Decision makers and citizens are heavily influenced these days by information they obtain from online resources — from news portals, online encyclopedias, blogs and forums, product websites and reviews, etc. However, the lack of control over what gets published online can often lead to dissemination of unreliable and misleading information. In such a scenario, how can one verify if a claim is true or determine which sources are trustworthy?

In this tutorial, the instructors systematically consider various approaches that address this problem. This tutorial is aimed at AI researchers who are interested in future research on trustworthiness of information. The goal of the tutorial is to present the audience with an exhaustive survey of relevant research on this topic and to outline future directions of research that may interest the AI community. No specific background knowledge is assumed of the audience.

**Tutorial outline**

The intended duration of the tutorial is four hours. The detailed outline and syllabus of the tutorial is given below.

1. **Introduction & Motivation [15 min]**
   - Real-life examples showing need to verify claims and study information trustworthiness across different domains

2. **Trustworthiness approaches based solely on source features [45 min]**
   - Graph based approaches: PageRank [3], topic-sensitive PageRank [7], propagation of trust and distrust [6]
   - Medical source reliability using feature-based classification approaches [14]
   - Text-centric approaches to check claims, using community knowledge [17]
   - Using reputation of sources [2, 1]
3 Fact-finders: Trustworthiness approaches based on sources and claims [30 min]

- Voting, “Sums”
- TruthFinder [20]
- “Investment” and variants [10].
- Probabilistic models [21, 22]

4. Extensions to basic models, adding structure [60 min]

- Incorporating prior knowledge into models, Generalized fact-finders [12, 11]
- Incorporating context: Evidence-based Trust Framework [16]
- Handling source dependence [5, 4]
- Latent credibility analysis [13]

5. Presenting trustworthy information [30 min]

- External factors influence human perception of trust [15, 8].
- Presenting credible information on controversial issues [18, 19, 9]

6. Conclusion and Future Research Directions [30 min]

Tutorial Instructors

Jeff Pasternack is a research scientist at Facebook. His dissertation work with advisor Dan Roth focused on computing the trustworthiness of information via both non-probabilistic models such as fact-finders which iteratively evaluate the believability of claims given the trustworthiness of their sources (and vice versa) and, more recently, generative Latent Credibility Analysis models that, by describing how a source “decides” to assert a claim, provide a principled and highly effective method of assessing trust; making efficient use of prior knowledge has been of substantial interest in both approaches. Earlier work includes metrics for measuring the collective trustworthiness of sets of claims (such as news articles), information extraction, transliteration, and constrained learning.

Dan Roth is a Professor in the Department of Computer Science at the University of Illinois at Urbana-Champaign and the Beckman Institute of Advanced Science and Technology, and a University of Illinois Scholar. He is a fellow of AAAI and the ACM. Roth has published broadly in machine learning, natural language processing, knowledge representation and reasoning and received several paper, teaching and research awards. He has developed several machine learning based natural language processing systems that are widely used in the computational linguistics community and in industry and has presented invited talks and tutorials in several major conferences.

V.G. Vinod Vydiswaran is a post-doctoral research associate at the Information Trust Institute at the University of Illinois at Urbana-Champaign. His dissertation research focused on modeling and predicting trustworthiness on online textual information, and was advised by Prof. Dan Roth and Prof. ChengXiang Zhai. As part of his research, he addressed various challenges of computing trustworthiness of online textual information, viz. algorithmic, data-driven, and user-centric. His earlier work has included developing a textual entailment system and applying entailment techniques to relation extraction and information retrieval. His research interests include text informatics, natural language processing, machine learning, and information extraction.
References


[20] Xiaoxin Yin, Jiawei Han, and Philip S. Yu. Truth Discovery with Multiple Conflicting Information Providers on the Web. IEEE Transactions on Knowledge and Data Engineering, 20(6):796–808, 2008.
