

Content-driven Trust Propagation Framework

V.G.Vinod Vydiswaran
University of Illinois
Urbana, IL
vgvinodv@illinois.edu

ChengXiang Zhai
University of Illinois
Urbana, IL
czhai@cs.uiuc.edu

Dan Roth
University of Illinois
Urbana, IL
danr@illinois.edu

ABSTRACT

Existing fact-finding models assume availability of structured data or accurate information extraction. However, as online data gets more unstructured, these assumptions are no longer valid. To overcome this, we propose a novel, content-based, trust propagation framework that relies on signals from the textual content to ascertain veracity of free-text claims and compute trustworthiness of their sources. We incorporate the quality of relevant content into the framework and present an iterative algorithm for propagation of trust scores. We show that existing fact finders on structured data can be modeled as specific instances of this framework. Using a retrieval-based approach to find relevant articles, we instantiate the framework to compute trustworthiness of news sources and articles. We show that the proposed framework helps ascertain trustworthiness of sources better. We also show that ranking news articles based on trustworthiness learned from the content-driven framework is significantly better than baselines that ignore either the content quality or the trust framework.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—*Information filtering*; I.2 [Computing Methodologies]: Artificial Intelligence

General Terms

Design, Algorithms, Measurement, Experimentation

Keywords

Trust models, credibility, graph algorithms, fact-finders

1. INTRODUCTION

As online content grows exponentially, reliance on the Web content is also growing in many domains. Surveys show that more people, independent of age, now rely on online content for news, opinion, social networking, and other

aspects of personal well-being, including health. An early-2009 poll [3] found that 62% of Americans got their information about HIV and AIDS through media sources, including news websites and the Internet, as compared to only 13% who received the information from their doctors. Similarly, while the readership of traditional newspapers is declining, accessing news online has steadily increased over the past few years. In a March 2010 survey of US Internet users [1] on the primary source used to find news, it was found that the Web/Internet is, by far, the most popular source (49%) as compared to television (32%) and newspapers (9%).

The relative ease of publishing online, however, has led to a significant impact on the overall quality of information accessible to the consumers. A user searching online for news would not be able to easily distinguish a trustworthy news article from another that is biased or nonfactual. In such a scenario, users would like to know whether a claim or an article they find online is indeed trustworthy and which sources are more trustworthy than others.

To address this problem, in this paper, we study how to predict trustworthiness of textual claims. We propose a novel, content-driven, trust propagation framework that helps ascertain the veracity of free-text claims and compute trustworthiness of their sources based on the quality of evidence. As a specific instance of such a framework, we measure the trustworthiness of news sources and stories. Given a collection of news text articles from multiple sources, we would like to ascertain relevant, trustworthy news articles and sources to help users gauge the veracity of given claims.

Although trustworthiness frameworks have been studied earlier, they were mainly focused on structured “facts” (claims) and the underlying information network. There are two major lines of previous work: the first is TruthFinder [24] that aims at computing the veracity of “facts” and trustworthiness of sources, utilizing the relationship between source and claims and the interaction between claims. The second relevant framework is proposed by Pasternack and Roth [19], who extend the fact-finding models by incorporating prior knowledge about claims into an existing fact-finding algorithm and formulate it as a linear program. Both have assumed structured data, i.e., they assume that there exists explicit correspondence between sources and claims.

The problem setup addressed in this work, however, is significantly different. We postulate that sources do not necessarily state the claims they make explicitly, making it difficult for information extraction algorithms to extract them accurately. Instead, sources give certain evidence in favor of or against claims, which are in free-text form. Existing trust

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

KDD'11, August 21–24, 2011, San Diego, California, USA.

Copyright 2011 ACM 978-1-4503-0813-7/11/08 ...\$10.00.

models do not handle such scenarios. We propose that trust computation frameworks should assimilate the relevance of evidence content found for a claim, uncertainty in quality of these evidence artifacts, and the information network structure to compute trustworthiness. This paper presents such a “content-driven” framework for computing trust, which is an enhancement to previous fact-finding models.

The key contributions of this paper are as follows: (i) we propose a novel, content-driven, trust computation framework (Sec. 3) and describe factors that affect computation of trustworthiness; (ii) we present approaches to address uncertainty of evidence and similarity between evidence documents (Sec. 4); and (iii) we instantiate a model to compute trustworthiness of news sources (Sec. 5.1), utilizing trust scores assigned to news stories by humans. Our experiments show that the instantiated model helps assess trustworthiness of news sources better and that ranking news articles based on our model is significantly better than baselines that ignore either the evidence quality or the trust framework.

2. PROBLEM DEFINITION

The overall goal is to estimate the veracity of textual claims. A **claim** is a statement, topic, or a document whose veracity is unknown. It can be a “fact” (alternatively, “true claim”) or a “hoax” (alternatively, “false claim”). Previous literature have sometimes called these “facts”, but we prefer a more unbiased nomenclature.

In a typical instantiation of the problem, the claim is input to the system and the framework would estimate the veracity of the claim. One of the key factors in determining the veracity is the characteristics of the sources that support the claim. A **source** gives evidence for (or “expresses”) a claim. Trustworthy sources support true claims or contradict false claims, while untrustworthy sources support false claims or contradict true claims. In addition, a piece of **evidence** is a passage, document, or artifact that supports or contradicts a claim. “Good” (trusted) evidence boosts the source’s trustworthiness and the claim’s veracity.

An overall judgment on the veracity of the claim is made based on the trustworthiness of the sources giving evidence for the claim and the relevance and nature of evidence found. Instead of classifying claims as true or false, the proposed framework assigns confidence scores to claims and ranks evidence and sources based on trustworthiness, to help users decide the claims’ veracity for themselves. The typical questions that the framework helps answer are:

- **Veracity of claims:** How truthful is this claim?
- **Source trustworthiness:** What are the trusted sources for this claim?
- **Evidence trustworthiness:** Which pieces of evidence are more trustworthy than others for this claim?

2.1 Sample instantiation of the problem

As a specific instance of the problem, consider the need to judge the trustworthiness of news available online. We can find many news articles on a topic from various sources, but they are not equally trustworthy. The news articles may be written by biased media sources or may be opinion articles that are more biased than, say, news reports. Our overall goal is to assign veracity of reporting on a given topic, in addition to inferring which sources and news articles are more trustworthy than others.

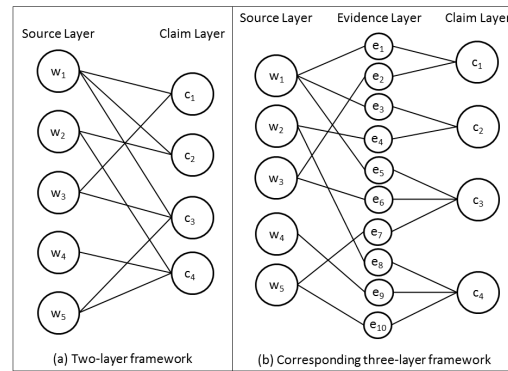


Figure 1: Two-layer trust framework and its corresponding three-layer representation

3. CONTENT-DRIVEN TRUST FRAMEWORK

3.1 Modeling trustworthiness

As mentioned in Sec. 1, the trust frameworks proposed so far are built on structured data and assume accurate information extraction. Previous work on computing trustworthiness [24, 19] are based on a bi-partite graph structure, consisting of the source layer and the claim layer (Fig. 1(a)). The nodes in the claim layer are linked to all the source nodes that express the claim. Typically, the claim score depends solely on the score of the sources linked to the claim, and conversely, the source score depends only on the score of the claims the source is linked to. A source is assumed to contribute uniformly to all the claims it expresses.

The two-layer architecture ignores the content and the context in which the source expresses the claim. To overcome this limitation, we propose a framework that includes content nodes as an intermediate layer. The framework, hence, is a three-tier graph consisting of source, evidence (content), and claim layers (see Fig. 1(b)). Each content node represents the evidence given by a source for a claim, and links to one source node and one claim node. This allows the trust framework to explicitly capture the textual context in which a source provides evidence to a claim.

Formally, let $\mathcal{W} = \{w_1, w_2, \dots, w_M\}$ be the set of M source websites. Each website w is assigned a trustworthiness score $\tau(w) \in [0, 1]$. If a source is completely trusted, its trustworthiness score $\tau(w) = 1$. Let $\mathcal{C} = \{c_1, c_2, \dots, c_N\}$ be the N claims in the dataset. Each claim c has an overall veracity score, $\sigma(c) \in [0, 1]$. $\sigma(c) = 1$ for facts and 0 for hoaxes. The veracity score may also be interpreted as the confidence of classifying a claim as true.

A claim c_i is associated with p pieces of evidence, $\mathcal{E}(c_i) = \{e_{i1}, e_{i2}, \dots, e_{ip}\}$. A piece of evidence, e , links one source website, $w(e) \in \mathcal{W}$, to one claim, $c(e) \in \mathcal{C}$, and has a confidence score, $\psi(e) \in [0, 1]$. $\psi(e)$ represents the confidence in the evidence’s trustworthiness. Finally, let $\mathcal{C}(w_i) = \{c \in \mathcal{C} \mid \exists e \in \mathcal{E} : (c(e) = c) \wedge (w(e) = w_i)\}$ denote the claims for which the source w_i gives some evidence.

In addition to the nodes, the framework also consists of interaction edges between the graph nodes. There are three main types of interactions in the proposed framework:

1. Evidence - Claim interaction (or Content Relevance): The evidence - claim edges represent the relevance

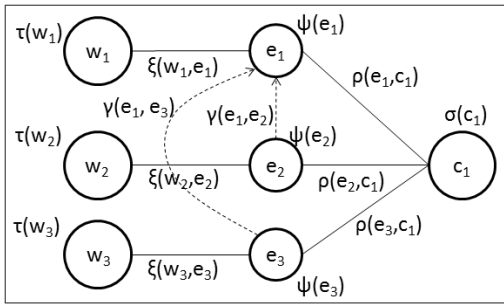


Figure 2: Factors influencing trustworthiness of claims and sources. Inter-evidence interactions are shown only with respect to evidence e_1 for clarity.

of evidence for a claim. Let $\rho(e, c)$ measure the extent of relevance of a piece of evidence e to claim c . The measure allows for highly relevant pieces of evidence to be weighted more than other evidence artifacts.

2. Evidence - Evidence interaction (or Content Equivalence): Evidence artifacts are not independent of each other. In fact, multiple sources may give very similar evidence for a claim. We define $\gamma(e_1, e_2)$ to be a measure of influence on evidence e_i due to evidence e_j . This may be either symmetric or asymmetric. Adding this measure of influence constrains the framework to make similar evidence artifacts get similar trustworthiness score, even though they may come from different sources that are not equally trustable.

3. Source - Evidence interaction: Another measure that affects the trustworthiness of claims is the confidence of the source w on a piece of evidence e , and by extension on the claim. Let $\xi(w, e)$ denote this measure. This is especially useful to model the confidence of extraction module (in the structured setup). Note that this has a different interpretation than $\tau(w)$, the trustworthiness of the source. $\tau(w)$ denotes how trustworthy are the claims made by the source: “is w a reliable source?” On the other hand, $\xi(w, e)$ denotes how confident is the fact that source w gave evidence e for the claim, and hence it can be different for different pieces of evidence given by the same source w .

Fig. 2 summarizes the factors influencing trustworthiness of claims. It shows the various parameters affecting $\sigma(c_1)$, and inter-evidence influence with respect to evidence e_1 .

3.2 Computing trust scores over framework

Next, we present techniques to compute the trustworthiness scores over the trust framework. We start with the evidence confidence scores, $\psi(e)$, which are initialized solely based on features from the evidence artifacts. Let $\psi^{(0)}(e)$ denote this initial estimate of the quality of evidence. Next, we assume uniform initial trustworthiness, $\tau^{(0)}(w)$, for all sources. If there is some prior knowledge on source trustworthiness, i.e. if some sources are “preferred” over others, then this information can be used to estimate $\tau^{(0)}(w)$. Once these nodes are initialized, the scores for nodes in other layers are computed iteratively as follows, until convergence.

Eq. 1 computes the average estimate of the claim veracity score based on the pieces of evidence relevant to the claim and the current estimate of the trustworthiness of their sources. Then, Eq. 2 re-estimates the source trust-

worthiness, based on new claim veracity scores. Finally, the evidence score is recomputed (Eq. 3) as a linear interpolation of the previous and new estimates of source trustworthiness:

$$\sigma^{(n+1)}(c_i) = \frac{\sum_{e_j \in \mathcal{E}(c_i)} [\psi^{(n)}(e_j) \times \tau^{(n)}(w(e_j))]}{|\mathcal{E}(c_i)|} \quad (1)$$

$$\tau^{(n+1)}(w_i) = \frac{\sum_{c_j \in \mathcal{C}(w_i)} \sigma^{(n+1)}(c_j)}{|\mathcal{C}(w_i)|} \quad (2)$$

$$\psi^{(n+1)}(e_i) = \mu \psi^{(n)}(e_i) + (1 - \mu) \tau^{(n+1)}(w(e_i)) \quad (3)$$

where μ denotes the interpolation parameter to control the bias of prior knowledge of $\psi(e)$ on future estimates. Once the algorithm converges, $\sigma(c)$ gives an estimate of the veracity of claim c , $\tau(w)$ shows the trustworthiness of source w , and $\psi(e)$ gives a goodness score for the piece of evidence e .

3.3 Modeling inter-evidence influence on trust scores

In addition to the source and other intrinsic properties of the evidence artifact, other pieces of evidence similar to an evidence artifact also influence its trustworthiness. To capture this effect, we introduce a similarity metric, $\gamma(e_i, e_j)$ as a measure of influence on e_i due to e_j . It may be either symmetric or asymmetric. Typical examples of $\gamma(e_i, e_j)$ between textual pieces of evidence include cosine and TF-IDF similarity, and advanced similarity metrics such as topic-sensitive similarity based on LDA [5]. Specific domain-level knowledge can also be incorporated into this function, such as rule-based named entity similarity or using thesauri for finding similarity between medical treatments.

The inter-evidence similarity $\gamma(e_i, e_j)$ can be incorporated to improve the estimate of $\psi(e_i)$ (Eq. 3), as follows:

$$\tilde{\psi}^{(n+1)}(e_i) = \lambda \left(\frac{\sum_{e_j \in \mathcal{E}(c(e_i))} [\psi^{(n)}(e_j) \times \gamma(e_i, e_j)]}{\frac{e_j \neq e_i}{|\mathcal{E}(c(e_i))|} - 1} \right) + (1 - \lambda) \psi^{(n+1)}(e_i) \quad (4)$$

where e_j spans over all pieces of evidence except e_i , that give evidence for claim $c(e_i)$. λ denotes the interpolation parameter that controls the effect of inter-evidence similarity in estimating $\psi(e)$. Such a formulation allows us to vary λ to model existing trust formulations. For example, setting $\lambda = 0$ effectively ignores the effect of evidence on each other, mirroring fact-finding algorithms such as Sums [19].

3.4 Modeling relevance of evidence

The formulation explained so far simplistically assumes all evidence artifacts are uniformly relevant to the given claim. However, an evidence may either strongly assert a claim or refer to it only in passing. In such scenarios, it may be prudent to weight the claim veracity score towards more relevant pieces of evidence. This factor is incorporated in the computation of claim veracity (Eq. 1) by adding the factor $\rho(e, c)$, as follows:

$$\tilde{\sigma}^{(n+1)}(c_i) = \frac{\sum_{e_j \in \mathcal{E}(c_i)} \phi(\psi^{(n)}(e_j), \tau^{(n)}(w(e_j)), \rho(e_j, c_i))}{|\mathcal{E}(c_i)|}$$

In this work, we have formulated the combination function ϕ as a product of the factors:

$$\tilde{\sigma}^{(n+1)}(c_i) = \frac{\sum_{e_j \in \mathcal{E}(c_i)} [\psi^{(n)}(e_j) \times \tau^{(n)}(w(e_j)) \times \rho(e_j, c_i)]}{|\mathcal{E}(c_i)|} \quad (5)$$

3.5 Modeling source-evidence interaction

As described in Sec. 3.1, the source-evidence interaction parameters, $\xi(w, e)$, are useful to model extraction accuracy in a structured domain instantiation. In trust propagation computation, the contribution of the trust score for a piece of evidence e from its source w is given by $\tau(w) \times \xi(w, e)$.

4. INSTANTIATING TRUST FRAMEWORK

Now that the details of the framework have been described, we can instantiate the framework for computing trustworthiness of claims. This requires the following steps:

1. Collecting relevant evidence for claims, and defining $\rho(e, c)$ and $\xi(w, e)$.
2. Assigning evidence scores, $\psi(e)$.
3. Defining similarity between evidence documents, $\gamma(e_i, e_j)$.

4.1 Collecting evidence for claims

When a claim is input to the system, relevant pieces of evidence are identified from a collection of documents. First, a text corpus of documents is collected and a retrieval index is built over these potential evidence artifacts. Then, for every claim c , the index is used to retrieve relevant pieces of evidence, which are then used to instantiate the evidence layer of the framework. For each retrieved document e , the retrieval score is normalized to $[0, 1]$ scale and used as $\rho(e, c)$.

For the sample instantiation to compute news trustworthiness (described in Sec. 2.1), the index is built over a collection of news stories. In this domain, since we do not have any extraction module, we uniformly set $\xi(w, e)$ to 1.

4.2 Assigning evidence scores

The initial evidence scores can be assigned in multiple ways, as follows:

1. **Uniform weight:** All pieces of evidence are set to a uniform weight, $\psi^{(0)}(e)$. If $\psi^{(0)}(e) = 1$, the framework reduces to the two-layer trust framework. $\psi^{(0)}(e) < 1$ denotes all evidence artifacts are uniformly, but not fully trusted.

2. **Assign weights based on retrieval score:** When the pieces of evidence are retrieved for a claim, they can be scored based on the rank and retrieval score given by the retrieval system such that top ranked artifacts get higher weight. In the experiments presented in this paper (Sec. 6), this represents the ranking baseline where sources and pieces of evidence are ranked based on a retrieval system, rather than the trust framework.

3. **Assign weights based on weak supervision:** It may be possible to assign non-uniform weights to the pieces of evidence based on human judgments, as humans may be able to review the artifacts to decide if they are trustworthy.

4. **Learn weights:** The evidence scores can be learned based on features from the evidence artifact, such as attribution of claims, style of writing, and reference to other articles that are trusted.

In this work, we will use the first three variants of assigning evidence scores. Our dataset consists of human-rated articles, where annotators judged stories in a stand-alone mode without looking at the source features. Hence, instead of learning the weights, we use these judgments as good signals for the score we want to learn. Learning scores using features from the evidence artifact is not considered in this paper, and is an interesting future direction to explore, especially for corpora where supervision is not available.

4.3 Finding similarity between documents

The effect of pieces of evidence on each other is captured in a similarity metric $\gamma(e_i, e_j)$, as follows:

1. **Strict similarity:** The strictest form of similarity between pieces of evidence is duplicate matching, where $\gamma(e_i, e_j) = 1$ only if $e_i = e_j$ and 0 otherwise. This may be required, for instance, to link up duplicate nodes if an evidence document is repeated as multiple nodes in the framework.

2. **Vector similarity:** The baseline similarity measure for text documents is the cosine similarity between documents, where the documents are represented as a binary vector of words and cosine similarity is given by the dot product of the two vectors. Another option is to use TF-IDF similarity, where a word’s frequency is weighted by the word’s IDF. If v_i and v_j are vector representations of the two documents, such that $v_i[w]$ is the product of frequency of word w in e_i and the IDF of w , then TF-IDF similarity is given by:

$$\text{TF-IDFsim}(v_i, v_j) = \frac{\sum_{w \in V} (v_i[w] \times v_j[w])}{\sqrt{\sum_{w \in V} (v_i[w])^2} \sqrt{\sum_{w \in V} (v_j[w])^2}}$$

3. **Topic-based similarity:** The similarity of documents can also be measured based on the topical similarity between them. A probabilistic mixture model is learned on the evidence documents for a claim, and documents are assigned to clusters based on the learned model. The details of the model are given in Sec. 4.3.1. Two documents that are clustered together have higher similarity than those from different clusters.

In the current work, these similarity factors were combined in the following formula:

$$\gamma(e_i, e_j) = \begin{cases} 1 & \text{if } e_i = e_j \\ \text{TF-IDFsim}(v_i, v_j) & \text{if } e_i \neq e_j \wedge e_j \in g(e_i) \\ \alpha \times \text{TF-IDFsim}(v_i, v_j) & \text{otherwise} \end{cases} \quad (6)$$

where v_i, v_j are TF-IDF vector representations of e_i, e_j , respectively, $g(e)$ is the cluster to which e belongs, and α is the average-length cluster similarity between $g(e_i)$ and $g(e_j)$.

4.3.1 Clustering evidence documents

Once the evidence documents are retrieved for a claim, we cluster them based on common aspects in the evidence text after learning a probabilistic mixture model. We used the PLSA model [13] with a background component [25]. Formally, let $\theta_1, \theta_2, \dots, \theta_k$ be k aspect-based unigram language models and θ_B be the background language model. A document d is generated word-by-word from the following mixture model:

$$p_d(w) = \lambda_B p(w|\theta_B) + (1 - \lambda_B) \sum_{j=1}^k [\pi_{d,j} p(w|\theta_j)]$$

where w is a word, $\pi_{d,j}$ is document-specific weight for the j^{th} aspect and $\sum_{j=1}^k \pi_{d,j} = 1$. λ_B is the mixture weight for the background model θ_B , and is set to a large value (typically, 0.95). The model parameters, $\Theta = (\theta_j, \pi_{d,j})_{j=1}^k$, can be estimated using Expectation Maximization [8] (details are omitted due to space limitation).

Once the model parameters are learned, the documents are clustered in one of k clusters based on which topic model best explains the document. A document d is assigned to cluster $j' = \arg \max_{j \in [1, k]} \pi_{d,j}$. As explained above, documents belonging to one cluster are considered more similar to each other than to documents from different clusters.

5. EXAMPLE INSTANTIATIONS

We outline how the proposed trust framework can be instantiated in different scenarios to demonstrate applicability of the framework across both structured and unstructured domains. We present sample instantiations in three domains, viz., news, books, and health forums. However, we will evaluate the framework only in the news domain.

5.1 Trusting Online News

Our first instantiation uses the proposed trust framework to capture the trustworthiness of news online (see Sec. 2.1). The objective is to check veracity of claims of the form: “News coverage on (a given topic) is fair and unbiased.” The input is, hence, in the form of topics (keyword queries). The framework outputs the veracity scores for claims and identifies most trusted news sources and stories. Here, the claim layer consists of topics of interest or genres of news that are given as input. For each topic/genre of interest, the evidence layer consists of news stories for the given topic. We define two models depending on the type of sources, such that the source layer consists of either the websites where the news was published (“*Source model*”), or the reporters writing the news story (“*Author model*”). To clarify, the *Source model* is modeled with $\langle \text{source, evidence, claim} \rangle$ layers, while the *Author model* is modeled with $\langle \text{author, evidence, claim} \rangle$ layers. As described in Sec. 4, the other parameters are set as follows: $\rho(e, c)$ is set based on retrieval score, normalized to $[0, 1]$; $\xi(w, e) = 1$ uniformly; $\psi^{(0)}(e)$ is computed based on user ratings; and $\gamma(e_i, e_j)$ is set according to Eq. 6. Clustering of news articles is done using a simple mixture model as described in Sec. 4.3.1, with $k = 4$.

The two models compute the veracity of news: how trustworthy is the news covering a particular topic of interest, or how trustworthy a news genre is, when compared to other news genres. It must be noted that trustworthiness in this domain signifies if the reporting of news is unbiased and fair, and if facts expressed in news stories are accurate with good coverage. This notably differs from finding trustworthiness of extracted numeric facts such as “Shakespeare was born in 1564.” or “The height of Mt. Everest is 8848 meters.”

5.2 Trusting Author lists on Online Bookstores

To show how the framework can be used to model trustworthiness in structured data, we present the framework with respect to the Books dataset used by Yin et al. [24]. The dataset consists of books with four extracted fields, viz., the source, the book title, ISBN, and the author list as given by the source.

To instantiate a model based on this dataset, the bookstores are the sources, and the claims are candidate lists of authors for a particular book. The evidence layer consists of the author lists given by a source for a particular book. The aim is to identify the most “trusted” candidate author list. To achieve this, the model loads the data for all books and builds the trust network. $\psi^{(0)}(e)$, $\tau^{(0)}(w)$, and $\xi(w, e)$ are set to 1. Content relevance and equivalence functions, $\rho(e, c)$ and $\gamma(e_i, e_j)$, show similarity of lists of author names, and can be built using person name similarity metrics [7]. In fact, by allowing the original text to appear in the model, our framework allows a source to support multiple claims to varying degrees, not committing to one (possibly faulty) person name normalization. This leads to a voting strategy that is more robust to data cleaning or extraction errors.

News Genre	# Stories	Avg. Rating
News Report	7,881	3.65
Opinion	6,611	3.69
News Analysis	2,634	3.73
Special Report	2,177	3.83
News	816	3.54
Editorial	634	3.73
Interview	583	3.87
Poll	301	3.73
Investigative Report	276	4.10
Review	169	3.71
Comment	141	3.67
Breaking News	124	3.66
Comedy News	114	3.75
Press Release	92	3.83
Advocacy	76	3.91
Speech	55	3.90
Statement	43	3.82
Research	34	3.96
Entertainment	17	3.39
Cartoon	12	3.74
Advertisement	3	2.43
Dramatization	2	3.48
Miscellaneous	369	3.52
Total	23,164	3.70

Table 1: News genres in the dataset, sorted by number of stories in each genre. The average rating for each genre is also shown.

5.3 Trusting Treatments from Health Forums

The framework can be also used in domains where relation extraction is needed, such as in question answering over the Web. Consider the scenario of trusting medical claims given in health forum posts. If a known disease-treatment pair is given as a relation query (the *claim*), the structured query is used to find relevant posts (the *evidence*) from medical forums and message boards (the *source*). The aim is to verify if the claims expressed in health forums are trustable, and if certain health forums are more trustable than others.

To achieve this, we can set up the framework by issuing a relation query [20] on an index over forum posts. The posts are rated on the sentiment expressed about the effectiveness of the query treatment (the *confidence in evidence*, $\psi(e)$). $\rho(e, c)$ is set depending on how strongly the treatment relation is expressed in the post, that is if the post is specifically about the treatment claim or talks about the treatment in passing. $\gamma(e_i, e_j)$ is set based on similarity of content between posts or grouping based on post authors.

6. EXPERIMENTS

The key research questions we want to evaluate are:

1. Does the trust framework help in identifying trustworthiness of claims and the most trusted sources for a claim?
2. Does the addition of content nodes in the proposed trust framework help users rank highly trusted evidence documents before other relevant, but not-as-trusted documents?
3. What is the effect of the evidence scoring strategy used? Specifically, how does evidence scoring strategy affect the overall trustworthiness of sources?
4. What is the effect of clustering articles? Specifically, does the source trustworthiness change with news genres?

6.1 Dataset characteristics

We instantiate and evaluate the framework in the news domain, as described in Sec. 5.1. News data was collected from a community-driven news review website, NewsTrust¹. NewsTrust allows members to analyze news articles on various tenets of good journalism. Members can rate stories on accuracy of facts, fairness and bias in reporting, style, depth

¹<http://www.newstrust.net/>

Topic	(1) Health care	(2) Obama administration	(3) Bush administration	(4) Democratic policy	(5) Republican policy
Trust score	0.538	0.558	0.646	0.462	0.443
Trusted Sources	OurFuture.org	Democracy Now	Democracy Now	Common Dreams	Consortium News
	TruthOut	Consortium News	TruthOut	Consortium News	Salon
	PolitiFact	Common Dreams	Consortium News	Democracy Now	Women's eNews
	FactCheck	TruthOut	Salon	FAIR	FactCheck
	AlterNet	Salon	Common Dreams	Salon	New Republic
	Media Matters	FactCheck	McClatchy	New Republic	Washington Post
	Huffington Post	Washington Independent	Huffington Post	San Francisco Chronicle	New York Times
	New York Times	Huffington Post	Boston Globe	FactCheck	Wall Street Journal
	New Republic	Washington Post	Washington Post	New York Times	Boston Globe
	CNN	New York Times	New York Times	Washington Post	The Guardian
	NPR	The Guardian	The Guardian	Women's eNews	Congressional Quarterly
	Washington Post	Los Angeles Times	AlterNet	The Guardian	Huffington Post
	Associated Press	The Hill	The Politico	Wall Street Journal	Bloomberg
	Time	The Politico	Newsweek	American Prospect	Christian Science Monitor
The Politico	Newsweek	The Hill	Los Angeles Times	The Hill	

Topic	(6) Immigration	(7) Gay rights	(8) Corruption	(9) Election reform	(10) WikiLeaks
Trust score	0.508	0.499	0.544	0.514	0.605
Trusted Sources	Democracy Now	Washington Post	Democracy Now	TruthOut	Democracy Now
	AlterNet	Slate	Consortium News	Consortium News	AlterNet
	NPR	New York Times	TruthOut	Huffington Post	McClatchy
	New York Times	Women's eNews	Seattle Times	Journalism.org	The Nation
	San Francisco Chronicle	Mother Jones	Common Dreams	AlterNet	New York Times
	Washington Post	Los Angeles Times	Talking Points Memo	New York Times	Salon
	Washington Independent	Salon	Inter Press Service	Washington Post	Wired
	Huffington Post	Huffington Post	Huffington Post	Los Angeles Times	Christian Science Monitor
	Associated Press	San Francisco Chronicle	Washington Post	Congressional Quarterly	The Guardian
	The Hill	Media Matters	Salon	Washington Independent	Economist
	Los Angeles Times	Washington Independent	New York Times	Think Progress	New Republic
	New America Media	Associated Press	Wall Street Journal	The Politico	Reuters
	CNN	The Politico	Mother Jones	Associated Press	Pajamas Media

Table 2: Variation of most trusted sources and overall trust score for various politics news topics. For each topic, only sources that contribute at least 10 articles for that topic are shown.

of coverage, attribution of quotes to sources, and other qualitative aspects of news. In addition, other popularity-based factors such as viewing, sharing, “liking”, and discussing a story are recorded. NewsTrust combines these ratings from multiple reviewers and assigns an overall score to each story. The scores are in the range of [1, 5] and are considered as “gold-standard” during evaluation. Further analysis showed that the scores were biased towards higher values (i.e., news stories were generally rated as trustworthy) and they followed a Gaussian distribution with mean 3.7 and standard deviation 0.59.

We crawled all the stories under “Politics” section from the NewsTrust website in October 2010, and identified the source (website), author, and genre for the news story. In total, we collected 24,388 stories from 23 news genres. After removing stories for which no reviews were posted, we had a set of 23,164 stories. Table 1 gives the distribution of number of stories and average rating per genre. Each story belongs to one genre, and is written by one or more authors and is posted on one or more websites, based on the authors’ affiliations. In all, we had 8,603 unique authors and 1,960 unique sources in the dataset. We built a search index over the news stories using the Lemur toolkit².

In the collected dataset, since a news story might link to multiple sources (multiple reporters working on a story, say), we could not directly treat each story as evidence when instantiating our trust framework. Instead, we duplicated the news stories such that each story links to exactly one source. Further, these duplicates were tied together with inter-evidence similarity, $\gamma(e_i, e_j) = 1$. As we observed that the scores followed a Gaussian distribution, with $\mu = 3.7$ and $\sigma = 0.59$, we normalized the scores to the range [0, 1] using a probabilistic interpretation of cumulative distribution function of the Gaussian distribution. Stating mathematically, if \mathcal{D} is a data corpus and x is the original score, the new score \tilde{x} is given by $\tilde{x} = \Pr(\text{score}(d) < x \mid d \in \mathcal{D})$.

²<http://www.lemurproject.org/>

#	Topic	Retrieval	Two-stage model	Our model
1	Health care	0.886	0.895	0.932
2	Obama administration	0.852	0.876	0.927
3	Bush administration	0.931	0.921	0.971
4	Democratic policy	0.894	0.769	0.922
5	Republican policy	0.774	0.848	0.936
6	Immigration	0.820	0.952	0.983
7	Gay rights	0.832	0.864	0.807
8	Corruption	0.874	0.841	0.941
9	Election reform	0.864	0.889	0.908
10	WikiLeaks	0.886	0.860	0.825
	Average	0.861	0.869	0.915

Table 3: NDCG values for ten news topics in politics. The improvement of the ranking using our *Trust model* is significant at $p = 0.05$ using Wilcoxon’s signed rank test, over both the *Retrieval* and *Two-stage model* runs.

6.2 Evaluating veracity of claims

We first verify if the framework can be used to judge if the claim is trustworthy or not. To evaluate this, we collected a set of ten political news topics and retrieved evidence documents for these topics from the dataset. Then, we initialized the *Source model* with these pieces of evidence, and ran the model to rate the trustworthiness of the claim topics.

Table 2 shows the overall trust score and top trusted sources for each political topic. We see that, according to the model, the news coverage of topics such as “Bush administration” and “WikiLeaks” are fairly trustworthy. In contrast, coverage on “Republican policy” and “Democratic policy” is not as trustworthy. It is also interesting to see that although the coverage of current and past presidential administration is fairly trustworthy overall, the articles on “Obama administration” are significantly less trustworthy than the coverage of “Bush administration”. For most topics, fact verifying websites such as PolitiFact and FactCheck are among the top trusted sources.

6.3 Evaluating ranking of evidence documents

To quantitatively evaluate performance, we measure how well the proposed news trust model ranks documents on

Retrieval	Two-stage model		Our Trust model		Ideal
The Politico 3.28	AlterNet	3.95	Dem. Now	3.97	NPR 4.14
Wash. Indep. 3.30	NYT	3.97	AlterNet	3.95	Dem. Now 3.97
Huff. Post 3.38	Wash. Post	3.38	NPR	4.14	NYT 3.97
AP 3.97	LA Times	3.66	NYT	3.97	AP 3.97
NYT 3.97	Wash. Indep.	3.30	SFO Chron.	3.73	AlterNet 3.95

Table 4: Top five results for a sample topic, “Immigration”, for *Retrieval*, *Two-stage model*, our proposed *Trust model*, and *Ideal* runs. Instead of docID, the document source is shown, along with the trust score. Some names have been abbreviated.

trustworthiness. Consider a use-case scenario where a user searches for a topic. Ideally, we would want to show more trusted documents higher in the ranked list. Traditional search algorithms rank documents on relevance rather than trustworthiness. When the user searches for news on a particular topic, a ranked list of relevant news stories is retrieved. Let us call this the *Retrieval* run. Using the trust model we learned, we can rerank these documents based on the trustworthiness of the source, to obtain a new document ranking (the *Trust model* run). We also rerank these document using an instantiation that mimics existing two-layer trust models (the *Two-stage model* run). Note that in the *Two-stage model* run, the graph structure is similar to that of the *Trust model*, except that all pieces of evidence get uniform weight and are not allowed to change across iterations. Using the gold-standard document scores as weights, we compute the normalized discounted cumulative gain (NDCG) measure for all three runs. NDCG [14], a ranking performance measure, computes the ratio of discounted cumulative gain of a run (where the weight of a relevant document is discounted by its rank) to the *ideal* discounted cumulative gain (assuming documents are ranked in decreasing order of their weights). NDCG values are in $[0, 1]$ and higher numbers are preferred.

Table 3 shows the NDCG values for ten news topics in politics. We see that the *Trust model* run has the highest NDCG value for eight of the ten topics. The improvement is statistically significant at $p = 0.05$ level using Wilcoxon’s signed rank test, compared to both the *Retrieval* run and the *Two-stage model* run. Table 4 shows a sample comparison of ranking in the *Retrieval*, *Two-stage model*, *Trust model*, and *ideal* runs for the topic “Immigration”. For each run, the source name of the document and the document’s rating is shown. We see that the *Trust model* based ranking is significantly better than ranking based on either the *Retrieval* run or the *Two-stage model*, and has ranked documents from more trusted sources higher. It is not surprising that our method outperforms the *Retrieval* run as the latter doesn’t model trustworthiness. However, the fact that our method also outperforms the *Two-stage model* significantly shows that incorporating content information into the trust framework is beneficial.

6.4 Effect of varying evidence scoring scheme

In prior research on trust models, the trustworthiness of claims is based on trustworthiness of sources, and the links between sources and claims are assumed to be ideal (completely trustworthy). In contrast, we propose a framework where the links (representing evidence for a claim from a source) are assumed to have a non-uniform trust rating.

To evaluate the effect of using evidence-based scoring, we

Voting (Popularity)	Average rating	Our Trust model
News Report	Investigative Report	Speech
Opinion	Research	Investigative Report
News Analysis	Advocacy	Interview
Special Report	Speech	Press Release
News	Interview	Poll
Editorial	Press Release	Statement
Interview	Special Report	Opinion
Poll	Statement	Research
Investigative Report	Comedy News	Comedy News
Review	Cartoon	Review

Table 5: Top ten news genres based on different scoring schemes, viz. (i) voting (popularity), (ii) average rating, and (iii) our trust model.

set up two model variants – one in which all pieces of evidence have a uniform score (of 1.0), while in the other, the scores are based on the trust scores based on reviews given by users of NewsTrust. The former corresponds to previous fact-finding models, such as TruthFinder [24] and Sums [19], where the evidence layer is essentially ignored.

Table 6(a) shows the effect of variation of evidence scoring scheme for the *Source model*. The first column shows the results when all pieces of evidence get the same weight and the sources are ranked by popularity. The second column shows results when pieces of evidence get non-uniform scores, but the sources are ranked on the average of trust-rating score of pieces of evidence from the source (without using the trust framework). The third column shows how the ranking varies when the trust rating is used to score evidence artifacts in the *Source model* and the model is let to iterate. We observe that when the model is allowed to converge³, more well-known sources tend to come up higher in the ranked list and relatively unknown sources that have few, but trusted, stories are pushed further down the list.

Unfortunately, we do not have gold labeling of true trustworthiness of sources to give a complete quantitative evaluation. Instead, we found the results of a 2010 poll [2] that asked participants which TV news sources are “mostly fair” (as compared to “liberal” or “conservative”). Based on this partial ranking, we find that our trust framework gives a Kendall’s rank correlation measure⁴, $\tau = 0.714$ when compared to the “ideal” ranking given by user polls, better than using voting or average rating ($\tau = 0.143$ for both).

Table 6(b) shows the results for a similar setup for the *Author model* and we see that the top results from the *Author model* (column 3) are all award-winning investigative reporters, unlike in the other two variants.

6.5 Effect of trust scores on news genres

We now try to understand how the proposed framework using trust-based evidence scores affects the rating of news genres in general. We use the *Source model* to rate news genres based on trustworthiness of news articles in that genre.

³Empirically, the model converges in 3–5 iterations.

⁴Kendall’s rank correlation measure [16] compares two ranked lists, L_1 and L_2 , by computing the following:

$$\text{Kendall's } \tau = \frac{\#(\text{concordant pairs}) - \#(\text{discordant pairs})}{\#(\text{total pairs})} = \frac{n}{\binom{n}{2}}$$

where a pair of entries (o_1, o_2) are said to be concordant if o_1 appears either before or after o_2 in both lists L_1 and L_2 , and is called a discordant pair if the order of the pair (o_1, o_2) is reversed from L_1 to L_2 . The value of τ ranges from -1 for reversed lists (fully discordant) to $+1$ for concordant lists.

(a) Top News Sources (using <i>Source model</i>)		
Voting (Popularity)	Average rating	Our Trust model
New York Times	3BlueDudes	ABC News
Washington Post	Democracy Now	CNN
Huffington Post	Consortium News	Wall Street Journal
Associated Press	Media Matters	Reuters
Los Angeles Times	TruthOut	Associated Press
The Politico	Common Dreams	San Francisco Chronicle
Wall Street Journal	The Nation	The Hill
AlterNet	FactCheck	Newsweek
Democracy Now	AlterNet	Seattle Times
Salon	Salon	Seattle Post Intelligencer
TruthOut	Seattle Post Intelligencer	The Politico
Seattle Post Intelligencer	New Yorker	Huffington Post
The Guardian	Mother Jones	MSNBC
Consortium News	McClatchy	Atlantic Monthly
Reuters	Huffington Post	BBC News
The Nation	The Guardian	NPR
CNN	Seattle Times	The Independent
McClatchy	New York Times	New Republic
Newsweek	Washington Post	Time

(b) Top News Reporters (using <i>Author model</i>)		
Voting (Popularity)	Average rating	Our Trust model
Amy Goodman	Jim Hightower	Juan Gonzalez
Glenn Greenwald	Sharif Abdel Kouddous	Robert Parry
Paul Krugman	Juan Gonzalez	Amy Goodman
Robert Parry	Ray McGovern	Jason Leopold
Sam Stein	Robert Parry	Glenn Greenwald
Juan Gonzalez	Amy Goodman	Paul Krugman
Jonathan Weisman	Jason Leopold	Frank Rich
Dan Balz	Glenn Greenwald	Dan Froomkin
Jason Leopold	Bill Moyers	John Nichols
Shailagh Murray	Greg Palast	Jon Stewart
Adam Nagourney	Paul Krugman	Paul Kane
Dan Eggen	Scott Horton	Dan Eggen
Paul Kane	Matt Taibbi	Juliet Eilperin
Peter Baker	Robert Reich	Michael D. Shear
Michael D. Shear	Lori Robertson	Arianna Huffington
Jeff Zeleny	Frank Rich	Sam Stein
Frank Rich	Michael Winship	Peter Baker
Jon Stewart	David Sirota	Jason Linkins
Ed Morrissey	Dan Froomkin	Shailagh Murray

Table 6: Top news sources and reporters as computed by three scoring schemes, viz., (i) voting (popularity), (ii) average rating, and (iii) our *Trust model* with evidence influence; after filtering out sources with less than 100 (or reporters with less than 30) stories each in the dataset.

Overall	News Report	Opinion	News Analysis	Special Report	Editorial	Interview
ABC News	Washington Times	Weekly Standard	Associated Press	Newsweek	Wall Street Journal	Huffington Post
CNN	ABC News	The Moderate Voice	Wall Street Journal	The Politico	Economist	Washington Post
Wall Street Journal	Seattle Post Intelligencer	Think Progress	Los Angeles Times	Associated Press	The Nation	PBS
Reuters	Agence France-Presse	Wall Street Journal	Huffington Post	Los Angeles Times	Washington Post	Bill Moyers Journal
Associated Press	CBS News	Daily Kos	AlterNet	New York Times	Seattle Post Intelligencer	MSNBC
San Francisco Chronicle	San Francisco Chronicle	Op Ed News	Time	Washington Post	New York Times	Democracy Now

Table 7: Variation of most trusted sources for different genres of news. For each genre, only sources that contribute at least 30 articles for that genre are included in the analysis.

We varied the scoring scheme, similar to those explained in Sec. 6.4, and Table 5 summarizes the findings.

We see that although “News Report” and “Opinion” are the most popular news genres, the stories in those genres do not tend to be highly trustworthy, and are not in top 10 news genres based on average rating of news stories. In contrast, “Investigative Reports” and “Press Releases” are more trustworthy genres. It was interesting to observe that “News Analysis” rates quite poorly, even lower than “Comedy News” in the ratings given by NewsTrust members and in our trust model results.

6.6 Effect of genre on source trustworthiness

Next, we wanted to analyze if the trusted sources tend to remain the same or vary across genres. Do some news sources specialize in specific genres and are more trusted than others for these specific genres? We compared the top news sources for different genres of news to study the pattern and Table 7 summarizes the results for six genres.

We find that the top-ranked sources per genre vary significantly across genres. For example, in the “Interview” genre, *PBS* and *Bill Moyers Journal* that specialize in this genre are rated high on trust scores. Similarly, *Wall Street Journal* and *Economist* are most trusted in the “Editorial” genre, while they do not occur in top trusted sources for “News Report” and “Special Report” genres. Interestingly, for the “Opinion” genre, the most trusted sources are primarily online news magazines and blogs.

7. RELATED WORK

In recent years, researchers have worked on different notions of trust for specific scenarios. Similar to the work on trust models introduced earlier, most of the research has focused on structured domain. Dong et al. [9, 10] propose models to detect copying of information and aggregate in-

formation from multiple sources to detect true value of a data item. Wu and Marian [23] find most suitable numeric answers by corroborating evidence from multiple sources. In contrast, our work proposes a general framework that uses content to estimate trustworthiness of claims, and can be extended to find most suitable answer (claim) by comparing claim trustworthiness as a post-processing step. In other works, Gallard et al. [11] incorporate the “hardness” of facts in the trust model, such that knowing easy facts earns less trust than knowing hard facts. WikiTrust [4] measures trustworthiness of Wikipedia pages based on revision history.

There has been some work in measuring quality of text in community-generated data. Shah and Pomerantz [21] measure the “quality” of answers in community QA setting using machine learning approaches. The definition of quality is, however, akin to popularity rather than trust. Su et al. [22] try to detect trustworthiness of community QA in terms of textual and linguistic features alone.

Formulating trust frameworks as iterative algorithms is not new. Search engines learn “quality” of a website based on reputation-based models such as PageRank [6] and HITS [17] using similar computation on the Web graph. In HITS, for instance, each website has a hub score and an authority score. The hub score of a webpage is computed using the authority scores of neighboring pages, and vice versa. Topic-specific extensions have been proposed to compute PageRank over websites per topic [12].

Researchers in other fields, including social scientists, have studied the theoretical and social aspects of trust. Lankes [18] studied how the advent of Internet has shifted the user’s perception of credibility from authoritative figures, such as librarians, to reliable entities, such as Web search engines. Kelton et al. [15] developed principles of trust by integrating behavioral and social science studies and research on information quality. Our work proposes a computational framework to facilitate such an integration.

8. CONCLUSION AND FUTURE WORK

In this paper, we proposed a content-driven trust computation framework that incorporates the quality of evidence content to compute the trustworthiness of sources and determine the veracity of free-text claims. We instantiated the framework in the news domain using human ratings as supervision for determining quality of evidence. We showed that incorporating evidence into the trust computation not only allows modeling trustworthiness in unstructured domain, but also improves the performance of the framework to find trustworthy sources. Reranking news articles based on trustworthiness computed by the proposed framework shows statistically significant improvement over retrieval-based ranking that ignores trustworthiness of sources and over trust models that ignore variation in evidence quality.

The current instantiation of the framework utilizes weak supervision at the evidence level in the form of human judgment on the trustworthiness of news articles. However, such supervision may not be available in other domains or for all news articles. In future, we plan to learn the weights in a supervised setting. We also plan to apply the trust framework on other domains with claims in any free-text sentence form, and use linguistic resources to compute content similarity.

Acknowledgments

The survey results reported in this paper were obtained from searches of the iPOLL Databank and other resources provided by the Roper Center for Public Opinion Research, University of Connecticut. This research was supported by the Multimodal Information Access and Synthesis Center at UIUC, part of CCICADA, a DHS Science and Technology Center of Excellence and the Army Research Laboratory under agreement W911NF-09-2-0053. Any opinions, findings, conclusions, or recommendations are those of the authors and do not necessarily reflect the view of the sponsors.

9. REFERENCES

- [1] Gather: The Changing Face of News Media, May 25th, 2010. <http://www.emarketer.com/>.
- [2] Polls: PBS Most Trusted News Source. TVNewsCheck, Feb 18th, 2010 (Retrieved Feb 16th, 2011). <http://www.tvnewscheck.com/>.
- [3] Survey by Henry J. Kaiser Family Foundation, January 26-March 8, 2009.
- [4] B. T. Adler and L. de Alfaro. A Content-driven Reputation System for the Wikipedia. In *Proc. of World Wide Web (WWW)*, pages 261–270, 2007.
- [5] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.
- [6] S. Brin and L. Page. The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Computer Networks*, 30(1–7):107–117, 1998.
- [7] W. W. Cohen, P. Ravikumar, and S. E. Fienberg. A Comparison of String Distance Metrics for Name-Matching Tasks. In *IJCAI Workshop on Information Integration on the Web*, 2003.
- [8] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of Royal Statistical Society, Series B*, 39(1):1–38, 1977.
- [9] X. L. Dong, L. Berti-Equille, Y. Hu, and D. Srivastava. Global Detection of Complex Copying Relationships Between Sources. *Proc. of VLDB Endowment (PVLDB)*, 3(1):1358–1369, 2010.
- [10] X. L. Dong, L. Berti-Equille, and D. Srivastava. Truth Discovery and Copying Detection in a Dynamic World. *Proc. of VLDB Endowment (PVLDB)*, 2(1):562–573, 2009.
- [11] A. Galland, S. Abiteboul, A. Marian, and P. Senellart. Corroborating Information from Disagreeing Views. In *Proc. of WSDM*, pages 131–140, 2010.
- [12] T. H. Haveliwala. Topic-sensitive PageRank. In *Proc. of 11th Intl. Conf. on World Wide Web (WWW)*, pages 517–526, 2002.
- [13] T. Hofmann. Probabilistic Latent Semantic Indexing. In *Proc. of 22nd Intl. ACM Conf. on Research and development in Information Retrieval (SIGIR)*, pages 50–57, 1999.
- [14] K. Järvelin and J. Kekäläinen. Cumulated Gain-based Evaluation of IR Techniques. *ACM Transactions on Information Systems*, 20(4):422–446, 2002.
- [15] K. Kelton, K. R. Fleischmann, and W. A. Wallace. Trust in Digital Information. *Journal of the American Society for Information Science and Technology*, 59(3):363–374, 2008.
- [16] M. G. Kendall. A New Measure of Rank Correlation. *Biometrika*, 30:81–89, 1938.
- [17] J. M. Kleinberg. Authoritative Sources in a Hyperlinked Environment. *Journal of ACM*, 46(5):604–632, 1999.
- [18] R. D. Lankes. Credibility on the Internet: Shifting from Authority to Reliability. *Journal of Documentation*, 64(5):667–686, 2007.
- [19] J. Pasternack and D. Roth. Knowing what to believe (when you already know something). In *Proc. of Intl. Conf. on Computational Linguistics (COLING)*, pages 877–885, 2010.
- [20] D. Roth, M. Sammons, and V. Vydiswaran. A Framework for Entailed Relation Recognition. In *Proc. of 47th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 57–60, 2009.
- [21] C. Shah and J. Pomerantz. Evaluating and Predicting Answer Quality in Community QA. In *Proc. of 33rd Intl. ACM SIGIR Conf. on Research and development in Information Retrieval*, pages 411–418, 2010.
- [22] Q. Su, C.-R. Huang, and H. K. Yun Chen. Evidentiality for Text Trustworthiness Detection. In *Proc. of the Workshop on NLP and Linguistics: Finding the Common Ground*, pages 10–17, 2010.
- [23] M. Wu and A. Marian. Corroborating Answers from Multiple Web Sources. In *Proc. of the 10th Intl. Workshop on Web and Databases (WebDB)*, pages 1–6, 2007.
- [24] X. Yin, J. Han, and P. S. Yu. Truth Discovery with Multiple Conflicting Information Providers on the Web. *IEEE Transactions on Knowledge and Data Engineering*, 20(6):796–808, 2008.
- [25] C. Zhai, A. Velivelli, and B. Yu. A Cross-Collection Mixture Model for Comparative Text Mining. In *Proc. of Intl. Conf. on Knowledge Discovery and Data Mining (KDD)*, pages 743–748, 2004.