Human Computation and Crowdsourcing: Works in Progress Abstracts:
An Adjunct to the Proceedings of the Third AAAI Conference on Human Computation and Crowdsourcing

Understanding Socially Constructed Concepts Using Blogs Data

Alastair Gill
King's College London
London, U.K.
alastair.gill@kcl.ac.uk

Francisco Iacobelli
Northeastern Illinois University
Chicago, U.S.A.
f-iacobelli@neiu.edu

Abstract

In this paper we propose a methodology to understand complex concepts, and which captures aspects of the contextual—and collaboratively constructed—meaning of these concepts with considerably less effort than manual coding. We use the word "quality" as one such concept to exemplify our methodology. By using unsupervised topic models along with a small corpus of human labeled data we explore the different uses of the concept "quality" in a large number of blogs. Our methodology is validated, qualitatively, by comparing our results to previous research. Finally, we note limitations and future directions of this work.

Introduction

Meaning in communication depends upon context and is negotiated by those participating in the interaction. When a concept is contextually dependent, or just generally not well understood, traditional techniques for analysis are more difficult to apply successfully. They may be costly in terms of coders or building dictionaries for automatic analysis (Quinn et al. 2010). One such difficult concept is "quality". Although everyone 'knows' what quality is, few can give a precise definition, and many argue about what it exactly means.

Methodology

The methodology proposed in this paper consists of three stages: (1) building a corpus and extracting topics, (2) gathering people's reported perceptions of quality; and (3) comparing the results of both stages with an established automatic similarity metric to understand use of "quality" in different contexts. Stages (1) and (2) are independent from each other and therefore, not necessarily sequential.

Building the blog corpus

Blogs discussing "quality" in English were collected from Blogspot as part of a larger project from January 2000 to December 2011 (cf. Kramer and Rodden 2008). This resulted in 33,238 posts, containing over 19.7 million words. Closed class/stop words were removed before applying topic modeling. We then ran LDA on the blog corpus to extract 100 topics: This number of topics can capture nuanced notions of quality, and at the same time it is a manageable number of topics to quickly examine, interpret and assign human-readable labels. We then retained the top 100 words (topic keys) of each topic for subsequent analysis.

Response (+)	Ave. score	Freq.	Response (–)	Ave. score	Freq.
well made	3	22	broken	-3	3
excellence	3	19	very bad	-3	2
long lasting	3	15	shoddy	-2.8	5
high standard	3	13	horrible	-2.5	2
perfection	3	12	bad	-2.1	13
luxury	3	7	flimsy	-2	2
positive	3	7	poor	-1.8	11
robust	3	7	time consuming	-1.8	4
perfect	3	6	low	-1.6	5
value for money	3	6	made in china	-1.5	2

Table 1: Human rated quality words

Reported perceptions of quality

To aid interpretation of topics discussing quality, we also collected words and short phrases that relate to different aspects of quality, which were manually categorised based on (Ghylin et al. 2008).Using an online survey, we asked participants to 'write down the words and phrases that you associate with quality', both positive and negative, and evaluate them on a 7-point scale ("3 Very Positive" to "–3 Very Negative"). 342 participants gave a total of 1684 responses (mean=4.9 responses each). Table 1 shows the 10 most strongly rated positive and negative responses.Two expert judges then grouped similar items and excluded low frequency ones: In total, 395 items were grouped into categories by content analysts; differences in grouping were resolved by mutual agreement; low frequency categories were excluded. In total, 20 categories were retained (Table 2).

Comparing Perceptions and Topics

To understand which categories of quality (Table 2) related to particular topics, we computed the similarity between all the words in each of the quality categories and the top 100 words of each topic using Latent Semantic Analysis (Furnas et al. 1988), a method based on low dimension word-

Category	Items	Example category items	No. related topics
Evaluation	40	good quality, excellence, exceptional	14
Value	33	worth the money, high value, class	13
Appearance	25	ergonomic, fashionable, image	8
Detail	6	attention to detail, craftsmanship, meticulous	8
Substance	30	materials used, thickness, weight	7
Application	27	safe, practical, performance	4
Authenticity	16	pure, exclusive, originality	4
Accuracy	18	correct, well thought out, accuracy	3
Process	16	process, well produced, organised	3
Trust	7	honesty, trusted, integrity	3
Ethics	5	thoughtful, fair, respect	3
Assurance	37	calibre, audit, guarantee	2
Product	13	functional, effective, complete	2
Feeling	11	care, passion, happiness	2
Comparative	16	premium, finest, better than normal	1
Resilience	30	stands the test of time, durability, classic	0
Individual	15	skills, character, individual	0
New	10	state of the art, modern, interesting	0
Reputation	10	reputable, recommended, status	0
Reliability	5	reliability, dependable, confidence	0

Table 2: Human coded quality categories.

Domain	No. Topics	Topic Information
Countries	4	East Asia, Indonesia, Indian Subcontinent, Philippines
Personal	4	(Features of personal writing)
Drink	3	Beer, Tea, Wine
Food	3	Desserts, Cooking, Asian
Imaging	3	Cameras, Camera Features, Image Libraries
IT	3	Infrastructure, (Spam), Web
Music	3	Album, Performance, Songs
Nature	3	Animals, Environment, Pets
Scholarship	3	Science, Universities, Writing
Books	2	Comics, Novels
Business	2	Finance, Services
Crafts	2	Materials, Pens
Fashion	2	Clothes, Colour
Quality	2	Experience, Personal
Travel	2	Countryside, Cities
Audio	1	Audio
Cars	1	Cars
Film	1	Film
Home Cinema	1	Home Cinema

Table 3: Domains and constituent LDA topics

vectors, built based on co-occurrence information and singular value decomposition. The semantic space (low dimension word vector matrix) used was trained on the same blogs in order to better represent the relationships found in our data; Topics were considered to be related to a human coded quality category if their similarity score was greater than two standard deviations from the mean similarity. This to ensure a statistically significant similarity score ($p < 0.05$). The number of topics relating to each human coded category is shown in column 4 of Table 2; For example, our methodology related the topic of cameras (words in this topic include *camera,digital, canon, x, d, sony*) to the quality category of 'Value'. We also found a relation between a topic about wine (*wine, wines, red, bottle, grapes, tasting*) and the quality categories of 'Appearance' and 'Evaluation.'

Results and Discussion

Examining briefly our human rated and coded data (Tables 1 and 2), we note that the self-report data provides a large variety of quality concepts which match a number of competing theoretical and experimental constructions of quality (product-based, manufacturing-based, absolutist; Garvin 1987). This suggests, as expected, that quality is simultaneously perceived from a variety of perspectives making its understanding and measurement difficult without more concrete contexts. We therefore use our quality concepts as a way to frame quality-related blog topics, thus providing a more concrete context in which to interpret quality. In doing so, we observe several patterns in the data, that are in line with previous research on quality, across the three most popular domains (Garvin 1987). The first, topics concerned with scholarship and writing, span four topic clusters: They relate to quality in terms of attention to detail (Detail), evaluation, value for money (Value) and appearance. The second, discussions of food, describe quality in terms of three topics: excellence (Evaluation), its relation with other food (Comparative) and in terms of function (Product). Finally in discussions about imaging, we note that these topics relate quality to cost (Value), the actual usage and performance as well as with excellence and their properties (Evaluation and Substance).

Conclusions and Future Work

We have presented a methodology that attempts to grasp different dimensions of complex concepts, such as "quality", by relying on the social construction of these concepts. Blending generative topic modeling techniques with modest human effort, we were able to analyze large quantities of blogs to dissect the meanings of "quality". This method is considerably less expensive than a complete human analysis of the text, with most of the effort being put on obtaining preliminary, human perceptions of the concept, quality and later interpreting the topics that were automatically selected as related to quality. We validate our results against other research on "quality". We believe that these results are promising and provide additional insight into the meaning of complex concepts. However, we note that these results may be limited to the blog genre and specific to the concept of "quality"; future work would validate these experimentally.

References

Furnas, G. W.; Deerwester, S.; Dumais, S. T.; Landauer, T. K.; Harshman, R. A.; Streeter, L. A.; and Lochbaum, K. E. 1988. Information retrieval using a singular value decomposition model of latent semantic structure. In *Proc. 11th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 465–480.

Garvin, D. A. 1987. Competing on the Eight Dimensions of Quality. *Harvard Business Review*.

Ghylin, K. M.; Green, B. D.; Drury, C. G.; Chen, J.; Schultz, J. L.; Uggirala, A.; Abraham, J. K.; and Lawson, T. A. 2008. Clarifying the dimensions of four concepts of quality. *Theoretical Issues in Ergonomics Science* 9(1):73–94.

Kramer, A. D. I., and Rodden, K. 2008. Word usage and posting behaviors: modeling blogs with unobtrusive data collection methods. In *Proceedings of the 26th Annual SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, 1125–1128. New York, NY, USA: ACM.

Quinn, K. M.; Monroe, B. L.; Colaresi, M.; Crespin, M. H.; and Radev, D. R. 2010. How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science* 54(1):209–228.