of the training 100 news reports was performed using the gCLUTO clustering package [62], [63] (see Figure 5). This procedure was done to help differentiate news reports based on their similarity according to a chosen clustering algorithm.

By experimenting with the data set and various clustering methods, 4 similarity clusters were formed using the Agglomerative clustering with *k*-nearest neighbor approach, clustering similar news reports based on the normalized frequency of relations.

The distance between a pair of peaks on the plane represents the relative similarity of their clusters. The height of each peak on the plane is proportional to the internal similarity of the cluster, calculated by the average pair-wise similarity between objects. The color of a peak represents the internal standard deviation of the cluster's objects [62].

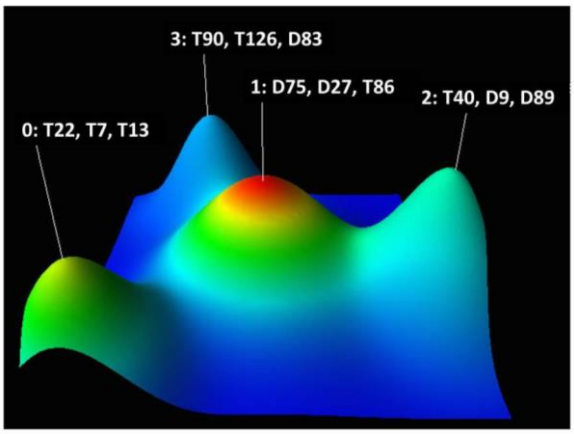

Figure 5. Clustering visualization in gCLUTO [62,63].

A clustering visualization produced clusters of size 41, 32, 20, 7 stories respectfully. Of note is the formation of certain clusters comprised entirely of truthful stories (e.g., Group 0: T22, T7, T13, Figure 5). This grouping of news reports with similar values indicates areas of further exploration to determine common characteristics including discriminating relations.

The validity of the model, that is, its ability to determine the deceptive value of a new story was measured based on the principle of co-ordinate distances. After deceptive and non-deceptive cluster centers were computed, new incoming stories were assessed of their deceptive values based on the Euclidean distances to these centers. For instance, if the co-ordinate of the story was closer to the deceptive center than the truthful center, it was deemed deceptive according to the model. Likewise, if the co-ordinate of the story was closer to the truthful center than to the deceptive center, it was deemed truthful. The outcome of comparing the actual deceptive value of a new story to its predicted deceptive value produced a success rate based on the test set of 32 news reports. The results indicate that the model was able to correctly assess 63% (20 out of 32 stories).

5.3. Predictive Modeling

The following logistic regression model was selected based on the training lumped dataset (of a 100 out of 132 news reports) (Table 1). Condition 1 is Truth and 0 is Deception: the positive coefficients increase the probability of the truth, and negative ones increase the probability of deception.

Four logistic regression indicators from a set of 18 pointed to Truth (Disjunction, Purpose, Restatement, and Solutionhood), while another predictor (Condition) pointed to Deception (Table 1).

Table 1. Coefficients of the selected logistic regression model to predict truthful or deceptive news reports.

Source RST relation	Coefficient	p-value
(Intercept)	-0.7109	0.0403
Condition	-3.6316	0.0676
Disjunction	10.6244	0.0523
Purpose	3.4383	0.1023
Restatement	6.0902	0.0219
Solutionhood	5.2755	0.0526

When tested for accuracy of the model predictions, the training set overall accuracy was 70% (Table 2). The test dataset accuracy, however, was 56%. Eighteen 18 out of 32 news reports that were predicted correctly (Table 3). This is not significantly better than chance (chi-square (1 df) = 0.0339, p = 0.854).

Table 2. Accuracy of the logistic regression model on the training set (n=100).

	Observed Deception	Observed Truth
Predicted Deception	**37**	17
Predicted Truth	13	**33**

Table 3. Accuracy of the logistic regression model on the test set (n=32).

	Observed Deception	Observed Truth
Predicted Deception	**12**	8
Predicted Truth	6	**6**

6. Discussion

While the RST-SVM clustering technique for the NPR's "Bluff the Listener" news reports was only in part successful (63% accuracy), further steps need to be taken to find predictors of deception for a news verification system. We deem it important however to report our results to the deception detection community and point out potential stumbling blocks in the data and analytical process. We now discuss the nature of our data sample and come back to the problem of subjectivity of RST assignments.

Are "Bluff the Listener" news reports entirely suitable for modeling deceptive news reports? It is possible that "Bluff the Listeners" writers' intent to deceive their listeners is mitigated by their goal to

entertain the audience. The question remains: is bluffing for entertainment similar enough to news reporting for misinformation? The elements of humor and intent to entertain may cause interferences in showing verbal differences between truths and lies.

In addition we observed that most "Bluff the Listener" news pieces were of a highly unlikely nature. They appear unbelievable or at least surprising, which makes the task of selecting the actual truthful event (out of three unlikely reports) more difficult. Does the plausibility of reported events interfere with deceptive clues? Perhaps other news venues (intended to strictly misinform) are more appropriate for predictive modeling. For instance, certain news outlets or websites openly declare their intentions to produce fake news (e.g., CBC's "This is That", Huffington Post, the onion, the Muse, etc.), have been known to misinform (e.g., Politifact.com employs investigative journalists to uncover misinformation in news) or have been caught fabricating (e.g., cases in [35], [64], [7], also see Section 1.2.4 for concrete examples).

Yet another possibility is that deception detection methods based on discourse structure nuances are not as effective for discourse types with pre-defined structures (such as news, ads, and weather reports) as compared to free form discourse types (such as personal narratives). Each of these confounding factors requires further investigation and additional analyses.

Lastly, as evidenced by our difficulties in achieving intercoder agreement, assignment of RST relations to text can be strongly affected by individual differences in coders' interpretations. Several RST relations have ambiguous or overlapping definitions, which can have a compounding effect on disagreements. This problem of subjectivity in RST was critiqued in the past, leading to several authors proposing different annotation and visualization schemes as alternatives [65], [66], [67]. However, none of them seem to have gained widespread adoption, nor do they resolve the fundamental problem of intercoder subjectivity. Rather than abandoning it in favor of as-yet unproven alternatives, we will continue improving robustness of the RST framework for potential future automation.

How might accuracy be improved? Based on the increase in coder agreement between the two reliability tests, continued coder training and consensus-building (such as through discussion of problematic cases) should help to improve consistency. It may be that the set of original RST relations is over-differentiated, forcing coders to make unnecessary distinctions between conceptually similar relations. The next step is to manually reapply the simplified (lumped) scheme with the reduced overall number of relations.

7. Conclusions

In the context of news consumption by lay people and professional analysts, it is critical to distinguish truthful news reports from deceptive ones. With few news verification mechanisms currently available, this research lays the groundwork towards an automated deception detection approach for news verification.

We sought to provide evidence of stable discourse differences between deceptive (fabricated) and truthful (authentic) news, specifically in terms of their rhetorical structures and coherence relation patterns. To make the veracity prediction (whether the news is truthful or not), we considered it to be useful to look at how news reports are structured. We described NPR's "Buff the Listener" news reports, written by professional news writers with the intention to bluff the audience, as a promising source of data for the deception detection task for news verification.

We applied a vector space model to cluster the news by discourse feature similarity and achieved 63% accuracy on our test set. Though our predictive model is comparable to average human lie detection abilities (54% accuracy) and performed at 70% accuracy on the training set, it performed at only 56% accuracy on the test set which is not significantly better than chance (chi-square (1 df) = 0.0339, p = 0.854). Thus, our results are promising but inconclusive, specifically in terms of data suitability and the method's predictive powers. There were several confounding issues (such news discourse specificity) and methodological limitation (such as the subjectivity of the RST relation assignments) that need further research on the path towards news verification system development.

The idea behind the news verification system is for it to take in a previously unseen news report from an incoming news stream, analyze its rhetorical structure, convert it mathematically into an abstract truth-deception vector space model, and estimate its (Euclidian distance) closeness to the truth and deception centers. Then, if the news report rating falls beyond an established threshold of veracity, an alert calls users to fact-check potentially deceptive content.

Though this work is technologically and methodologically challenging, it is timely and carries potential benefits to news consumers. In principle, news verification system can improve credibility assessment of digital news sources. The mere awareness of potential deception can increase new media literacy and prevent undesirable costs of mistaking fake news for authentic reports.

8. Acknowledgements

This research was funded by the GRAND Canadian Network of Excellence Award for "News Authentication".

9. References

[1] J.T. Hancock, "Digital deception: When, where & how people lie online," *The Oxford handbook of Internet psychology*, A. N. Joinson, et al., eds., Oxford University Press, 2012, pp. 508.

[2] C.J. Fox, *Information & misinformation: An investigation of the notions of information, misinformation, informing, & misinforming*, Greenwood, 1983.

[3] K. Ahmed, et al., *Community & trust-aware fake media detection*, Springer Science & Business Media LLC 2012.

[4] D. Gillmore, "We the media: The rise of citizen journalists," *National Civic Review*, vol. 93, no. 3, 2004, pp. 58-63.

[5] CBC Radio "And the Winner Is", News 2.0, Part II, Retrieved from http://www.cbc.ca/andthewinneris/," 31/03/12

[6] A.J. Waskey, "News Media: Internet," *Encyclopedia of Deception* 2, T. R. Levine, ed., Sage Reference, 2014, p. 710.

[7] A. Mintz, *Web of Deception: Misinformation on the Internet*, CyberAge Books, 2002.

[8] "Fictitious Entry," 2013; http://en.wikipedia.org/wiki/Fictitious_entry.

[9] A. Vrij, *Detecting Lies & Deceit*, Wiley, 2000.

[10] A. Vrij, "Why professionals fail to catch liars & how they can improve," *Legal & criminological psychology*, vol. 9, no. 2, 2004, pp. 159-181.

[11] A. Vrij, et al., "Deception Traits in Psychological Interviewing," *Journal of Police & Criminal Psychology*, vol. 28, no. 2, 2012, pp. 115-126.

[12] M.G. Frank, et al., "Individual & Small Group Accuracy in Judging Truthful & Deceptive Communication," *Group Decision & Negotiation*, vol. 13, 2004, pp. 45-59.

[13] B.M. DePaulo, et al., "The Accuracy-Confidence Correlation in the Detection of Deception," *Personality & Social Psychology Review*, vol. 1, no. 4, 1997, pp. 346-357.

[14] V.L. Rubin & N. Conroy, "Discerning truth from deception: Human judgments & automation efforts," *First Monday*2012; http://firstmonday.org.

[15] M. Ali & T.R. Levine, "The language of truthful & deceptive denials & confessions," *Communication Reports*, vol. 21, 2008, pp. 82-91.

[16] M.L. Newman, et al., "Lying words: Predicting deception from linguistic styles," *Personality & Social Psychology Bulletin*, vol. 29, no. 5, 2003, pp. 665-675.

[17] J.K. Burgoon, et al., "Detecting deception through linguistic analysis," *Intelligence & Security Informatics, Proceedings*, vol. 2665, 2003, pp. 91-101.

[18] J. Bachenko, et al., "Verification & implementation of language-based deception indicators in civil & criminal narratives," *Proc. 22nd International Conference on Computational Linguistics*, ACL, 2008.

[19] C.M. Fuller, et al., "Decision support for determining veracity via linguistic-based cues " *Decision Support Systems* vol. 46, no. 3, 2009, pp. 695-703.

[20] J.T. Hancock, et al., "On lying & being lied to: A linguistic analysis of deception in computer-mediated communication," *Discourse Processes*, vol. 45, no. 1, 2008, pp. 1-23.

[21] V.L. Rubin, "On deception & deception detection: content analysis of computer-mediated stated beliefs," *Proc. 73rd ASIS&T Annual Meeting: Navigating Streams in an Information Ecosystem*, American Society for Information Science, 2010.

[22] L. Zhou, et al., "Automating Linguistics-Based Cues for Detecting Deception in Text-Based Asynchronous Computer-Mediated Communications," *Group Decision & Negotiation*, vol. 13, no. 1, 2004, pp. 81-106.

[23] S. Porter & J.C. Yuille, "The language of deceit: An investigation of the verbal clues to deception in the interrogation context," *Law & Human Behavior*, vol. 20, no. 4, 1996, pp. 443-458.

[24] J. Pennebaker & M. Francis, *Linguistic inquiry & word count: LIWC.* , Erlbaum Publishers, 1999.

[25] R. Mihalcea & C. Strapparava, "The Lie Detector: Explorations in the Automatic Recognition of Deceptive Language," *Proc. 47th Annual Meeting of the Association for Computational Linguistics, Singapore*, 2009, pp. 309-312.

[26] J. Bachenko, et al., "Verification & implementation of language-based deception indicators in civil & criminal narratives," *Proc. ACL*, ACL, 2008.

[27] V.L. Rubin & T. Lukoianova, "Truth & deception at the rhetorical structure level," *Journal of the Association for Information Science & Technology*, 2014, pp. n/a-n/a.

[28] V.L. Rubin & T. Vashchilko, "Identification of truth & deception in text: application of vector space model to rhetorical structure theory," *Proc. Workshop on Computational Approaches to Deception Detection, EACL Annual Meeting*, ACL, 2012.

[29] A. Mukherjee, et al., *Fake Review Detection: Classification & Analysis of Real & Pseudo Reviews. Technical Report.*, Department of Computer Science, University of Illinois at Chicago, & Google Inc., 2013.

[30] N. Jindal & B. Liu, "Opinion spam & analysis," *Book Opinion spam & analysis*, Series Opinion spam & analysis, ed., Editor ed.^eds., ACM, 2008, pp.

[31] P. Keila & D. Skillicorn, *Detecting unusual & deceptive communication in email. Technical Report*, School of Computing, Queen's University, Kingston, Canada, 2005.

[32] N. Kumar & R.N. Reddy, "Automatic Detection of Fake Profiles in Online Social Networks," BTech Thesis, 2012.

[33] C.L. Toma & J.T. Hancock, "What Lies Beneath: The Linguistic Traces of Deception in Online Dating Profiles," *Journal of Communication*, vol. 62, no. 1, 2012, pp. 78-97.

[34] J. Guillory & J.T. Hancock, "The Effect of Linkedin on Deception in Resumes," *Cyberpsychology, Behavior, & Social Networking*, vol. 15, no. 3, 2012, pp. 135-140.

[35] D.L. Lasorsa & J. Dai, "Newsroom's Normal Accident?," *Journalism Practice*, 1 (2), 2007, pp. 159-174.

[36] L. Dalecki, et al., "The News Readability Problem," *Journalism Practice*, vol. 3, no. 1, 2009.

[37] V.L. Rubin, "Stating with Certainty or Stating with Doubt: Intercoder Reliability Results for Manual Annotation of Epistemically Modalized Statements," *Proc. Human Language Technologies Conference*, 2007, pp. 141-144.

[38] V.L. Rubin, et al., "Certainty Categorization Model," *Proc. AAAI Symposium on Exploring Attitude & Affect in Text*, 2004.

[39] V.L. Rubin, et al., "Certainty Identification in Texts: Categorization Model & Manual Tagging Results.," *Computing Attitude & Affect in Text: Theory & Applications*, The Information Retrieval Series, J. G. Shanahan, et al., eds., Springer-Verlag, 2005, pp. 61-76.
[40] R. Sauri & J. Pustejovsky, "FactBank: a corpus annotated with event factuality," *Language Resources & Evaluation*, vol. 43, no. 3, 2009, pp. 227-268.
[41] R. Sauri & J. Pustejovsky, "Are You Sure That This Happened? Assessing the Factuality Degree of Events in Text," *Computational Linguistics*, 2012, pp. 1-39.
[42] T.A. van Dijk, *News as Discourse*, Lawrence Erlbaum Associates Inc., 1988, p. 200.
[43] T.A. van Dijk, *Studies in the Pragmatics of Discourse*, Mouton Publishers, 1981.
[44] W.C. Mann & S.A. Thompson, "Rhetorical Structure Theory: Toward a Functional Theory of Text Organization," *Text*, vol. 8, no. 3, 1988, pp. 243-281.
[45] C.D. Manning & H. Schutze, *Foundations of Statistical Natural Language Processing*, The MIT Press, 1999.
[46] G. Salton & M.J. McGill, *Introduction to Modern Information Retrieval*, McGraw-Hill, 1983.
[47] R. Baeza-Yates & B. Ribeiro-Neto, *Modern information retrieval*, Addison-Wesley, 1999.
[48] P. Berkhin, "Survey of clustering data mining techniques.," 2002; DOI: 10.1.1.18.3739.
[49] A. Strehl, et al., "Impact of similarity measures on web-page clustering.," *Book Impact of similarity measures on web-page clustering.*, Series Impact of similarity measures on web-page clustering., 2000, pp. 58-64.
[50] A. Strehl, et al., "Impact of similarity measures on web-page clustering," *Proc. AAAI-2000 Workshop of Artificial Intelligence for Web Search*, 2000, pp. 58-64.
[51] V.L. Rubin & N. Conroy, "Challenges in Automated Deception Detection in Computer-Mediated Communication," ASIST2011.
[52] N.C. Rowe, "Automatic Detection of Fake File Systems Report," 2013; <http://faculty.nps.edu/ncrowe/fakeintel.htm.
[53] T.R. Levine, et al., "The Impact of Lie to Me on Viewers Actual Ability to Detect Deception," *Communication Research*.
[54] D. Marcu, "The Rhetorical Parsing of Natural Language Texts," *Proc. ACL/EACL1997*, 1997, pp. 96-103.
[55] D. Marcu & A. Echihabi, "An Unsupervised Approach to Recognizing Discourse Relations," *Proc. ACL2002*, 2002, pp. 368-375.
[56] H. Hernault, et al., "HILDA: A discourse parser using support vector machine classification," *Dialogue & Discourse*, vol. 1, no. 3, 2010, pp. 1-33.
[57] V.W. Feng & G. Hirst, "Text-level discourse parsing with rich linguistic features," *Proc. Association for Computational Linguistics: Human Language Technologies (ACL2012)*, 2012, pp. 60-68.
[58] Z. Lin, et al., "Recognizing implicit discourse relations in the Penn Discourse Treebank," *Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP2009)*, 2009, pp. 343-351.
[59] J. Cohen, "A coefficient of agreement for nominal scales.," *Educational & Psychological Measurement*, vol. 20, 1960, pp. 37-46.
[60] R Core Team, *R: A language & environment for statistical computing*, R Foundation for Statistical Computing, 2014.
[61] H.P. Grice, "Logic & conversation," *Syntax & semantics 3: Speech acts*, P. Cole & J. Morgan, eds., Academic Press, 1975, pp. 41-58.
[62] M. Rasmussen & G. Karypis, *gCLUTO: An interactive clustering, visualization & analysis system. UMN-CS TR-04-021*, 2004.
[63] G. Karypis, *CLUTO: A Clustering toolkit*, Computer Science Department, 2002.
[64] P. Levinson, *New new Media*, Pearson-Penguin Academics, 2013.
[65] A. Lascarides & N. Asher, " Segmented discourse representation theory: Dynamic semantics with discourse structure," *Computing meaning*, Springer, 2007, pp. 87-124.
[66] M. Stede, "Disambiguating rhetorical structure," *Research on Language & Computation*, 6 (3-4), 2008, pp. 311-332.
[67] F. Wolf & E. Gibson, "Representing discourse coherence: A corpus-based study," *Computational Linguistics*, vol. 31, no. 2, 2005, pp. 249-287.