

Annotating Claims in Online Customer Reviews

Annotation Manual

Version 1.0

Shilpa Arora, Carolyn P. Rosé and Mahesh Joshi
Language Technologies Institute
School of Computer Science
Carnegie Mellon University
{shilpaa, cprose, maheshj}@cs.cmu.edu

April 6, 2009

Abstract

This manual provides guidelines for categorizing comments in online customer reviews as either a *qualified* claim or a *bald* claim. These guidelines were used to annotate a dataset of amazon product reviews ¹.

1 Introduction

Comments in a customer review are claims the reviewer makes based on his/her experience with the product. These reviews are used by other potential customers to make purchasing decisions. However, these comments are often open to interpretation and customers hesitate to rely on such *bald* claims. For example, *this camera is small* is a bald claim, since *small* unless expressed relatively or based on a poll, is difficult to interpret. Other claims such as *it is small enough to fit easily in a coat pocket* are easier to interpret and we define such claims as *qualified* claims.

¹The dataset is available at www.cs.cmu.edu/~shilpaa/datasets/opinion-claims/

2 Annotation Unit

For this annotation task, an annotation unit is a comment in an online review posting. A comment usually corresponds to a sentence but sometimes comments are not complete sentences or are long sentences that could be broken down into multiple sentences. Given a review, the user annotates each comment as *qualified* or *bald* claim. In the next section, we discuss about the data we annotated and the annotation procedure we followed. In Section 4, we provide guidelines for annotating comments in a review as qualified or bald claim, followed by some examples in Section 5.

3 Data and Annotation Procedure

We applied our annotation scheme to a collection of customer reviews from the opinion mining research group at University of Illinois, Chicago² [Hu and Liu, 2004]. This dataset is a collection of amazon’s customer reviews for 5 products : Canon G3, Nikon coolpix 4300, Nokia 6610, Creative Labs Nomad Jukebox Zen Xtra 40GB, Apex AD2600 Progressive-scan DVD player. We annotated the data for 3 out of 5 products (Nikon coolpix 4300, Nokia 6610, Apex AD2600). Each comment in the review is evaluated as being qualified or bald claim.

The data was completely double coded such that each review comment received a code from the two annotators. For a total of 1,252 review comments, the Cohen’s kappa [Cohen, 1960] agreement was 0.465. On a separate dataset (365 review comments)³, we evaluated our agreement after removing the borderline cases (only about 14%) and there was a statistically significant improvement in kappa to 0.532. Since the agreement was low, we resolved our conflict by consensus coding on the data that is made available⁴.

Some general guidelines that we followed are:

- Context: The context should be used only if the given sentence is incomplete such as when it uses coreference.
- Introducing a list: The sentence introducing the list (list header) is implicitly appended to every point in the list when labeling them. For example, in the sentence below ‘Drawbacks of this camera are’ is appended to each of the points in the list being introduced.

²<http://www.cs.uic.edu/~liub/FBS/CustomerReviewData.zip>

³from the same pool but not included in the dataset for training yet

⁴The dataset is available at www.cs.cmu.edu/~shilpaa/datasets/opinion-claims/

Drawbacks of this camera are:

1. *It has a low resolution*
2. *The buttons are not very intuitive*

4 Annotation Scheme

In this section, we define the two types of claims and guidelines that were used to distinguish between them. Examples presented are taken from the customer product review dataset we annotated.

4.1 Qualified Claim

A claim is called *qualified* if its validity or scope is limited by making the conditions of its applicability more explicit. It could be either a fact or a statement that is well-defined and attributed to some source. Qualified claims can be broadly divided into following types:

1. **Fact:** A fact about the product is well qualified. For example, *The camera comes with a lexar 16mb starter card, which stores about 10 images in fine mode at the highest resolution* is a qualified claim.
2. **Reporting an event:** A comment reporting an event related to the product is a qualified claim. For example, *I sent my camera to nikon for servicing, took them a whole 6 weeks to diagnose the problem.*
3. **Self-attributed opinion or belief:** A comment where the reviewer expresses his/her opinion or belief about the product and attributes it to himself/herself is a qualified claim. For example, *I find this to be a great feature* is well qualified.
4. **Verifiable claim:** A claim that can be verified to be true or not is a qualified claim. For example, *It did not work.* Something working or not here means whether it's functional or not which is verifiable.

4.2 Bald Claim

Bald claims are non-factual claims that are open to interpretation and thus cannot be verified. In this work, we identify following types of claims as bald claims:

1. **Unattributed opinion or belief:** A comment that implicitly expresses an opinion or belief without qualifying it with an explicit attribution is a bald claim. For example, the following comments are bald:

- *Expectation is that camera automatically figures out when to use the flash.*
- *This camera is perfect for an enthusiastic amateur photographer.*
- *Higher zoom is known to be better.*

2. **Gradable words:** Wikipedia⁵ defines semantic gradability as follows:

“Adjectives and adverbs typically have the semantic feature of being *gradable*, that is the quality or state that they describe exists on a gradual scale between two opposite poles. For example, there is a gradable scale between the antonyms cold and hot.”

Here, we discuss how gradable words are used in bald claims. These categories are inspired by the *vagueness, non-objective measurability and imprecision* distinctions made by Chen (2008).

- (a) *Not quantifiable gradable words:* Non quantifiable gradable words like ‘nice’, ‘good’, ‘better’, ‘best’ etc. usually make a claim bald, as there is no clear definition of being good, nice or better.
- (b) *Quantifiable gradable words:* Quantifiable gradable words make a claim bald in the following context:
 - i. *With no frame of reference:* Quantifiable gradable words like ‘small’, ‘long’, ‘hot’, ‘cold’ make a claim bald when used without any frame of reference. For example, a comment *this desk is small* is a bald claim whereas *this desk is smaller than what I had earlier* is a qualified claim, since the comparative *smaller* can be verified by observation or actual measurement, but whether something is *small* in general is open to interpretation.
 - ii. *Imprecise Estimates:* Estimates for quantifiable gradable words or quantities such as *a lot smaller, a few cars, many cars, much larger, top ranked* etc. are not qualified. However, estimates like *about 100 people, about 6 feet tall* are not bald even though they express imprecision, since usually the acceptable error in these estimate expressions is known and hence they can be called qualified.

3. **Exaggeration:** Wikipedia⁶ defines *exaggeration* as

“Exaggeration means the describing of something and making it more than it really is. An example of exaggeration is: *I was walking along when suddenly this enormous dog came up to me. It was as big as an elephant.* The dog may have been big, but it was certainly not as big as that.”

⁵http://en.wikipedia.org/wiki/English_grammar#Semantic_gradability

⁶<http://simple.wikipedia.org/wiki/Exaggeration>

People sometime use exaggeration in reviews. For example, *On every visit, the food has blown us away.* Other examples of exaggeration are *I have told you a million times how to do that* or *countless hours*. Exaggerations do not have a well defined scope and hence are not qualified

4. **Sarcasm:** Wikipedia⁷ defines sarcasm as follows:

“Sarcasm is a form of irony that is bitter or cutting, being intended to taunt its target.”.

For example⁸, *When you get inside and sit down at your desk, you notice that one of your co-workers is talking loudly on his phone. When he hangs up, you say, “I think you should talk a little bit louder next time – the entire office didn’t hear it.” Your co-worker apologizes.*

People sometime use sarcasm in their reviews. For example, *Yeah - this program protects your computer all right - by locking you off the internet.*

5 Examples

We now present some examples of qualified and bald claims from customer product review dataset we annotated. For each example, in brackets, we provide a reference to the category to which this claim belongs.

5.1 Qualified Claims

1. I do like this phone overall(would give it a 3.5), but unfortunatly, I am returning it. [self-attributed opinion or belief]
2. I was surprised by t-mobile’s service. [self-attributed opinion or belief]
3. It has fm radio option and I can listen to music anytime . [fact, verifiable claim]
4. The menu options appear one at a time, taking up the whole screen, and one has to scroll down one by one, or have memorized where they are in the menu order, to select them by using the number keypad. [verifiable claim]
5. Like some other reviewers, I recieved this phone using t-mobiles upgrade program. [reported event]
6. I like the automatic key lock function. [self-attributed opinion or belief]

⁷<http://en.wikipedia.org/wiki/Sarcasm>

⁸<http://people.howstuffworks.com/sarcasm.htm>

7. They piggy-back on cingular's service. [fact] (piggy-back here means fall back and its presented as an objective fact about t-mobile falling back on cingular's service in case of no signal.)
8. But, then again, the ringer can be so loud that I heard it ringing inside my office, when I was already out on the street. [verifiable claim]

5.2 Bald Claims

1. The buttons on the phone are small, even for my small fingertips, but you get used to them rather quickly. [no frame of reference]
2. I have excellent hearing but the volume level on this phone is especially quiet. [non-quantifiable gradable (*excellent*), no frame of reference (*quite*)]
3. Forget talking on it in public, you just can't hear the other person! [exaggeration]
4. Also, their t-zones, although cheap(\$ 4.99 / mo.) never works . [exaggeration]
5. Since I received the phone, I spent countless hours on the phone with customer service reps who promised t-zones would work in 24 hours [exaggeration]
6. I must have heard this about a dozen times over the span of 2 weeks, when tzones never worked. [imprecise estimate (*a dozen times*), exaggeration]
7. Thumbs down on that one. [unattributed opinion or belief]
8. The vibrate setting is loud ! [no frame of reference]
9. So loud, really, that it doesn't work terribly well as a silent ringer option. [sarcasm]
10. Which was a good thing, because I realized I had forgotten my phone ...([sarcasm] - using context from previous comment for co-reference expressed with which)

6 Conclusion

In this annotation manual, we presented guidelines for classifying claims in online customer reviews as qualified or bald. In the future, we aim to extend this annotation scheme to a more fine grained annotation, where we annotate sources, intensity etc. of a claim as in other work [Wiebe et al., 2005].

References

- [Chen, 2008] Chen Wei. 2008. *Dimensions of Subjectivity in Natural Language*. In proceedings of ACL 2008, Human Language Technologies. Short Papers.
- [Hu and Liu, 2004] Minqing Hu and Bing Liu. 2004. *Mining and Summarizing Customer Reviews*. In proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.
- [Cohen, 1960] Jacob Cohen. 1960. *A Coefficient of Agreement for Nominal Scales*. Educational and Psychological Measurement, Vol. 20, No. 1., pp. 37-46.
- [Wiebe et al., 2005] Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. *Annotating expressions of opinions and emotions in language*. Language Resources and Evaluation, volume 39, issue 2-3, pp. 165-210.