MIS Quarterly

RESEARCH ARTICLE

EXPECTING THE UNEXPECTED: EFFECTS OF DATA COLLECTION DESIGN CHOICES ON THE QUALITY OF CROWDSOURCED USER-GENERATED CONTENT

Roman Lukyanenko
Department of Information Technologies, HEC Montréal,
Montréal, QC CANADA H3T 2A7 {roman.lukyanenko@hec.ca}

Jeffrey Parsons
Faculty of Business Administration, Memorial University of Newfoundland,
St. John's, NL CANADA A1B 3X5 {jeffreyp@mun.ca}

Yolanda F. Wiersma
Department of Biology, Memorial University of Newfoundland,
St. John's, NL CANADA A1B 3X5 {ywiersma@mun.ca}

Mahed Maddah
Sawyer Business School, Suffolk University,
Boston, MA 02108 U.S.A. {mmaddah@suffolk.edu}

Appendix A

Applicability Check Details

This appendix describes our applicability check in more detail. The purpose of the applicability check (Rosemann and Vessey 2008) was to determine whether attribute data could be transformed to a form (in this case, species level classification) useful to data consumers (in this case, biologists). We also used the applicability check to explore perceptions that biologists in a university setting held about the potential uses and usefulness of data collected using an instance-based approach (versus a class-based approach). The applicability check is discussed briefly in the main manuscript; here, we provide details about the method we used to collect data, and the feedback we received from participants.

Method

We tested the applicability of an attribute-based data collection approach to users of UGC via an interactive seminar presentation. We made the presentation as part of a seminar series in the Department of Geography at Memorial University of Newfoundland, as geography is a field in which there is considerable interest in crowdsourced UGC (referred to by geographers as volunteered geographic information) and none of the authors are affiliated with the department. We developed a questionnaire, which was distributed in paper form, on the tables where audience members sat. The questionnaire asked six open-ended questions about the perceived benefits and limitations of both instance-based versus class-based approaches, as well as about potential applications of the instance-based approach to the respondent's own research. In addition, there were two questions asking respondents to rank (on a seven-point Likert scale) their agreement with two statements, one about the relevance and applicability of the instance-based data collection approach and the other about the relevance and applicability of the experimental findings we presented. The questionnaire also included some biographical questions (gender, position, research field, highest degree obtained).

The third author introduced the topic and format of the seminar and outlined how the website NL Nature had been harnessed by the research team to examine questions about data quality in citizen science. The second author then proceeded to outline the concept of instance-based versus class-based approaches to citizen science data collection. About 20 minutes into the presentation, a slide with the question "From your perspective (please state that perspective), in what situations (if any) would each approach be useful for data collection?" was shown and the presenter opened the floor to discussion/feedback. We then proceeded to present detailed results from our suite of experiments, followed by a slide with the question "Do the results we summarized change anything about your original perceptions of class-based versus instance-based data collection?," which prompted further discussion. The third author took notes on the discussion, and the entire discussion was audio recorded.

Results

The seminar was attended by 21 people. The majority (18) were from the Geography Department. Three members of the Biology Department also attended. The audience was a mix of faculty members, graduate students, and visiting researchers. Ten people returned questionnaires.

In response to our question about the extent to which participants agreed that the instance-based data collection approach is relevant and applicable to the practice of citizen science, the mean response was 6.0 on a seven-point scale (where seven was labeled "strongly agree"). Likewise, the mean response to our question about the extent to which participants agreed that the results of the experiments we presented were relevant and applicable to the practice of citizen science was also 6.0 on a seven-point scale. *These results provide a clear picture that participants viewed the instance-based approach as potentially valuable in the context of collecting citizen science data.*

Turning to the results of the open-ended questions we asked, Table A1 summarizes the responses we received to each question.

From our perspective, the feedback from the presentation reaffirmed our findings about the advantages and limitations of instance-based data collection. Specifically, participants saw the flexibility of the instance-based approach in accommodating unanticipated data, encouraging participation by people who are not familiar with the classes of interest to the researchers, and recognizing the potential to capture nonstandard forms of data. At the same time, concerns were expressed about the need for post-processing of data to make it useful for the goals of data consumers (scientists), and the likelihood that the collected data would be messy.

Table A1. Responses to Applicability Check Questions

Question	Summary of Responses	Examples Provided by Participants
Describe potential uses or applications of instance-based data collection	• When classes cannot be predetermined • When phenomena cannot reliably be classified • When participants knowledge (including traditional knowledge) is very different from that of researchers	• In a project on toxic waste, land users noted changes in the color of flames generated by burning wood as evidence a site was toxic. Researchers had not anticipated this.
(How) Could you make use of instance-based data collection in your research?	• To obtain new insights • To collect (more) information about organisms difficult to identify visually • To integrate nontraditional sources of knowledge	• A botanist was engaged in a project to classify plants, but was unable to in many cases, only realizing afterward that, because it was fall, the plants looked different than during summer. Identifying attributes would have been much easier.
Advantages of instance-based data collection	• Collect organism-specific data • Easy to use • Greater accuracy • Capture unanticipated data • Enhance reuse • Capture citizen scientists' categories • Capture subjective data • Allow nonexperts to participate	• In a project in which fishers reported on ocean sponges, one participant indicated the IB approach would generate much richer data than "coral, other" or "sponge, unidentified" that frequently showed up in their data, due to difficulty in identifying sponge species
Limitations of instance-based data collection	• Difficulties post-processing data to a standard format • No common measures across instances • Complete dataset unlikely to be useful to an individual project • Hard to implement where classification/standardization is very important	• In a project on caribou behavior, the interest is in a single class and additional instances might be considered noise
Advantages of class-based data collection	• Uniform data requires less post-processing and supports easy analysis • Useful with "informed" citizens who know what they are collecting • Useful for specific goals/questions • Tractable quantities of data • Data is uniform	• Standardization/classification makes data more accessible and readable across the scientific community
Limitations of class-based data collection	• Categories must be known and well-defined • Low accuracy • More training required for amateurs • May miss relevant phenomena • Does not accommodate unanticipated data	• In a project working with local residents, class-based data collection imposes the view or schema of the researchers on the contributors, thereby missing potential valuable perspectives of those closest to the phenomena of interest.

Appendix B

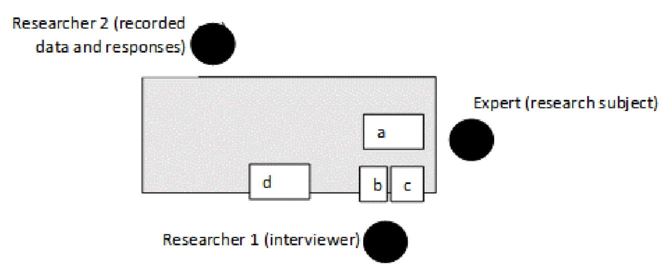
Expert and Machine Classification Study Details

Exploring Usefulness of Instance-Based Data: Classification by Experts and Machines

To answer the question whether additional processing can produce accurate classifications of attribute-based data at a level useful to data consumers, we conducted a study in which domain experts in biology were asked to infer classes based on attributes of observed organisms generated by citizen scientists. In addition, because domain expertise is a scarce resource, we also investigated the potential for classification using machine learning. We used a large dataset collected previously in a controlled laboratory setting.[1] In this previous study, we elicited free-form attributes from 390 nonexperts in biology (business students). The participants were shown images of plants and animals from the local region and were asked to describe those using attributes and classes.

To assess whether experts in natural history could accurately classify species based on attribute data, we conducted one-on-one interviews with local natural history experts having strong knowledge of the flora and fauna of the region (e.g., professors of biology at Memorial University of Newfoundland, members of the natural history society). We selected 16 organisms, a reasonable maximum for participants before risking expert fatigue (determined via a pretest). We designed the interview as a guessing game. For each organism, we had tallied the top 11 attributes provided by the nonexperts in the study in which the data were collected, along with the top most-frequent class (always a basic-level category, such as bird, tree, or fish) that had been provided by the research participants in that earlier experiment. The cutoff at 11 attributes did not affect our results, as all experts reached one of the stopping rules (see below) before the eleventh attribute.

After informed consent was obtained, the interviewer read out the instructions. To start, the interviewer presented the first sheet of the questionnaire to the natural history expert with the basic-level category and revealed only the first (most common) attribute. The expert was asked to provide a "best guess" of what the organism was. The expert was then asked to rank how well s/he could narrow down the list of possible species the organisms could be (as a measure of precision), as well as the confidence in the answer at the precision level given, based on the values given on the cards. For example, if the first attribute shown made the natural history expert think of a possible five species but s/he was unable to narrow it further, the precision was ranked "a." If the expert had sufficient attributes that matched only one species (correct or incorrect) in their mental map, then precision was coded as "e." Following the initial response, the interviewer revealed the second attribute. The experts were given the opportunity to revise their answer and were again asked to provide precision and confidence rankings. This procedure continued until one of the following stopping rules (not revealed to participants) was reached: (1) expert remained at precision level a, b, or c for three attributes; (2) expert reached precision level (d) or (e), even if incorrect; or (3) expert correctly identified the species. Once the stopping rule was achieved, the interviewer revealed the image and asked the natural history expert to identify it. This was followed by a short debriefing. Then the interviewer moved to the second organism on the questionnaire and the process continued for approximately one hour. After completing this process, participants were asked to complete a short exit questionnaire to collect biographical information. The entire interview was recorded on a digital recorder and later transcribed and a second researcher sat in on the interview (but did not speak) and took notes/recorded responses (see Figure B1 for the experimental setup).



Note: Boxes in the figure represent (a) expert questionnaire, (b) precision scale, (c) confidence scale, (d) images of plants and animals used for debriefing.

Figure B1. Set-Up for Expert Classification Experiment (Aerial View of Table and Participants)

---

[1]Our study only considers a subset of the data obtained through this experiment; the full experiment is beyond the scope of the paper and is reported elsewhere (Lukyanenko et al. 2014).

The contents of each of the sheets shown to the expert are described below. (a) is a copy of the attribute list, the example below is for a single species—there was one sheet per species; the list was initially covered with an opaque card and the attributes revealed to the expert one at a time. (b) is a copy of the "precision scale"; the expert was asked to give a rank for their answer each time a new attribute was revealed. (c) is a copy of the "confidence scale"; the expert was asked to give a rank for their answer with each guess. (d) is an envelope of 8.5 × 11" color photographs of each species that was revealed to the expert when they reached one of the stopping rules (see text).

A. Experts Questionnaire

Sample Item [Common Tern, Sterna hirundo]
[Attributes sequentially revealed]:
White; Orange beak; Black head; Orange feet; Large wings; Long wings; Long tail; Orange legs; Black top of head; Pointy beak; Grey

B. Precision Scale

Based on the attributes given thus far, could you narrow down this species to one of a possible?
More than 10 species?
5–10 species?
3 or 4 species?
2 species?
1 species?

C. Confidence Scale

Based on the attributes given thus far, how confident are you that you can identify the species to the specificity level you indicated? Use a scale from 1 to 5:
1 = Not at all confident
5 = Completely confident

D. Sample Photograph (Common tern—source: Wikimedia Commons)

We prepared questionnaires by printing a list of the 11 attributes for each organism on a single sheet of paper, as well as a sheet with precision rankings and one with confidence rankings. These were laid out on the interview table in full view of the expert participating (Figure B1).

The attributes were ordered based on the reported frequency by the nonexperts (most frequent first).[2] At the top of each page, the basic-level category for the organism was given. The order in which organisms (but not attributes) were presented was randomized across participants. We prepared 8.5x11" color photos of the same images that had been shown on screen to the nonexperts in the experiment that generated the attribute data used here, but kept these hidden from view initially (Figure B1). Each interview session lasted approximately one hour.

Responses were coded as correct (scored as 1) if they matched the common names or the genus/species. We assigned a score of 0.5 for answers containing the correct general name (e.g., "orchid" as final response for "Calypso orchid"). If the final response included two or three possible species, including the correct one, we also coded this as 0.5. Overall scores were tallied for each expert, including partial scores (thus a participant could score 9.5 out of 16 responses, for example).

The 16 natural history experts had a mean of 28 years (s.d. 20 years) of experience in natural history and 33.5 years (s.d. 17.5 years) living in the region. Self-identified areas of expertise varied, but included fish, mushrooms, birds, plants, and mammals, covering all the kinds of organisms used in the study. We further quantified expertise based on how well participants were able to identify the organism from the photograph (shown after the stopping rule was reached). When shown the photograph, our interview participants had a mean number of correct classifications of the item shown (at the genus/species-level) of 59.4% (s.d 14.7). This is considerably higher than that generally achieved by nonexperts (e.g., in our lab experiment).[3]

Participants' ability to identify organisms varied based on the attributes (see Table B1). There was a high correlation between the confidence level reported with the final guess and the percentage of times the guess was correct (Spearman's rho = 0.68, $p < 0.01$). There was also a high correlation between the precision reported with the final guess and the percentage of times the item was guessed correctly (Spearman's rho = 0.87, $p \ll 0.001$). However, even for organisms for which experts had low to no correct classification, final precision was quite high), meaning that experts could come up with a limited list (usually less than five) of species that fit the set of attributes provided. While perfect species-level identification may not always be possible based on attributes provided by nonexperts (in taxonomy the diagnostic attributes to discriminate closely related species can be cryptic), a limited list (usually of similar species) can have utility for many ecological research questions, even if the true species-level identity is unknown. The results provide strong evidence of the utility of the instance-based approach for reducing the classification uncertainty from the basic level, which is the level at which nonexperts in general can accurately classify. Our set of species encompassed a range of taxonomic groups (plants, birds, mammals) and not all natural history experts are necessarily well-versed in all taxa.

To prepare the data for machine learning (ML), we converted the attribute data set into a matrix of attributes where the attributes provided in the study were assigned 1 if a particular participant used that attribute to describe the species of interest and 0 otherwise. The resulting matrix contained 119 columns and 1,839 rows with 5,129 (2.34%) of the attributes coded as "1" and 213,712 (97.66%) coded as "0." Each row represents attributes (with associated basic-level category) provided by one of the 125 nonexpert data contributors who were asked to describe the organisms.

As expected, the attribute data are sparse, making it potentially challenging for the machine learning algorithm to discover patterns and classify species correctly based on the attributes. The sparsity of the dataset is consistent with other research findings showing that crowds generate "noisy" data (Brynjolfsson et al. 2016; Sheng et al. 2008).

To ensure accessibility of our approach to data consumers, we applied a variety of common ML algorithms, which are available in popular ML software, including neural networks, support vector machines, random forests, boosting using decision trees and naïve Bayes algorithms (Provost and Fawcett 2013). In each case, the average classification accuracy was above 70%. The top performing algorithm (Table C1) was a boosted decision tree classifier, which achieved an average F-measure (a widely-used machine learning accuracy metric) of 0.76 (± 0.12 s.d.) across 16 species (based on 10 fold cross-validation and 50 boosting iterations).

A direct comparison between human and machine performance is not meaningful since ML worked with 16 finite targets (species), whereas experts had to draw from their knowledge of all possible organisms in a local area. The results, however, suggest that, while the immediate data obtained from the instance-based approach may have low precision, it can indeed be improved by human annotation and/or applying common, off-the-shelf ML techniques.

---

[2]Note that this procedure might inflate the apparent ability of experts to identify species relative to data collected in a field setting, as there would be no way to assess attribute frequency in the latter case. However, the goal in this study is simply to assess the feasibility of using attribute data for identification, and we determined this was a reasonable way to order the attributes as it reflected the likelihood such attributes would appear in practice.

[3]A less than perfect identification accuracy by experts demonstrates the limit of expertise in broad domains. High expertise is usually narrow in scope (e.g., a single subordinate category). For example, a person who owns a collie could be considered a "collie expert," but not expert in other dog breeds (Tanaka and Taylor 1991). Yet in practice, data consumers frequently have to make decisions on broader issues (e.g., all local birds).

Table B1. Classification Accuracy of Experts and F-Measure Obtained by Machines

Species	% Correct responses (humans) corrected for expertise based on ability to identify species in photo	Mean confidence on scale 1 (low) to 5 (high) by humans	Mean precision on scale 1 (low, can think of 5–10 species) to 5 (high, can think of 1 species) by humans	F-measure – harmonic mean (boosted tree)*
American Robin	65.60	3.69	3.34	0.88
Blue Jay	78.10	3.93	4.06	0.86
Blue Winged Teal	50.00	3.07	3.33	0.85
Calypso Orchid	71.40	2.87	3.07	0.64
Caribou	85.70	4.43	4.71	0.89
Caspian Tern	42.90	3.60	3.53	0.63
Common Tern	45.50	3.53	3.20	0.70
Fireweed	0.00	3.21	2.93	0.62
Greater Yellowlegs	100.00	3.80	3.80	0.84
Indian Pipe	10.00	3.15	2.62	0.82
Lung lichen	33.30	3.27	2.82	0.80
Mallard Duck	80.80	3.60	4.27	0.85
Moose	93.30	3.93	4.47	0.91
Old Man’s Beard	100.00	3.50	3.57	0.68
Sheep laurel	0.00	3.31	2.38	0.50
Spotted sandpiper	100.00	3.64	2.86	0.71
Mean	**59.79**	**3.53**	**3.44**	**0.76**

*The results were obtained from a boosted decision tree classifier. The tree implemented a Chi-square automatic interaction detection algorithm (Geurts et al. 2006). The trees were boosted using an adaptive boosting (AdaBoost) classifier that combined the outputs of decision trees such that the trees were sequentially modified to reduce misclassification rates of the predecessor trees (Freund and Schapire 1997). The metrics were evaluated based on 10-fold cross-validation and 50 boosting iterations. AdaBoost algorithm is commonly available in such data mining software packages as RapidMiner (rapidminer.com) or SAS Enterprise Miner (sas.com).

Finally, we highlight several challenges of applying machine learning in open crowdsourcing.

First, for projects that are starting from scratch, there could be difficulties in obtaining a suitable training sample from which to build decision models. The key difficulty is that unlike applications of machine learning, where previous decisions by humans can be leveraged, in open crowdsourcing, data sets produced by crowds can be unique and equivalent data for training may be difficult or impossible to obtain. At the same time, several strategies can be used to generate a training set that approximates the actual crowdsourced data. For example, one could use a “verified” subset of the data created by experts. Another approach is to use uploaded photographs as a way to positively identify down to the desired level of precision, and then use these labels as the training set. Finally, one could also conduct a laboratory procedure as we did above by asking participants to observe sample instances and provide free-list attributes and classes, and then use of these for training. The three approaches can also be combined. There is ongoing research on learning from insufficient training data in computer science and related fields (e.g., Sommer and Paxson 2010; Webb et al. 2001; Wuest et al. 2016). Our paper provides an additional use case for these efforts, which, with more progress, can better support practical application of our ideas.

Second, we also note the difficulty in handling multiple target classes using machine learning. Typical machine learning activities involve predicting a small subset of classes (e.g., yes/no) (Provost and Fawcett 2013). However, some domains in crowdsourcing may contain hundreds or even thousands of classes that need to be learned and automatically predicted based on sparse crowdsourced UGC. Here, we anticipate that progress in machine learning algorithms will increase the ability to handle and predict large numbers of classes. We hope our paper motivates research in machine learning on improving ability to make a large number (i.e., tens, hundreds, or thousands) classification decisions at once (see Ou and Murphey 2007; Shalaginov and Franke 2016; Vincent and Hansen 2014).

We also note, however, that although a domain may contain a large number of classes, for a given task, data consumers are typically interested in a small subset of these classes (e.g., instances of wildfires in California, earthquakes in Japan, specific invasive species in Queensland). As we show, off-the-shelf machine learning techniques are quite effective at handling few (in our case 16) classification categories. Furthermore, there are established strategies for zeroing in on a single class out of many in machine learning (see Bishop 2006).

Many of the challenges related to the use of machine learning in support of the contributor-centric IQ may not be severe for all projects and should become easier to address over time.

Appendix C

Additional Use Cases for Instance-Based Data Collection

We discuss additional use cases for instance-based data collection (beyond the citizen science context used throughout the paper).

Case 1: Urban Sensing in Smart Cities

As more people are living in urban areas, research shows that human well-being in cities is dependent on the quality of urban infrastructure and municipal services (Clarkson and Kirby 2016; Hartig and Kahn 2016). The development of "smart cities" helps urban planners, managers and decision makers collect a range of environmental and human-use data related to urban life (Cardone et al. 2013; Hivon and Titah 2017; Kalay 2017; Ramaswami et al. 2016).

Sensors such as traffic counters and air, noise, and water monitoring devices can be deployed to gather data in urban spaces. In addition, smart cities increasingly benefit from *human sensors*. Unlike automated sensors, human sensors have the capacity to interpret real-world events and act upon them. In addition to reporting on typical things and events, humans can make sense of unanticipated phenomena that would get coded as "errors" or "outliers" by most electronic sensors or not captured at all. However, to take advantage of this ability, more flexible data collection may be needed.

Potential data consumers in this case include municipal police, municipal public health and safety agencies, city planners, architects, planning consultants, local businesses and organizations, and citizens themselves.

Municipalities and urban agencies increasingly develop or subscribe to platforms that collect UGC from urban sensors. For example, a popular crowdsourcing project, CitySourced.com, asks ordinary citizens to provide reports based on predefined categories, as shown in Figure C1. However, this predefined schema may be inadequate to capture all phenomena of potential interest to municipalities. Consider the example of an overturned ammonia truck on a city road. Potentially affected residents might experience and report different phenomena, "I'm stuck in traffic but I don't know why," "I saw an overturned tanker truck," "I was working in my garden and suddenly had difficulty breathing," and "I saw clouds billowing from the other side of my backyard." All of these may be manifestations of the incident. However, it would be virtually impossible to anticipate all such events in advance, or to incorporate them in the schema of an app such as CitySourced.com.

Figure C1. Categories of Urban Phenomena Provided by CitySourced.com (as of October 20, 2017)

Case 2: Disaster Management and Response

User-generated content is becoming a major source of information on natural and anthropogenic disasters. Disaster management and response applications are increasingly used by people in areas affected by natural and man-made disasters, as well as by volunteers outside the affected areas who wish to assist in recovery and restoration efforts (Goodchild 2007; Goodchild and Glennon 2010). Frequently, the observations made by ordinary people provide real-time input for disaster response and management agencies and result in more efficient and timely rescue and relief efforts (Johnson and Sieber 2012; Majchrzak and More 2011; Pultar et al. 2009).

Data consumers in this domain include governments at different levels, relief agencies, first responders, affected and interested local public and private organizations, citizens in the affected areas, as well as scientists, engineers, and planners.

In this context, timeliness, accuracy, and completeness of the report on an unfolding natural disaster are of the essence. It is, therefore, vital to ensure that the data collection processes facilitate, rather than impede, the ability of people to provide their observations. Many existing platforms designed to capture UGC related to disaster response rely on predefined categories. Consider a prominent project of this kind, the U.S. Federal Emergency Management Agency Disaster Reporter, which seeks "to crowdsource and share disaster-related information for events occurring within the United States" (https://www.fema.gov/disaster-reporter). Based on the predefined schema for the app (see Figure C2), we see that people may have difficulties reporting on anthropogenic events, as there are virtually no categories for human-made events (e.g., chemical explosion) provided. Similarly, every event requires an observer to have a positive classification, which may not be possible in all circumstances. For example, people may observe wilted plants and dried-up creeks, but may not necessarily conclude that drought is the cause, and thus may not post an otherwise valid observation. Finally, underscoring the difficulties in creating predefined classes, we note that there are multiple categories for the same event (e.g., "Home Fires" and "Floods" may be caused by "Wildfires" and "Tsunamis," respectively). This may create difficulty for observers in using appropriate categories when reporting (as well as integrating related observations).

Figure C2. Categories of Disaster Events Provided by FEMA Reporter (as of October 20, 2017)

Case 3: Product Improvement and Customer Service Based on Customer Feedback

Part of the appeal of crowdsourced UGC is finding something unexpected and new. It has been long known that front-line employees, being in direct contact with day-to-day situations, are well-equipped at spotting unusual activities, manufacturing defects, or process failures (Tax and Brown 1998; Trevino and Victor 1992; Tucker and Edmondson 2003). In a UGC setting, a notable ability of contributors is that they can report individual experiences with objects of interest to data consumers. Increasingly, companies are taking advantage of UGC to seek customer feedback on consumer product impressions, malfunctions, general usage, and suggestions for improvement or ideas for future products (Abbasi et al. 2018; Ordenes et al. 2014; Stelzer et al. 2016; Voss et al. 2004). Data consumers in this domain are mainly within the organization and include product designers and engineers, marketers, production line managers, customer service specialists, business analysts, and top executives responsible for shaping company's strategy and new product development.

Customers should be able to provide feedback by communicating their experiences as seamlessly, accurately, and completely as they would like. Having observed several existing interfaces for collecting customer product feedback, we noticed that they employ predefined categories as primary units of data collection. Although relying on predefined categories can be helpful in focusing consumers on the aspects of the products or services of most interest to the organizational data consumers, and can enable auto-directing of feedback to the unit in the organization that can act on it promptly, we expect lower accuracy and completeness (including discovering something new or unexpected) to occur in this setting.

To illustrate, Apple's Customer Feedback website (https://getsupport.apple.com) offers a selection of predefined categories for getting hardware help with an iPhone, as shown in Figure C3. While these categories cover many of the common issues, a perusal of iPhone online communities and media reports reveals many more categories not that do not fit the current predefined schema, such as a cracked camera lens, debris under the screen, camera foam misalignment, missing camera lenses, and iPhone getting hot.[4] Considering the multitude of potential interactions with an iPhone, it would be difficult, if not impossible, to predict and include every possible issue that customers might encounter. Naturally, the option to report another topic is also provided, but having the predefined choices may dissuade consumers from communicating something that does not fall into these categories.

[4]For example, https://unlockgadget.com/blog/post/computer-repair/different-types-of-phone-damage-and-ways-to-repair-phones/36.

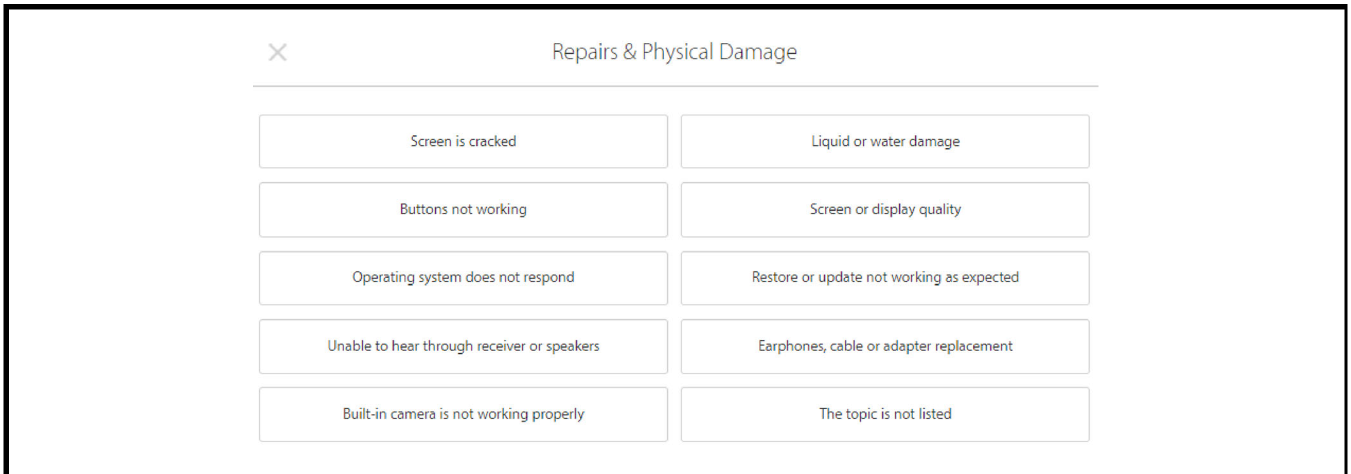


Figure C3. Categories of Hardware Issues by Apple Inc. (as of January 12, 2018)

References

Abbasi, A., Zhou, Y., Deng, S., and Zhang, P. 2018. "Text Analytics to Support Sense-Making in Social Media: A Language-Action Perspective," *MIS Quarterly* (42:2), pp. 1-38.

Bishop, C. 2006. *Pattern Recognition and Machine Learning*, New York : Springer-Verlag.

Brynjolfsson, E., Geva, T., and Reichman, S. 2016. "Crowd-Squared: Amplifying the Predictive Power of Search Trend Data," *MIS Quarterly* (40:4), pp. 941-961.

Cardone, G., Foschini, L., Bellavista, P., Corradi, A., Borcea, C., Talasila, M., and Curtmola, R. 2013. "Fostering Participation in Smart Cities: A Geo-Social Crowdsensing Platform," *IEEE Communications Magazine* (51:6), pp. 112-119.

Clarkson, B. D., and Kirby, C. L. 2016. "Ecological Restoration in Urban Environments in New Zealand," *Ecological Management & Restoration* (17:3), pp. 180-190.

Freund, Y., and Schapire, R. E. 1997. "A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting," *Journal of Computer and System Sciences* (55:1), pp. 119-139.

Geurts, P., Ernst, D., and Wehenkel, L. 2006. "Extremely Randomized Trees," *Machine Learning* (63:1), pp. 3-42.

Goodchild, M. 2007. "Citizens as Sensors: The World of Volunteered Geography," *GeoJournal* (69:4), pp. 211-221.

Goodchild, M. F., and Glennon, J. A. 2010. "Crowdsourcing Geographic Information for Disaster Response: A Research Frontier," *International Journal of Digital Earth* (3:3), pp. 231-241.

Hartig, T., and Kahn, P. H. 2016. "Living in Cities, Naturally," *Science* (352:6288), pp. 938-940.

Hivon, J., and Titah, R. 2017. "Conceptualizing Citizen Participation in Open Data Use at the City Level," *Transforming Government: People, Process and Policy* (11:1), pp. 99-118.

Johnson, P. A., and Sieber, R. E. 2012. "Situating the Adoption of VGI by Government," in *Crowdsourcing Geographic Knowledge*, D. Sui, S. Elwood, and M. Goodchild (eds.), Dordrecht, Netherlands: Springer, pp. 65-81.

Kalay, Y. E. 2017. "How Smart Is the Smart City? Assessing the Impact of ICT on Cities," in *Proceedings of the Agent Based Modelling of Urban Systems: First International Workshop*, M.-R. Namazi-Rad, L. Padgham, P. Perez, K. Nagel, and A. Bazzan (eds.), Singapore, May 10, 2016, Springer International Publishing Switzerland, pp. 189-207.

Lukyanenko, R., Parsons, J., and Wiersma, Y. 2014. "The IQ of the Crowd: Understanding and Improving Information Quality in Structured User-Generated Content," *Information Systems Research* (25:4), pp. 669-689.

Majchrzak, A., and More, P. H. B. 2011. "Emergency! Web 2.0 to the Rescue!," *Communications of the ACM* (54:4), pp. 125-132.

Ordenes, F. V., Theodoulidis, B., Burton, J., Gruber, T., and Zaki, M. 2014. "Analyzing Customer Experience Feedback Using Text Mining: A Linguistics-Based Approach," *Journal of Service Research* (17:3), pp. 278-295.

Ou, G., and Murphey, Y. L. 2007. "Multi-Class Pattern Classification Using Neural Networks," *Pattern Recognition* (40:1), pp. 4-18.

Provost, F., and Fawcett, T. 2013. *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*, Sebastopol, CA: O'Reilly Media, Inc.

Pultar, E., Raubal, M., Cova, T. J., and Goodchild, M. F. 2009. "Dynamic GIS Case Studies: Wildfire Evacuation and Volunteered Geographic Information," *Transactions in GIS* (13:s1), pp. 85-104.

Ramaswami, A., Russell, A. G., Culligan, P. J., Sharma, K. R., and Kumar, E. 2016. "Meta-Principles for Developing Smart, Sustainable, and Healthy Cities," *Science* (352:6288), pp. 940-943.

Rosemann, M., and Vessey, I. 2008. "Toward Improving the Relevance of Information Systems Research to Practice: The Role of Applicability Checks," *MIS Quarterly* (32:1), pp. 1-22.

Shalaginov, A., and Franke, K. 2016. "Towards Improvement of Multinomial Classification Accuracy of Neuro-Fuzzy for Digital Forensics Applications," in *Proceedings of the 16th International Conference on Hybrid Intelligent Systems*, A. Abraham, A. Haqiq, A. M. Alimi, G. Mezzour, N. Rokbani, and A. K. Muda (eds.), Springer International Publishing Switzerland, pp. 199-210.

Sheng, V. S., Provost, F., and Ipeirotis, P. G. 2008. "Get Another Label? Improving Data Quality and Data Mining Using Multiple, Noisy Labelers," in *Proceedings of the 14th International Conference on Knowledge Discovery and Data Mining*, New York: ACM Press, pp. 614-622.

Sommer, R., and Paxson, V. 2010. Outside the Closed World: On Using Machine Learning for Network Intrusion Detection," in *Proceedings of the IEEE Symposium on Security and Privacy*, Los Alamitos, CA: IEEE Computer Society Press, pp. 305-316.

Stelzer, A., Englert, F., Hörold, S., and Mayas, C. 2016. "Improving Service Quality in Public Transportation Systems Using Automated Customer Feedback," *Transportation Research Part E: Logistics and Transportation Review* (89), pp. 259-271.

Tax, S. S., and Brown, S. W. 1998. "Recovering and Learning from Service Failure," *MIT Sloan Management Review* (40:1), pp. 75-88.

Tanaka, J. W., and Taylor, M. 1991. "Object Categories and Expertise: Is the Basic Level in the Eye of the Beholder?," *Cognitive Psychology* (23:3), pp. 457-482.

Trevino, L. K., and Victor, B. 1992. "Peer Reporting of Unethical Behavior: A Social Context Perspective," *Academy of Management Journal* (35:1), pp. 38-64.

Tucker, A. L., and Edmondson, A. C. 2003. "Why Hospitals Don't Learn from Failures: Organizational and Psychological Dynamics That Inhibit System Change," *California Management Review* (45:2), pp. 55-72.

Vincent, M., and Hansen, N. R. 2014. "Sparse Group Lasso and High Dimensional Multinomial Classification," *Computational Statistics & Data Analysis* (71), pp. 771-786.

Voss, C. A., Roth, A. V., Rosenzweig, E. D., Blackmon, K., and Chase, R. B. 2004. "A Tale of Two Countries' Conservatism, Service Quality, and Feedback on Customer Satisfaction," *Journal of Service Research* (6:3), pp. 212-230.

Webb, G. I., Pazzani, M. J., and Billsus, D. 2001. "Machine Learning for User Modeling," *User Modeling and User-Adapted Interaction* (11:1-2), pp. 19-29.

Wuest, T., Weimer, D., Irgens, C., and Thoben, K.-D. 2016. "Machine Learning in Manufacturing: Advantages, Challenges, and Applications," *Production & Manufacturing Research* (4:1), pp. 23-45.