The Quantum Collective

Lora Aroyo*
Vrije Universiteit Amsterdam
Amsterdam, The Netherlands
lmaroyo@gmail.com

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

ABSTRACT

AI and collective intelligence systems universally suffer from a deficiency of context. There are innumerable possible contexts that may possibly change the interpretation of some signal, that may change the proper response to some stimulus. For example, an image understanding system that does not recognize an arrest event in a zoomed image of a person's face. How is it possible to know there is more information, outside of what the system can access, that affects the interpretation of data?

The solution to the context problem in practice today is a pragmatic, engineering one: analyze errors (in recommendations, question answers, image recognition, etc.), classify the kinds of contextual information that caused the wrong behavior, find the most common type of context that causes errors, and add information about that kind of context to the system. Clearly this approach is neither general nor scalable, and ignores the infamous long tail of possible contextual information that may affect a system's understanding and its behavior.

In this paper we outline a new, more general, approach to recognizing context. The approach is grounded in a fairly simple intuition: the mathematics underlying quantum mechanics is far more appropriate for modeling, and therefore simulating, human cognitive behavior than the standard toolset from classical statistics. Notions such as Heisenberg's uncertainty principle, superpositions of states, and entanglement have direct and measurable analogs in collective intelligence.

CCS CONCEPTS

• Human-centered computing → HCI design and evaluation methods; **HCI design and evaluation methods**; **HCI design and evaluation methods**; **• Theory of computation → Semantics and reasoning**; **Semantics and reasoning**; **• Information systems → World Wide Web**;

KEYWORDS

Ambiguity; Context modelling; Crowdsourcing; Collective Intelligence; Quantum mathematics; Cognitive behavior; Human Computation; Disagreement; Quality metrics

*This is the corresponding author
†This is the corresponding author

WWW'18 Companion, April 23–27, 2018, Lyon, France

ACM ISBN 978-1-4503-5640-4/18/04.
https://doi.org/10.1145/3184558.3191550

ACM Reference Format:
Lora Aroyo and Chris Welty. 2018. The Quantum Collective. In *The 2018 Web Conference Companion, April 23–27, 2018, Lyon, France.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3184558.3191550

1 INTRODUCTION

Other fields of science are questioning whether classical probability theory is the right basis for explaining and predicting complex behavior. For example, in psychology, in studies of cognition and rationality, a new trend is being set, using quantum math to explain 'irrational' (or contradicting) human behavior, such as using quantum math to model human decision-making processes [5]. These researchers do not propose that human brains are quantum computers, but rather that the mathematical principles of quantum theory can help understand and model human cognition and behaviors more accurately than probability and statistics. In economics, quantum based theories of entanglement are being used to more accurately predict complex markets.

One aspect of quantum formalisms that make them more appropriate for collective intelligence is to break the assumption that probabilities must sum to 1, a notion that perfectly models games of chance (e.g. the probabilities of a die landing on each particular side must sum to 1). The mathematical properties of all statistical laws are derived from discrete or continuous modeling of events such that the sum of all possibilities is 1. However, this is a drastic simplification of reality, because we draw statistics and model the world based on existing data, but there is always possible context which is outside that data that may affect the outcome, unlike in dice games.

For example, in a famous paper that revealed a "paradox" of human behavior, a Gallup poll was discussed that was conducted during September, 1997. Half of the 1,002 respondents were asked the following pair of questions: âĂIJDo you generally think Bill Clinton is honest and trustworthy?âĂİ and subsequently, the same question about Al Gore [7]. Half the respondents were asked their opinion of Clinton first, then Gore, the other half were asked their opinion in the other order. The results are completely different, and the experiment has been repeated many times with the same result. This problem was considered a statistical paradox until fairly recently, when quantum math was used to accurately model what was happening [8].

Intuitively, what was happening should be clear: a person's "opinion" of a politician is a complex thing, it is neither binary nor discrete; we do not simply "like" or "dislike" them, nor do we like them on a scale of 1-5. Generally we may agree with some things, disagree with others, be disgusted, excited, or ambivalent about their policies, speeches, personal lives, friends, the way they look, etc. When we are asked about one person first, it evokes a certain association that may bring to bear a different subset of these overall impressions of the other person. When the problem is boiled down to a simple "like" or "dislike" question, then people seem to believe opposite things at the same time. This is neither a paradox nor irrational (as was originally suggested), but actually a problem of modeling human opinions using a mathematical formalism devised for dice games. As it turns out, a mathematical formalism devised for describing the superposition of states works far better. We can still simplify complex situations such as human opinion to "liking" or "disliking", as long as we don't assume that they are statistical complements; in other words, as long as we allow a superposition that "collapses" when the question is answered. And of course the idea works no matter what the granularity of the simplification is: one can like and dislike "Al Gore's environmental policy" because there may be some aspect of it they like and others they dislike. Note, however, we would expect that as more explicit context is added to the decision, the less the effect of context would be.

2 APROACH

The mathematical formalisms used in quantum mechanics are significantly more complex than classical probability and statistics. We start simply, however, and provide an example that demonstrates the intuitive argument. We begin with the idea of a superposition, and in this short paper we show how a human-in-the-loop approach can be updated to provide data that informs this superposition and helps identify the effect of context.

Our basic approach to superposition is to drop the "dice game assumption" that the probabilities of an event and its apparent logical contradiction must sum to 1, to allow for this possible superposition of states. To do this, *we collect and measure the probability of an event happening separately from measuring the probability of it not happening*. That extra measurement gives us insight into the context effect.

We have been working on experiments to use quantum modeling for understanding medical natural language, understanding sounds, understanding historical events, and several others. We collect human annotations in an active learning framework that identifies the most problematic cases for human judgment, and leaves the easier cases to the machine [4], [6]. In previous work we have demonstrated our approach achieves expert quality in knowledge intensive domains like medicine [3]. In the case of understanding medical language, the state of the art in machine processing of natural language is almost exclusively one sentence at a time. It shouldn't be surprising to learn that many sentences simply can't be understood without including the previous sentences.

The problem for NLP has been that, when processing one sentence at a time, there is no way to tell the difference between a sentence that the machine *misunderstood* and a sentence that *can't be understood* without more context. Before our observations, as it is practiced today, there was no way to measure these errors or to model them. Our approach allows us to first measure how big a problem it is and second to model such sentences so that they can be recognized.

3 EXPERIMENT

In a simple and preliminary experiment, we gave a crowd of unskilled workers 88 sentences from the medical domain and asked them to select which of a set of 12 *a-priori* semantic relations was being expressed in each sentence (e.g. treats, symptom-of, causes, etc.). They also could choose "none" or "other". We gave another set of workers the same sentences but asked them to choose which of the relations was **not** expressed. In both cases, the workers were encouraged to select "all that apply". For example, in the sentence, "ANTIBIOTICS are a first-line treatment for symptoms of TYPHUS," we would expect all the workers in the positive task to select the *treats* relation, whereas for the sentence "Patients with GITELMAN SYNDROME can demonstrate normal urinary concentrating ability and have HYPOCALCIURIA", we would expect to see *symptom-of* and *associated-with* and *causes* to be selected. For the negative task, we would expect the behavior to be the inverse, i.e. that most people would *not* choose those options.

15 workers annotate each sentence, and they see the sentences with the candidate relation arguments highlighted; full details of the tasks can be found in previous work[1] [2],[1]. We normalize the results by treating the sum of all worker selections on a sentence as a vector and using cosine similarity as a comparison.

If we assume the fraction of workers that selected a particular relation to be the probability the relation is expressed in the sentence (in the positive case) or not expressed in the sentence (in the negative case), then we would expect the observed probabilities from each task, $P(R)$ and $P(\neg R)$, to sum to 1 for each sentence-relation pair. Therefore we would expect the sentence vectors from each task, which represents a distribution across the possible relations for the sentence, to have a cosine similarity of 0 (the vectors are at right angles).

In previous experiments, when fixing the number of workers per sentence at 15 (11-15 per sentence after eliminating spam), fewer than 1% of the sentence vectors changed significantly when repeating the positive task on the same datasets. But for the negative task, 28% (24/88) of the sentences had significantly different annotations from the positive task (cosine similarity > .30), and 12.5% (11/88) had the top scoring relation change. Clearly, gathering the information for the complementary task caused the workers to behave differently. Consider this sentence:

> *"Success rates were similar between treatment groups, but there were fewer documented breakthrough FUNGAL INFECTIONS in the LIPOSOMAL AMPHOTERICIN B group."*

The sentence has been pulled out of a context in which the treatment groups are explained, but one cannot correctly assume from the sentence alone that Liposomal Amphotericin–B treats fungal infections. In the positive task *treats* was the most popular single relation, with $P(Tr) = .45$, whereas in the negative task $P(\neg Tr) = .75$, and the highest score went to *associated-with*, $P(\neg AW) = .58$ compared to $P(AW) = .09$. It is noteworthy that *these probabilities, when sampled separately, do not come close to 1*, and informative to see that for the negative task, the disagreement points the vector in a different direction (the cosine similarity is .45). For the 72% of

[1] https://github.com/CrowdTruth/Medical-Relation-Extraction

sentences whose individual relation probabilities do sum to 1, the sentences are understandable without any context, like the simple example sentences given above.

As with the Gallup Poll example, what's going on here should be clear, by forcing the workers to think about the complementary task, they think about the selections slightly differently. In the original task, it is more work to select an extra relation, so workers tend towards selecting only the most obvious relation being expressed, 1.04 relations per sentence on average. In the negative task, it is more work to eliminate a relation, so they tend to be more inclusive to possible interpretations, averaging 1.44 (non negative) relations per sentence. This difference in the number of annotations obviously shows up mainly in the sentences that were more confusing, either because of missing context, as with this example, or in other cases they tend to reconsider secondary or alternate meanings.

This experiment is not intended to be conclusive, rather it offers a good example of the intuitions behind our proposed approach to consider superposition as a way of characterizing differences in human interpretation due to context.

REFERENCES

[1] Chris Welty Anca Dumitrache, Lora Aroyo. 2015. Achieving Expert-Level Annotation Quality with CrowdTruth: The Case of Medical Relation Extraction. *Proceedings of BDM2I Workshop* ISWC2015 (2015).

[2] Chris Welty Anca Dumitrache, Lora Aroyo. 2015. CrowdTruth Measures for Language Ambiguity: The Case of Medical Relation Extraction. *Proceedings of LD4IE Workshop* ISWC2015 (2015).

[3] Lora Aroyo and Chris Welty. 2014. The Three Sides of CrowdTruth. *Journal of Human Computation* 1 (2014), 31–44.

[4] Lora Aroyo and Chris Welty. 2015. Truth Is a Lie: CrowdTruth and the Seven Myths of Human Annotation. *AI Magazine* 26 (2015), 15–24.

[5] Peter D Bruza, Zheng Wang, and Jerome R Busemeyer. 2015. Quantum cognition: a new theoretical approach to psychology. *Trends in Cognitive Science* 19 (2015), 383–393.

[6] Oana Inel, Khalid Khamkham, Tatiana Cristea, Anca Dumitrache, Arne Rutjes, Jelle van der Ploeg, Lukasz Romaszko, Lora Aroyo, and Robert-Jan Sips. 2014. CrowdTruth: Machine-Human Computation Framework for Harnessing Disagreement in Gathering Annotated Data. *The Semantic Web–ISWC 2014* 33 (2014), 486–504.

[7] Zheng Wang and Jerome R Busemeyer. 2013. A quantum question order model supported by empirical tests of an a priori and precise prediction. *Trends in Cognitive Science* 5 (2013), 689–710.

[8] Zheng Wang, Tyler Solloway, Richard M Shiffrin, and Jerome R Busemeyer. 2014. Context effects produced by question orders reveal quantum nature of human judgments. *Proceedings of National Academy of Science USA* 111 (2014), 9431–9436.