Crowdsourcing Inclusivity: Dealing with Diversity of Opinions, Perspectives and Ambiguity in Annotated Data

The CrowdTruth Tutorial

Lora Aroyo
Google
New York, New York, USA
l.m.aroyo@gmail.com

Anca Dumitrache
Vrije Universiteit Amsterdam
Netherlands
anca.dumitrache@vu.nl

Oana Inel
Vrije Universiteit Amsterdam
Netherlands
oana.inel@vu.nl

Zoltán Szlávik
IBM Benelux CAS
Amsterdam, Netherlands
zoltan.szlavik@nl.ibm.com

Benjamin Timmermans
IBM Benelux CAS
Amsterdam, Netherlands
benjamin.timmermans@nl.ibm.com

Chris Welty
Google Research
New York, New York, USA
cawelty@gmail.com

ABSTRACT

In this tutorial, we introduce a novel crowdsourcing methodology called *CrowdTruth* [1, 9]. The central characteristic of CrowdTruth is harnessing the diversity in human interpretation to capture the wide range of opinions and perspectives, and thus provide more reliable, realistic and inclusive real-world annotated data for training and evaluating machine learning components. Unlike other methods, we do not discard dissenting votes, but incorporate them into a richer and more continuous representation of truth. CrowdTruth is a widely used crowdsourcing methodology[1] adopted by industrial partners and public organizations such as Google, IBM, New York Times, Cleveland Clinic, Crowdynews, Sound and Vision archive, Rijksmuseum, and in a multitude of domains such as AI, news, medicine, social media, cultural heritage, and social sciences. The goal of this tutorial is to introduce the audience to a *novel approach to crowdsourcing* that takes advantage of the diversity of opinions and perspectives that is inherent to the Web, as methods that deal with disagreement and diversity in crowdsourcing have become increasingly popular. Creating this more *complex notion of truth* contributes directly to the larger discussion on *how to make the Web more reliable, diverse and inclusive.*

CCS CONCEPTS

• **Human-centered computing → Empirical studies in collaborative and social computing**; • **Computing methodologies** → *Information extraction*; *Language resources*; *Lexical semantics.*

KEYWORDS

Crowdsourcing; Ambiguity; Perspectives; Diversity; Inter-annotator Disagreement; Ground Truth; Computational Social Sciences; Digital Humanities; Medical Text Annotation

[1] http://crowdtruth.org

WWW '19 Companion, May 13–17, 2019, San Francisco, CA, USA

ACM ISBN 978-1-4503-6675-5/19/05.
https://doi.org/10.1145/3308558.3320096

ACM Reference Format:
Lora Aroyo, Anca Dumitrache, Oana Inel, Zoltán Szlávik, Benjamin Timmermans, and Chris Welty. 2019. Crowdsourcing Inclusivity: Dealing with Diversity of Opinions, Perspectives and Ambiguity in Annotated Data: The CrowdTruth Tutorial. In *Companion Proceedings of the 2019 World Wide Web Conference (WWW'19 Companion), May 13–17, 2019, San Francisco, CA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3308558.3320096

1 INTRODUCTION

The goal of the CrowdTruth tutorial is to introduce the audience to *a novel approach to crowdsourcing that takes advantage of the diversity of opinions and perspectives that is inherent to the Web.* It is relevant to the conference themes, as methods *dealing with disagreement and diversity in crowdsourcing* are increasingly popular [2, 3]; and a more complex notion of truth contributes directly to the discussion on *how to make the Web more reliable, diverse and inclusive.*

In this tutorial, we discuss how inter-annotator disagreement appears in crowdsourcing as a result of task design, ambiguous data, and crowd annotators with varying degrees of skills and reliability. We discuss elements of visual design and user interaction encouraging diversity of opinion, as well as data science methods for aggregating and interpreting crowdsourced data that accounts for inter-annotator disagreement.

The CrowdTruth methodology [1, 8] has proven to be a valuable method for gathering ground truth data for training and evaluating information extraction methods, as well as other machine learning models that facilitate data navigation and populating the Web, in a wide array of industrial and research applications at Google, IBM, Rijksmuseum, Sound and Vision, New York Times etc. These cover also a wide range of domains, from medical domain [5], to open domain [6], to cultural heritage [4], digital humanities and computational social sciences [11], to various information retrieval [12] and natural language processing tasks [7, 10].

2 DURATION AND SESSIONS

This is a hands-on tutorial with real-world examples where we discuss challenges, limitations, opportunities and open issues in an interactive manner. We provide a guided hands-on experience with Jupyter notebooks and actual crowdsourcing templates in FigureEight[2], to enhance the learning process during the tutorial.

The tutorial material consists of presentation slides, hands-on material (Jupyter notebooks[3] and crowdsourcing templates) and relevant bibliography and is available prior, during and after the tutorial on Github[4] (CC-BY). The hands-on exercises are prepared with data gathered from experiments in real-world use cases, e.g. medical text, news text, news video and social media reviews.

The program contains four sessions dedicated to different parts of the CrowdTruth methodology:

- **Introduction:** Here we introduce crowdsourcing for gathering semantic interpretation. We describe various crowdsourcing modalities, settings and platforms in order to provide an overview of existing capabilities. Furthermore, we focus on explaining the limitations and myths of the current practices to gathering ground truth through crowdsourcing.
- **Task Design:** Here we explain how to design an open-ended or a closed task in order to gather the full range of opinions and perspectives, as well as the present ambiguity in the data. We provide a variety of examples for different data modalities such as text, image and video.
- **Data Processing:** In the CrowdTruth methodology the crowd contributor annotations are translated into a vector space representation, which is called the annotation vector. In this session, we provide examples of different annotation vectors for both open-ended and closed tasks and explain how the data collected is processed into a vector space, where the opinions and perspectives are projected, and what kind of data transformations can be applied to optimally represent in this space different data types and modalities.
- **Disagreement-aware Metrics:** We present the CrowdTruth approach for crowdsourcing ground truth data that uses disagreement-aware metrics to capture the ambiguity and the diversity of opinions and perspectives inherent in semantic interpretation. Here we focus on the quality metrics to evaluate the media units, the annotations and the crowd workers. The hands-on material provide a closer look at the data and methodology presented.

3 PREVIOUS EDITIONS AND ORGANIZATION

The tutorial organizers are long-term collaborators in the context of the **CrowdTruth development and research**[5]. The first edition of this tutorial was presented at the International Semantic Web Conference 2018[6], Monterey, California. The material and the schedule of the tutorial can be found here[7]. In the second edition, we bring together most illustrative application examples and lessons learned with practical experiences in data collection and analysis. Since 2014 different versions of this tutorial have been presented:

- as part of the **MSc curriculum** for computer science, artificial intelligence, business and information science[8] as well as in the **Digital Humanities minor** at the VU University Amsterdam. In 2016, the course was awarded the **ICT-Project of the year in education** in the Netherlands.
- in **master classes for professionals**[9] at the VU University Amsterdam and IBM Benelux,
- as a core module in the Big Data in Society Summer School held at the VU University Amsterdam, in 2015[10] and 2016[11].
- at the **Data Science with Humans in the Loop symposium**[12], organized by the Human Computation Community in the Netherlands (HComp-NL).

4 AUDIENCE

The tutorial is intended for a broad range of participants - researchers and practitioners in different domains of computer science, social science and humanities - interested in the topics of crowdsourcing, data annotation and ground truth creation. It is suitable for all levels of knowledge in human computation. Participants that are already experienced with crowdsourcing can apply CrowdTruth on their own crowdsourced datasets.

REFERENCES

[1] Lora Aroyo and Chris Welty. 2014. The Three Sides of CrowdTruth. *Journal of Human Computation* 1 (2014), 31–34. Issue 1. https://doi.org/10.15346/hc.v1i1.3

[2] Joseph Chee Chang, Saleema Amershi, and Ece Kamar. 2017. Revolt: Collaborative Crowdsourcing for Labeling Machine Learning Datasets. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 2334–2346. https://doi.org/10.1145/3025453.3026044

[3] Alessandro Checco, Kevin Roitero, Eddy Maddalena, Stefano Mizzaro, and Gianluca Demartini. 2017. Let's Agree to Disagree: Fixing Agreement Measures for Crowdsourcing. , 11–20 pages. http://eprints.whiterose.ac.uk/122865/ © 2017, Association for the Advancement of Artificial Intelligence (www.aaai.org).

[4] Victor De Boer, Johan Oomen, Oana Inel, Lora Aroyo, Elco Van Staveren, Werner Helmich, and Dennis De Beurs. 2015. DIVE in the Event-Based Browsing of Linked Historical Media. *Web Semantics: Science, Services and Agents on WWW* 35 (2015), 152–158. http://www.websemanticsjournal.org/index.php/ps/article/view/427

[5] Anca Dumitrache, Lora Aroyo, and Chris Welty. 2017. Crowdsourcing Ground Truth for Medical Relation Extraction. *ACM Trans. Interact. Intell. Syst., Special Issue on Human-Centered Machine Learning (in publication)* 8, 2 (2017), 12. http://arxiv.org/abs/1701.02185

[6] Anca Dumitrache, Lora Aroyo, and Chris Welty. 2017. False positive and cross-relation signals in distant supervision data. *arXiv preprint arXiv:1711.05186* (2017).

[7] Anca Dumitrache, Lora Aroyo, and Chris Welty. 2018. Capturing ambiguity in crowdsourcing frame disambiguation. In *Sixth AAAI Conference on Human Computation and Crowdsourcing*.

[8] Anca Dumitrache, Oana Inel, Lora Aroyo, Benjamin Timmermans, and Chris Welty. 2018. CrowdTruth 2.0: Quality Metrics for Crowdsourcing with Disagreement. (2018). https://arxiv.org/abs/1808.06080

[9] Anca Dumitrache, Oana Inel, Benjamin Timmermans, Carlos Ortiz, Robert-Jan Sips, and Lora Aroyo. 2017. Empirical Methodology for Crowdsourcing Ground Truth. *Semantic Web Journal, Special Issue on Human Computation and Crowdsourcing (in review)* (2017). http://www.semantic-web-journal.net/content/empirical-methodology-crowdsourcing-ground-truth-0

[10] Oana Inel and Lora Aroyo. 2017. Harnessing diversity in crowds and machines for better ner performance. In *European Semantic Web Conference*. Springer, 289–304.

[11] Oana Inel, Tommaso Caselli, and Lora Aroyo. 2016. Crowdsourcing Salient Information from News and Tweets.. In *LREC*. European Language Resources Association (ELRA), 3959–3966.

[12] Oana Inel, Giannis Haralabopoulos, Dan Li, Christophe Van Gysel, ZoltÃąn Szlávik, Elena Simperl, Evangelos Kanoulas, and Lora Aroyo. 2018. Studying Topical Relevance with Evidence-based Crowdsourcing. In *CIKM*. ACM, 1253–1262.

[2] https://www.figure-eight.com

[3] http://data.crowdtruth.org

[4] https://github.com/CrowdTruth/CrowdTruth-core

[5] http://crowdtruth.org/team/

[6] http://iswc2018.semanticweb.org

[7] crowdtruth.org/tutorial

[8] http://crowdtruth.org/course/watson-innovation-course-2016/

[9] http://btimmermans.com/2017/10/05/ibm-watson-masterclasses-with-vu-amsterdam-and-tu-delft/

[10] http://crowdtruth.org/course/big-data-summerschool-2015/

[11] http://crowdtruth.org/course/big-data-summerschool-2016/

[12] http://amsterdamdatascience.nl/event/data-science-with-humans-in-the-loop/