

Avoiding Fool's Gold: Mining Rich Annotations from the Crowd

Stephen Purpura (sp559@cornell.edu)
Information Science, Cornell University

Punchline

To classify, one seeks to exclude alternative interpretations. This project investigates the use of diverse viewpoints as data for machine learning.

Introduction: "You Get What You Measure"

Researchers struggle with constructing a view of reality through annotation and classification. Inter-rater reliability is frequently the primary measure of annotation team success, and research teams will iterate on annotation schemes and training to produce a program for annotators to follow.

Realizing that each seemingly theory driven annotation rule added to the annotation process is, in itself, a source of potential over-fitting, this poster examines a new theory of annotation practice that seeks to maximize annotation utility through joint maximization of validity and reliability.

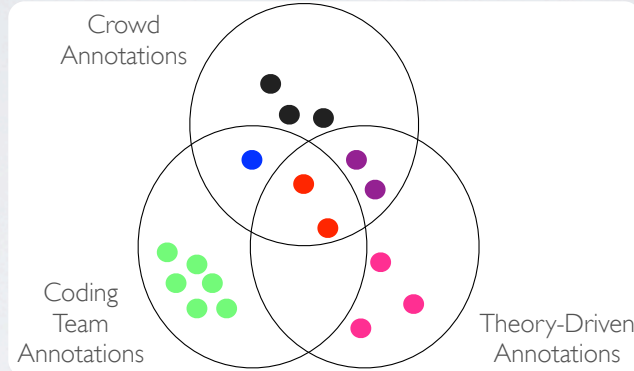
Observing the Problem in Language Use

Tweet: "OMG!! my apartment smells like bacon!!!!"

To the person writing the Tweet, is this an expression of positive emotion or negative emotion? How about to the person reading the Tweet? In our work, we've found different reactions based on whether a person is a vegetarian.

Tweet: "Cherry Cherry Boom Boom :)"

In this example, many adults will simply not know what the writer is trying to communicate. This is an example of **community language**. While this example has several community uses, one of the uses is to express admiration, appreciation, and excitement about the "Haus of Gaga" (a group that works with the popular singer Lady Gaga).



Optimizing Different Objective Functions

Theory-Driven Annotations

Theory-driven annotations are collected from "experts". These annotations are created using established methods.

Coding Team Annotations

Experts train annotation teams. But there is a limit to their ability to train. **Reliable annotations are collected from the coding team frequently by discarding alternative interpretations as invalid.** This process has an intended effect of maximizing reliability.

Crowd Annotations

Crowd annotations are collected from a representative sample of expected users. In search, this might be done through query-log analysis. For a political poll, this might be a true representative sample. You can think of the crowd as an approximation of target users.

What if Annotation Embraced Confusion?

One of the uses of data mining is to assist in finding patterns in data. Experts may have theories and models, but, when experts exclude multiple interpretations of data, they may actually be discarding key evidence that can be mined to improve theories and models. Instead, this projects wants all the sordid data!

Tell us your tawdry tales! We want the examples where you just can't train your coding team to recognize a concept. Where the expert labels don't match the crowd labels. We want each iteration of the coding scheme and the mistakes made by each of the annotators.

Reliability as a Machine Learning Feature

Using the confusion -- the interpretations from the different groups -- we seek to find patterns in the different interpretations that will give a machine learning algorithm insight into using the interpretation data to present a view for the crowd, the annotation team, or the expert (based on theory).

One-to-one marketing. In this sense, sometimes the view we present will be customized to the individual. When the system knows it wants to classify for the vegetarian, it will label "OMG!!! my apartment smells like bacon!!!!" as negative emotion. But, for an expert model, we may want the machine to classify this Tweet as indeterminate emotion.

Additional dimensionality. We accomplish this by expanding the dimensionality of the tag or label. It may be odd to think of a Tweet as labeled on a continuum of "appreciation" but thinking about these underlying traits can provide insight into whether social science models can be used effectively.

Acknowledgements

This project is developed under the guidance of Claire Cardie, Dan Cosley, Jeff Hancock, and Phoebe Sengers. Special thanks to Carl Lagoze for contributing the name and hopefully more.