Professional Credibility: Authority on the Web

Jack G. Conrad
Research & Development
Thomson Reuters
Professional
St. Paul, MN 55123 USA
Jack.G.Conrad@ThomsonReuters.com

Jochen L. Leidner
Research & Development
Thomson Reuters Prof.
Scientific Division
London, England EC1N 8JS
UK
Jochen.Leidner@ThomsonReuters.com

Frank Schilder
Research & Development
Thomson Reuters
Professional
St. Paul, MN 55123 USA
Frank.Schilder@ThomsonReuters.com

ABSTRACT

Opinion mining techniques add another dimension to search and summarization technology by actually identifying the author's opinion about a subject, rather than simply identifying the subject itself. Given the dramatic explosion of the blogosphere, both in terms of its data and its participants, it is becoming increasingly important to be able to measure the *authority* of these participants, especially when professional application areas are involved. After having performed preliminary investigations into sentiment analysis in the legal blogosphere, we are beginning a new direction of work which addresses representing, measuring, and monitoring the degree of authority and thus presumed credibility associated with various types of blog participants. In particular, we explore the utility of authority-detection layered atop opinion mining in the legal and financial domains.

Categories and Subject Descriptors

H.3.0.a [**Information Storage and Retrieval**]: General—*Web Search*; I.2.7.i [**Information Storage and Retrieval**]: Natural Language Processing—*Web Text Analysis*

General Terms

Experimentation, Measurement

Keywords

opinion mining, blog mining, blog monitoring, authority, credibility

1. INTRODUCTION

Analyzing text regarding its sentiment can be extremely valuable to a customer who is looking for information about a company, a brand, a product or a service. But it is also an increasingly important resource for other types of information, including professional expertise and actionable insights.

The blog-space is an expanding environment where consumers go to seek and share opinions that may or may not be useful. More and more customer viewpoints are posted on blogs. The content of the these blogs ranges from short product reviews by consumers to elaborate essays on legal issues by law professors.

Yet there is more than the recent development of Weblogs that offer content for sentiment analysis. Traditional forums for movie or product reviews, political news or legal discussions, provide material that may be analyzed according to expressed opinions that favor one viewpoint or another.

For enterprises such as Thomson Reuters that provide a wealth of information to business professionals practicing in fields such as finance, science, health care, and law, the sheer amount of data available today is making it increasingly important to be able to rank and filter this information by its credibility, which is often presumed to be correlated to authority. To this end, we have conducted a basic study to determine how closely human annotators would agree when assigning authority to highly opinionated blog commentaries. The rationale behind this study is that one cannot expect a machine to perform at a level that surpasses human agreement. The exercises described below are one means of establishing reasonable baselines for future authority-related monitoring and performance.

1.1 Definitions

In order to provide a foundation for the concepts upon which we focus as well as their relationships, we being with a set of definitions.

1. **Authority** — an accepted source of information or advice, either an expert on the subject or a persuasive force.

2. **Credibility** — a quality of being believable, trustworthy.

3. **Popularity** — the quality of garnering the favor of the general public or a particular group of people.

4. **Trust** — a reliance on the integrity, ability, credibility of a person or source of information.

5. **Sentiment** — an attitude or opinion towards something; a thought influenced by feeling or sensibility.

6. **Fact** — something known to be true based on observation or experience.

We discuss the prospects for exploiting what are arguably more objective, observable if not measurable qualities like authority, as a means of gauging more subjective, harder

to measure qualities like credibility and degree of trust. In short, we seek an operational definition of credibility, and the features investigated are decidedly more multifaceted and complex than simple attributes such as popularity.

The remainder of this paper is organized as follows: Section 2 briefly outlines prior work in credibility and authority and contrasts these with the related topics of trust and persuasion. Section 3 presents two salient use cases and how they may relate to earlier sentiment analysis work. Section 4 describes our experimental framework and the annotated blogs we developed for blog-related applications. In Sections 5 and 6, we draw our conclusions and discuss future work.

2. PRIOR WORK IN AUTHORITY AND CREDIBILITY

There has not been nearly as much research performed in the area of Web-based credibility as there has been in the area of sentiment analysis and opinion mining. Ulicny and Baclawski reported at ICWSM07 on seminal work which attempted to create profiles of users using a large and varied feature set that they showed to correlate with credibility [12], upon which an SVM model could be trained. Danielson and Fogg are involved in the Web Credibility Project at Stanford, a part of Stanford's Persuasive Technology Lab (credibility.stanford.edu) that has produced encyclopedia-type entries on user persuasion and Web credibility [3, 4]. Earlier works on Web-based authority have pre-dated the recent explosion of the blogosphere and other prominent social media applications [7]. Technorati[1] uses the log of the number of incoming blog links over six months as a measure of authority [10]. If one views authority as a proxy for credibility, in the sense that authority is correlated with credibility, then this measure may have merit. The question, however, is—is such a singular metric dependable? This is an important question which is addressed below.

An appreciable body of work has been devoted to the topics of trust and persuasion relative to information or arguments present on the Internet [6, 5]. These qualities tend to be degrees possessed by the respondent rather than properties of the contributor. Given this property, for the purposes of our focus in this short paper, we will consider these topics of marginal relevance and thus out of scope.

2.1 Extensions to Authority/Credibility Work

In lieu of an algorithm to track these related qualities, we will propose a supervised learning approach which involves a multi-dimensional model along with training data that possesses numeric scores (e.g., sentences that are assigned "authority scores"). A machine learning engine such as an SVM could be harnessed to "learn" the applicable model. In its most basic state, the default approach relies upon one's level of activity in the blogosphere as a gauge of authority. But given additional evidence, this baseline is promptly superseded by a more sophisticated, combinational model based on other evidence which is represented as features. Some prospective classes of features are illustrated in Table 1. The diversity of these feature sets underscores how a rich and varied collection of evidence is desirable in order to track authority reliably and not be misled by activity-levels alone. As this field matures, it may be possible to calculate confidence-levels for final authority scores, based on the number and quality of the features available.

[1] www.technorati.com

1.	Activity-level
	— E.g., cumulative blog participation
2.	Nature of Web alias
	— E.g., comical, witty vs. conventional
3.	Proper name features
	— E.g., Arthur C. Clarke vs. M. Mouse
4.	Title features
	— E.g., Dr., Prof., ... Ph.D., M.D.
5.	Citation features
	— E.g., authority levels of persons, URLs cited
6.	Internet domain features
	— E.g., transparent, relevant country
7.	Online resume-related features
	— E.g., professional affiliations
8.	Linguistic features
	— E.g., degree/quality of noun phrases, entities
9.	Grammatical features
	— E.g., level of diction
10.	Statistical features
	— E.g., average length of sentences

Table 1: Feature Classes for Measuring Authority

3. APPLICATIONS

We have identified two significant use cases that highlight the importance of robust, multi-faceted, and reliable mechanisms for determining the authority of contributors to the blogosphere.

3.1 Relevant Use Cases

3.1.1 Use Case 1: Authority in the Blawgosphere

One compelling use case in the legal blogosphere (a.k.a. the *blawgosphere*)[2] is found in blogs which discuss important decisions and trends generated by the court system. Simple measures of authority such as that used by Technorati basically reward Web-based activity and longevity. Given today's highly Web-engaged law students, it is possible to have, for instance, a second year law student who is this active on the Web and in particular the blawgosphere. By contrast, there could be a noted authority in a specific legal practice area such as Constitutional Law who has not had much of a presence on the Web, but who may have just started his own blawg. Basic authority measuring algorithms would rank the historically more active law student above the legal scholar, despite the clear disparity in professional credentials of the two. This phenomenon underscores the need for more robust approaches to capturing and representing such credentials when attempting to determine the authority, and, by extension, credibility, of the blog participant. Affiliation-related attributes, for example, identified from URLs tied to home pages or staff pages, could help.

3.1.2 Use Case 2: Authority in the Flogosphere

A relevant use case in the financial blogosphere (a.k.a. *flogosphere*)[3] comes from blogs which focus on activities and performance levels involving financial markets or financial commodities. In this professional discipline that is distinct from the legal-space, it would be possible to have an experienced financial analyst with a mediocre track record of predictions and recommendations who is a regular partici-

[2] *blawg* = legal blog

[3] *flog* = financial blog

pant in the flogosphere. By contrast, there may be other time and market-tested experts with significantly more impressive performance records, though much less active in such online forums. Here too, common techniques to gauge the "authority" and presumed credibility of such participants would err. Note that the second use case differs from the first insofar as it can be validated using external resources. This situation clearly invites techniques that leverage broader and more comprehensive means of capturing such professional credentials, from personal or corporate Web sites to linguistic features of the written text.

These illustrations represent realistic scenarios that are distinct from the general practices in blogging where anonymity rather than transparency is the norm. Determining authority and credibility become more problematic in this case and may also introduce further problems [1]. However, professional 'users' in the blogosphere will typically demand transparency in the material presented, because their actions have to rely on the retrieved information. Hence, authority and credibility are key to mining useful information from blogs for professional customers.

3.2 Related Sentiment-based Investigations

A recent, comprehensive, and thorough report on the state of the art in sentiment analysis can be found in Pang and Lee's FTIR article [9]. We have explored the utility and effectiveness of sentiment analysis resources in a preliminary study within the blawgosphere, regarding bloggers comments about the current state of legal research tools such as Westlaw[4] and Lexis-Nexis.[5] This was done by harnessing *Ling-Pipe* from Alias-I,[6] a toolkit which contains sentiment analyzing resources that includes character-based language models as well as token-based Naïve Bayes techniques [2]. One of the chief lessons learned was the importance of the credibility of the author, in this case, the authority of the author. A second-year law student, for example, who is bemoaning the fact that legal research engines are not user friendly or intuitive, might not carry with her the same degree of credibility as a law firm associate with six years of experience who has used the existing research tools for both small and large projects under a spectrum of timelines.

4. AUTHORITY ANNOTATION TRIAL

In order to be able to reliably measure and assign types of authority values, gold standards or tagged exemplar blog entries or comments can be indispensable. To this end, we harnessed two paralegal members of our staff to examine and classify the entries and comments to two legal blogs: the first from The Volokh Conspiracy [11],[7] the second from The Balkanization Blog [8].[8]

For the two sets of blog entries and associated comments, we asked our paralegal annotators to assign the following attributes. A four-point Likert scale was used for the inspection and assignment of authority. The annotators were asked to examine a wide variety of features in their determination of a respondent's level as evidenced by the description below.

---

[4] www.westlaw.com

[5] www.lexisnexis.com

[6] www.alias-i.com

[7] www.volokh.com

[8] balkin.blogspot.com

- REFERENT ID — which represents the ID a comment refers to, e.g., entry, c1, c2, c3 ....

- POLARITY — { AGREE, DISAGREE, BALANCED, UNRELATED }

- DEGREE — { 1, 2, 3 } at token or phrase level, where 1 is mild, 2 is medium, and 3 is intense.

- AUTHORITY — { 0, 1, 2, 3 } where the authority is judged for each person, not for each entry. If someone posts more than one entry, it would be judged based on all the comments the person wrote and the information offered by the mandatory link to the person's Web page or email address. Sometimes a real full name is given, but no link to a blogger entry. A Web search needs to be carried out to confirm the authority of this person. Possible scores are illustrated below.

  - 0, no evidence present that author is of any authority (the default)
  - 1, some indications of being an authority (e.g., writing style)
  - 2, more than one indication that the person is an authority (e.g., writing style, profession, law blogs)
  - 3, clear authority in the field (i.e. law professor, attorney with relevant practice area, etc.)

4.1 Results

Table 2 presents basic agreement figures between the annotators for the authority tagging task. There are at least three immediate take-aways from these results. The first is that the task can hardly be characterized as simple, given how frequent the paralegals disagreed, albeit, typically by a single unit. We did observe several reasons why the annotators did not assign identical authority tags as often as they might have. A majority of the respondents did not have a link to a Web-based profile or home page. Often this left reliance upon the language and mode of argumentation (internal evidence) as keys to judgment, and clearly such evidence is limited and leaves the ultimate assignment up to the subjectivity of the reviewer. It is worth noting that upon closer inspection, there was no clear pattern that was identified, for instance, of one annotator giving consistently higher scores than the other. The second is that given the ± 1 (within 1) agreement that is consistently achieved between them, such agreement may still be able to contribute to the overall assessment standards for a system which automatically assigns authority to the blog participants being examined. Lastly, the Kendall tau scores suggest that annotators may be able to achieve a higher degree of correlation when they have a larger set of entries to work with.

4.2 Open Research Questions

In performing the experiment outlined above, we identified several essential research questions involved in the quantification of human judgment of authority. Some of these include:

1. Are the propagation mechanisms for authority, credibility and trust the same, or do they require separate treatment?

Legal Blog	Number of Comments	Complete Agreement	Percentage Agreement	Within 1 Agreement	Percentage Agreement	Kendall tau
Volokh Conspiracy	18	4	22%	17	94%	0.49
Balkanization	30	22	73%	30	100%	0.88
Combined	48	26	54%	47	98%	0.69

Table 2: Inter-assessor Agreement for Blog Respondent Authority Levels

2. What are the risks and possible consequences when authority does not correlate well with credibility? Can it be empirically quantified how often this happens?

3. How are the pieces of evidence that contribute to establishing evidence for an author on the Web best aggregated?

4. How do we 'calibrate' these authority levels, given that to some extent someone who is a clear authority to *Party A* may not hold the same degree of authority to *Party B*?

5. What are the *external* characteristics (e.g., professional affiliation) versus *structural* characteristics (e.g., via links, comment indicators, etc.) of authority, and to what extent are the two correlated?

6. What would be an effective mark-up scheme in order to create a gold standard for learning how to score credibility, for instance, based on one's interaction with the blogosphere?

7. How can any trust-based scoring system be made resilient against adversarial behavior (e.g., spam attacks)?

Answers to questions like these would ideally be at least partially answered empirically together with a consensus formed through forums such as the *Workshop on Information Credibility on the Web (WICOW)* and other research venues.

5. CONCLUSIONS

Given the rapid and ongoing growth of the Internet, and particularly of the blogosphere, the ability to distill the views and opinions expressed according to an author's credibility is increasingly essential. This ability is especially significant for professional information providers like Thomson Reuters. That this author-attribute may be tracked and measured by authority at least partially validates the initial steps of the study we have outlined above. Although the concurrence achieved within this framework were mixed at best, they nonetheless point to the fact that consistent loose agreement is achievable and this can be leveraged to bolster systems which serve and rank content according to authority and thus presumed credibility. Just as sentiment analysis is proving to be an invaluable addition to certain types of information needs, knowledge of credibility, leveraging proxies such as authority metrics, also promises to deliver additional value to information retrieval needs, most notably in the blogosphere.

6. FUTURE WORK

Once a sufficient amount of gold data has been created using procedures similar to those described above, we can begin to test baseline approaches to automate means of determining and assigning authority scores. An important piece

of this research will include the identification of a diverse set of features with which to model the participants, a model which can be refined and optimized according to the measured performance levels of this authority-detecting component. Another essential piece is obviously access to the blogosphere data itself. To that end, we are in the process of accumulating substantial quantities of both blawgosphere and flogosphere data resulting from current crawls.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] Jonathan Alter. All umbrage all the time. *Newsweek*, page 29, July 28, 2008.

[2] Jack G. Conrad and Frank Schilder. Opinion mining in legal blogs. In *Proceedings of the 11th International Conference on Artificial Intelligence and Law (ICAIL 2007)*, pages 231–236, Palo Alto, CA, 2007. ACM Press.

[3] David R. Danielson. Web Credibility. *Encyclopedia of Human-Computer Interaction, C. Ghaoui (ed.)*, pages 713–721, 2005.

[4] B. J. Fogg, Gregory Cuellar, and David R. Danielson. Motivating, influencing, and persuading users. *The Human-Computer Interaction Handbook: Fundamentals, evolving technologies, and emerging applications, J. A. Jacko and A. Sears (eds.)*, 2008. In Press.

[5] Ramanatham Guha, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. Propagation of trust and distrust. In *Proceedings of the 13th International World Wide Web Conference (WWW04)*, pages 403–412, New York, NY, 2004. ACM Press.

[6] Amubhav Kale, Amit Karandikar, Pranam Kolari, Akshay Java, Tim Finin, and Anupam Joshi. Modeling trust and influence in the Blogosphere using link polarity. In *Proceedings of the First International Conference on Weblogs and Social Media (ICWSM07)*, Boulder, CO, 2007. www.icwsm.org.

[7] Jon M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, Sept. 1999.

[8] Sandy Levinson. "How Times Change re the Second Amendment". In *Balkinization blog*. http://balkin.blogspot.com/2007/11/how-times-change-re-second-amendment.html, Nov. 25, 2007.

[9] Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135, 2008.

[10] David L. Sifry. State of the Blogosphere. In *www.technorati.com/weblog/2006/11/161.html*, Nov. 6, 2006.

[11] Ilya Somin. "potential consequences of enacting the equal rights amendment". In *Volokh blog*. www.volokh.com/posts/1181081750.html, Nov. 27, 2007.

[12] Brian Ulicny and Ken Baclawski. New metrics for newsblog credibility. In *Proceedings of the First International Conference on Weblogs and Social Media (ICWSM07)*, Boulder, CO, 2007. www.icwsm.org.