Human Computation
AAAI Technical Report WS-12-08

Contextual Commonsense Knowledge Acquisition from Social Content by Crowd-Sourcing Explanations

Yen-Ling Kuo and **Jane Yung-jen Hsu**
Department of Computer Science and
Information Engineering
National Taiwan University
yjhsu@csie.ntu.edu.tw

Fuming Shih
Computer Science and Artificial Intelligence Lab
Massachusetts Institute of Technology
fuming@mit.edu

Abstract

Contextual knowledge is essential in answering questions given specific observations. While recent approaches to building commonsense knowledge bases via text mining and/or crowdsourcing are successful, contextual knowledge is largely missing. To address this gap, this paper presents *SocialExplain*, a novel approach to acquiring contextual commonsense knowledge from explanations of social content. The acquisition process is broken into two cognitively simple tasks: to identify contextual clues from the given social content, and to explain the content with the clues. An experiment was conducted to show that multiple pieces of contextual commonsense knowledge can be identified from a small number of tweets. Online users verified that 92.45% of the acquired sentences are good, and 95.92% are new sentences compared with existing crowd-sourced commonsense knowledge bases.

Introduction

Some commonsense knowledge is true only in specific contexts. Incorporating contextual commonsense knowledge into applications can help interprete the observed data for improved reasoning. For example, an indoor robot can recognize a room in the house based on objects observed in the scene using its domain knowledge about indoor locations (Gupta and Kochenderfer 2004). Similarly, an energy-saving agent can detect non-essential appliances from observed user activities and appliance states by reasoning with commonsense knowledge about power consuming activites (Lee et al. 2012).

Commonsense knowledge in Cyc (Lenat 1995) was laboriously encoded by knowledge engineers with rigorous treatment of contexts. In contrast, alternative approaches to acquiring commonsense knowledge using text mining (Schubert 2002) and crowd-sourcing (Singh et al. 2002) are effective and efficient, but they do not fare well in capturing contextual knowledge.

Both *object-location* and *appliance-activity* associations are specifically collected for the tasks at hand. Such contextual knowledge enables one to answer "kitchen" to the question "I see a refrigerator and a sink. Where am I?" correctly. However, in order to answer a similar question "I see a Giants fan and a football helmet. Where am I?" with equal success, it is necessary to acquire contextual knowledge about the sports domain. Existing text mining and crowd-sourcing techniques, e.g. mining high frequency terms from web corpus (Etzioni et al. 2004), coupled semi-supervised learning (Carlson et al. 2010), Wikipedia-style voluntary contribution (Singh et al. 2002), or human computation games (von Ahn, Kedia, and Blum 2006; Kuo et al. 2009) are domain-independent.

This research aims to harvest the associations between observations and knowledge used by readers in "contextualizing" the stream of content on social media. For example, given tweets "Going straight from the lab to The Garden tonight!" followed by "What a great comeback, Celtics rock!", one may postulate that the tweets are from "a basketball-loving college student living in Boston" by reasoning with observations, e.g. "lab", "Garden," and "Celtics", and commonsense knowledge such as "college students work in the lab," "The Celtics are an NBA team," and "The Garden is the home arena for the Boston Celtics." The key idea is to elicit contextual knowledge in the process of explaining the common interpretation of the given social content.

This paper proposes *SocialExplain*, a human computation algorithm that takes a small collection of social content as the input observations and returns a set of contextual knowledge in an attempt to interpret the content. In what follows, we will start by reviewing some related crowdsourcing techniques. The proposed contextual knowledge acquisition process, *SocialExplain*, is defined as two cognitively simple tasks for human contributors: to identify contextual clues from the given social content, and to explain the content by verifying the association of the clues and specific concepts observed. We then present the experiment conducted on Amazon's Mechanical Turk to show that multiple sentences of contextual commonsense knowledge can be acquired from updates by selected Twitter users. The explanations collected extend the OMCS knowledge network (Liu and Singh 2004) in nine domains. Online users verified that 92.45% of the acquired sentences are good, and 95.92% are new sentences compared with existing crowd-sourced commonsense knowledge bases.

Related Work

Human computation exploits the productivity of online users to solve problems that are simple to humans yet extremely difficult to computers. With the access of microtask market such as Amazon's Mechanical Turk, researchers are able to develop algorithms that involve humans in the loop (Parameswaran et al. 2011; Kittur et al. 2011; Law and Zhang 2011).

Crowd-assisted knowledge acquisition

Since natural language is more difficult to parse than structured data, crowdsourcing techniques are chosen by many researchers to acquire commonsense knowledge. The MIT Open Mind Common Sense (OMCS) project harvests commonsense via direct contribution of sentences from voluntary web users (Singh et al. 2002). Verbosity (von Ahn, Kedia, and Blum 2006) and Virtual Pets (Kuo et al. 2009) are games to collect and validate commonsense sentences. However, previous research found that unguided crowdsourcing suffers from high redundancy and small gains in new knowledge (Chklovski and Gil 2005; Kuo and Hsu 2010).

In order to sustainably build knowledge bases, systems have been proposed to incorporate machine intelligence into wisdom of the crowd (Maher and Fisher 2012; Chang, Kuo, and Hsu 2011). The systems utilize human cognitive ability to produce content and machine intelligence, e.g. topic modeling, to synthesize across human contributions. SocialExplain takes a similar approach to combine machine-driven and human-driven operations to elicit contextual knowledge from social content.

Crowd-source explanations

Interpreting datasets is fundamentally a human cognitive process. Recently, researchers try to leverage this human capability by crowd-sourcing explanations. Strategies are already developed to ask people for explanations to analyze charts (Willett, Heer, and Agrawala 2012) or interpret machine learning results (Hutton 2012).

According to the study of social psychology, people tend to apply lots of shared beliefs, i.e. commonsense knowledge, when they make explanations of the data they observe (McGarty, Yzerbyt, and Spears 2002). Instead of directly using the collected explanations as output, SocialExplain decomposes human explanation process into several operations to captures these shared beliefs.

SocialExplain

With the abundant social media content, e.g. tweets and Facebook status updates, the concepts embeded in these daily texts serve as good sources of observations to associate with commonsense knowledge. For example, if a person mentions football game many times in his status, we can easily know that he may be a *sports fan* and may go to *football games* or *watch sports games* on TV in his spare time. To better solicit contextual knowledge from people's explanations of social content, we introduce *SocialExplain* to help people frame their explanations. Based on these explanations, several domains of commonsense knowledge are populated from these observed concepts.

Task definition

The SocialExplain algorithm takes a piece of social content D, e.g. updates of status, bookmarks, or blog post of a user on social media website, to generate a commonsense network G using the explanations people provided to describe the content. A commonsense network consists of sentences that are true given a set of concepts C_o identified from social content D. Figure 1 depicts the SocialExplain workflow,

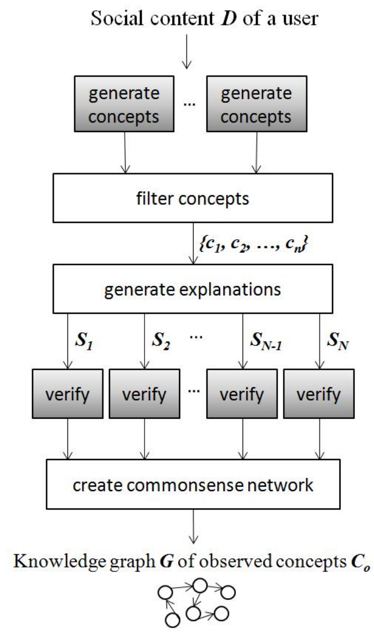

Figure 1: SocialExplain

showing the human-driven and machine-driven operations in grey and white boxes respectively, including:

- **generate concepts**: given a set of social content D, this operation generates concepts C_n which are concepts (1) people can directly observe from D and (2) people infer from D. Both kinds of concepts are contextual cues people use to understand the content.
- **filter concepts**: given generated concepts $C_1, ..., C_K$ provided by K workers, this operation aggregates the concepts and returns a set of high confidence concepts $c_1, ..., c_n$ for making explanations.
- **generate explanations**: given n concepts, this operation fills concepts in predefined explanation templates and returns all possible combinations of explanations.
- **verify explanations**: given social content D and a subset S_i of all possible explanations, this operation returns a verification vector $\vec{v} = \{0, 1\}^{|S_i|}$ where bit j indicates whether the jth sentence in S_i is a good explanation to

describe the user who generates D. The S_i in this operation consists of 20 randomly selected explanations from all possible explanations.

- **create commonsense network**: given a set of verification vectors $\vec{v_1}, ..., \vec{v_n}$ provided by n workers, this operation aggregates the votes, decomposes explanations to commonsense sentences, and returns commonsense network G that is associated with a set of observed concepts C_o where $C_o = \{c_i | c_i \in \{c_1, ..., c_n\}$ and c_i can be found in $D\}$.

The human-driven operations, "generate concepts" and "verify explanations", are associated with small HITs (human intelligence tasks) that are distributed to workers on Amazon's Mechanical Turk (Turkers).

Generate and filter concepts

In order to include the most related contextual cues people use to interprete social content D, we explicitly ask Turkers to generate three concepts for each of the following concept categories: *people/role* (`R`), *activity/event* (`E`), *time* (`T`), *location* (`L`), *object* (`O`), and *property* (`P`). For example, when seeing tweets like *"The NBA is a joke."* and *"UCLA Football just hired the former NFL coach... Crazy!"*, a Turker might put down *sport fan* (`R`), *watch ball games* (`E`), *weekend* (`T`), *stadium* (`L`), *football/basketball* (`O`), and *athletic* (`P`) respectively for each category. These concept categories are similar to the story representation in (Mulholland, Collins, and Zdrahal 2004) since social content is a user's social story. The questions designed to get concepts of these categories (see table 1) are also around the user who produced social content D.

The *filter* operation then aggregates the concepts generated from previous step. Each concept is associated with a concept category and a count, where the count is the number of Turkers who generate this concept. We filter out the concepts with count ≤ 1 to eliminate irrelevant concepts.

Generate and verify explanations

When a person makes explanations to describe another person, s/he is making associations between the contextual cues s/he identified. Initially, any pair of concepts output from previous operation can be connected in commonsense network. However, it is not true in every case. For example, it is not possible to "play football" in "living room".

Research in cognitive science shows that explanations are used to constrain inferences and guide concept generalizations (Tania and Lombrozo 2006). Accordingly, if people can generate explanations for chosen concepts, the relations embeded in explanations are considered *valid* or *reasonable* than any other arbitrary combination of concepts. To lessen the effort of producing useful explanations, SocialExplain seperates this process to two operations: *generate explanations* and *verify explanations*.

In *generate explanations* operation, the system generates possible explanations by filling concepts into predefined templates such as "[People/role] would be likely to appear at [Location]." These templates are rewritten from the combination of relations that connect concepts in ConceptNet. In our template design (see table 2), we use *AtLocation*, *HasProperty*, *HasSubevent*, *UsedFor*, and *CapableOf* relations. We do not consider *IsA*, *PartOf*, and *MadeOf* relation because they are easily introduced from the up-to-date lexical resources, e.g. WordNet (Miller 1995). The combinations of these relations are 1-step or 2-step reasoning task for people. The simplest ones are single commonsense sentences, e.g. given the role (c_R) and then find the location (c_L) the person would appear at $(c_R, AtLocation, c_L)$. After this operation, all combinations of concepts and templated are generated to be verified.

In *verify explanations* operation, Turkers are asked to verify if a sentence is reasonable to describe the user who produces social content D. A HIT in this operation contains 20 possible explanations to be verified. Every possible explanation is first verified by three Turkers. For the sentences that are accepted by at least two Turkers, they are verified by additional five Turkers so that the system can evaluate their relevance level to the observed concepts.

Probabilistic representation of templates Each explanation template implies a probabilistic representation, e.g. $P(c_L|c_R)$, people may use for reasoning out the explanations. Table 2 shows all the explanation templates and their corresponding probabilistic representations. When Turkers verifying the explanations, especially the explanations that require 2-step reasoning, they are verifying if the combinations of concepts provide a coherent explanation (Thagard 2000). Take the template "[People/role] would be likely to (do) [Activity/event] at [Location]" for example, for the composition of $(c_R, \mathit{CapableOf}, c_E)$ and $(c_E, AtLocation, c_L)$, people would select the concept combinations that maximize $P(c_L|c_E)P(c_E|c_R)$, e.g. "student would study in library" in a HIT. This characteristic helps us guarantee the quality of collected sentences.

Create commonsense network

After verifying all the possible explanations of given social content, we apply this machine-driven operation to create the corresponding commonsense network. We first normalize each concept to its canonical form so that concepts provided by different Turkers can be matched up with others. Every template is decomposed to the relations in table 2. The explanations are flattened to commonsense triples using the concepts and the corresponding relations. For example, "*student* would be likely to use *book* when *studying*" is normalized and turned to two sentences *(book, UsedFor, study)* and *(student, CapableOf, study)*. These sentences can be used to form the basic structure of the commonsense semantic network. The nodes in this network are the concepts and their labeled edges are relations connecting two concepts.

Semantics of edges in commonsense network If there is an edge between two concepts, the corresponding sentence may be true. Since the confidence of a sentence in crowd-sourced commonsense knowledge base may not be 100%, we can simply use the counts of a sentence's appearances in the collections as the score of an edge. The score only implies the confidence of the sentence. For a triple (c_1, r, c_2), there are two interpretations for its contextual meaning: (1)

Table 1: Concept categories

Category	Question	Sample answers
(R) **People/role**	What is the character or role of this person?	manager, mom
(E) **Activity/event**	What kinds of events do you think are important to this person?	watch movie, study
(T) **Time**	Name any concept about time observed from the data.	daytime, Christmas
(L) **Location**	What places do you think he/she would usually go?	office, library
(O) **Object**	What kinds of objects do you think are important to this person?	camera, book
(P) **Property**	Name some properties (adjective) to describe this person.	professional, creative

Table 2: Explanation templates

Template	**# of reasoning steps**	**Relations**	**Probabilistic representation**
R would be likely to appear at/in L.	1	*AtLocation*(c_R, c_L)	$P(c_L\|c_R)$
R would be likely to (do) E.	1	*CapableOf*(c_R, c_E)	$P(c_E\|c_R)$
R would be likely to (do) E at L.	2	*HasSubevent*(c_L, c_E), CapableOf(c_R, c_E)	$P(c_L\|c_E)P(c_E\|c_R)$
R would be likely to (use, see) O when (doing) E.	2	*UsedFor*(c_O, c_E), *CapableOf*(c_R, c_E)	$P(c_O\|c_E)P(c_E\|c_R)$
R who is/are P would be likely to appear at/in/on L.	2	*AtLocation*(c_R, c_L), *HasProperty*(c_R, c_P)	$P(c_L\|c_R)P(c_R\|c_P)$
R who is/are P would be likely to (do) E.	2	*CapableOf*(c_R, c_E), *HasProperty*(c_R, c_P)	$P(c_E\|c_R)P(c_R\|c_P)$
R would be likely to do E during/on/at T.	2	*AtTime*(c_E, c_T), *CapableOf*(c_R, c_E)	$P(c_T\|c_E)P(c_E\|c_R)$

given concept c_1, the probability concept c_2 relates to it with relation r; and (2) given concept c_2, the probability concept c_1 relates to it with relation r. We can estimate these probabilities from the counts of commonsense sentences:

$$P(c_1, r|c_2) = \frac{count(c_1, r, c_2)}{\sum_{i \in C, j \in R} count(i, j, c_2)} \quad (1)$$

and

$$P(r, c_2|c_1) = \frac{count(c_1, r, c_2)}{\sum_{i \in C, j \in R} count(c_1, j, i)} \quad (2)$$

With the conditional probability estimation, $P(c_1|c_2, r)$ and $P(c_2|c_1, r)$, as the scores of a link, we are able to make probabilistic inference with the human-contributed sentences. A set of contextual commonsense knowledge is identified by performing random walk algorithm on the network from the observed concepts.

Evaluation

In order to verify if the proposed SocialExplain workflow can help acquire good contextual knowledge, we performed experiments using Twitter updates as our source of social content.

Experiment setup

We compile two datasets of tweets from two selected Twitter accounts. We select accounts that have at least 1,000 but less than 10,000 followers to avoid branding or marketing dialogs. For verification reason, these accounts are familiar to the authors of this paper. In addition, we filtered retweets, reply tweets and links in the tweets, so that only contents intended to share publicly as personal opinions are captured. The final datasets used in this paper contain tweets arranged in chronological order. Table 3 is a part of tweets from one account used in this experiment.

Table 3: Sample tweets from a person

(1)	Just checked Redeemer Presbyterian Church in New York.
(2)	hmmm - now I'm wearing a sports coat.
(3)	Really was a good day for a prayer walk.
(4)	Star Wars: The trailer makes me want to watch the movie.
(5)	So apparently, at least at this gym, there is a trainer whose entire world is here at 24 hour fitness. Let's see if it is even half as busy by the end of the year.
(6)	Netflix releases Android app only for HTC phones & Nexus S.
(7)	Watching Star Trek with some good friends.

Deployment on MTurk A Turker is paid 0.03 USD for *generate concepts* task and 0.01 USD for *verify explanations* task. Considering the language fluency and the cultural orientation, only the workers from the U.S. and Canada can participate these HITs. We also add ground truth questions in both HITs for Turker screening. In *generate concepts* task, we ask for the name of a particular object (e.g. movie) mentioned in the tweets. In *verify explanations* task, we put two related commonsesne sentences that are already confirmed to be true/false in the verification questions. For example, "You can find movie in a movie theater" is one of the ground truth questions in this task. Only results from Turkers who give correct answers to these ground truth questions are considered to be the output of the human-driven operations. We also add an optional feedback box at the end of each HIT to gather participants' feedback such as the difficulty of the task and how they generate concepts.

Experimental result

In our experiment, we collected 728 valid explanations and turned them to 1,257 commonsense sentences. These sentences are grouped into different sets of contextual knowledge. These commonsense sentences are verified by human labelers to evaluate their qualities. The concepts and sentences are also compared with keyword matching and the current crowd-sourced knowledge base, ConceptNet, to see that SocialExplain can capture knowledge that is not found using state-of-the-art approaches.

Association between observations and acquired sentences

The collected sentences inherently form a commonsense semantic network as described in previous section. In this part of evaluation, we used the sentences with count > 1 to construct the network. Using the probability estimation of the edges (i.e. equation 1 and 2), we performed a simple random walk from observed concepts to find the contextual knowledge of the observed concepts. The probability of a path can be calculated by

$$P((c_1^1, r^1, c_2^1), \ldots, (c_1^n, r^i, c_2^n)) \\ = P(c_o) \prod_{i=1}^{n} P(c_2^i, r^i | c_1^i)$$

, where n is the path length, c_o is an observed concept, $c_1^1 = c_o$ and $c_1^i = c_2^{(i-1)}$ for $n \geq i > 1$.

The threshold of the path probability is set to 0.1 to avoid including too general and irrelevant concepts. The resulted contextual commonsense knowledge is a connected graph. Seven different sets of contextual commonsense knowledge are identified after this process. They are "Christian", "watch sports game", "watch movie", "gym", "reading", "office/work", "social event", "family", and "dog lover".

Figure 2 shows two different sets of contextual knowledge identified from the sample tweets in table 3. Four concepts, sport, gym, trainer, and watch movie, were observed by contributors from the tweets. The contributors generated and associated the four observed concepts with other inferred concepts. For example, *offer instruction*, *gym*, and *training equipment* are connected to *trainer*. Next time, when we observe the co-occurrence of *trainer* and *gym*, we will choose this set of sentences to answer questions instead of using irrelevant knowledge like "a trainer can condition a racehorse".

Quality of the acquired sentences

We rated the sentences of those with the answer counts > 1. Each sentence is rated as either good or bad by 3 human labelers, and it is treated as a good sentence if two or more labelers rated the sentence as good. Otherwise, it is considered as a bad sentence. A sentence is also rated as either related or not related to our twitter dataset using the same process.

In our result of the 265 unique sentences, 247 of them are good (precision = 92.45%), and 218 of them are relevant to the twitter dataset (relevance = 82.26%). Compared to the commonsense sentences collected from human computation game (precision = 80% for answer count > 1 (Kuo et al. 2009)), we think that the sentences generated by SocialExplain have higher precision because the workers try to make coherent explanations, so that most people would think they are reasonable/relevant explanations of the given tweets.

Informativeness of the concepts

The concepts collected from SocialExplain are compared with the baseline methods of text mining and crowd-sourcing.

Study 1: Observed concepts versus inferred concepts Since text mining techniques only acquire concepts that already exist in text, we compare the number of observed concepts and inferred concepts in this study. Only 51 out of 518 acquired concepts were found in the twitter dataset (hit rate = 9.85%). This number shows that people can infer much more concepts that coexist in a context from only a small number of observed concepts. Furthermore, most generated concepts, e.g. Christian, big screen TV, smart phone, etc, are inferred concepts in commonsense reasoning process and cannot acquire directly from ordinary text mining techniques.

Study 2: Concepts from SocialExpain versus concepts in ConceptNet Every concept collected by SocialExplain is also checked if it exists in ConceptNet. 220 out of 528 concepts in SocialExplain were found in ConceptNet (hit rate = 41.67%). From contributors' feedbacks, we find out that they tend to avoid some concepts, e.g. people, things, man, in this experiment because they think these concepts do not provide any information even if they are the correct ones. This characteristic lessens the redundent concept problem in the crowd-sourcing of commonsense knowledge. The concepts introduced by SocialExplain are mostly new to existing crowd-sourced commonsense KBs.

Informativeness of the sentences

There are 760 unique triples in the 1,257 collected commonsense sentences. For every unique triple, we use ConceptNet API to check if the triple already exists in ConceptNet. In this experiment, 31 out of 760 unique triples are found in ConceptNet (hit rate = 4.08%). The results show that the provision of observation data in SocialExplain can guide contributors come up the sentences that would not be provided in ordinary crowd-sourcing process. Even for the top 100 commonsense sentences, only 8 sentences found in ConceptNet. Most of these sentences are contextual knowledge, such as *(UCLA alumni, CapableOf, social networking)*, *(student, HasProperty, diligent)*, and *(TV, UsedFor,*

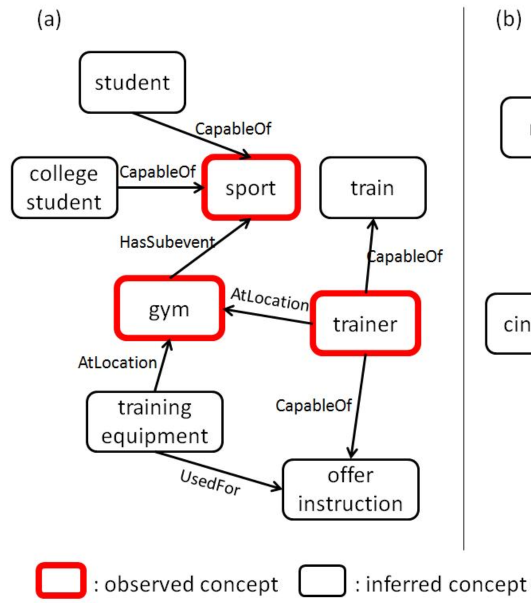

Figure 2: Contextual commonsense knowledge of different observed concepts: (a) gym (b) watch movie domain.

watch ball game), rather than the general facts of the world, e.g. *(people, CapableOf, eat)*. So, we can conclude that the collected sentences have more expected contribution to the commonsense KB than the sentences collected without any guidance.

Feedback and Discussion

To improve the process of SocialExplain, we collected feedbacks from participants and recorded notes about their task performance. In general, most participants finished the tasks within 5 minutes. Additionally, the effects of explanations have improved both the quantity and quality of the contextual knowledge collected. We found that the participants tend to avoid concepts that are too general in order to preserve the informativeness of the explanation by constraining the degree of generalization of the concepts (Patalano, Chin-Parker, and Ross 2006). Consequently, contextual knowledge obtained by SocialExplain is at a more appropriate level of granularity for reasoning about a person.

Several participants reported that the first thing they do is to group similar sentences together. To reduce such efforts, we can first apply topic modeling techniques (Blei, Ng, and Jordan 2003) to first identify possible topics within a person's tweets; those topics can represent different personas of a person. As a result, human can find relevant cues easier in grouped sentences for generating contextual knowledge in different categories.

In addition to microblogging texts, SocialExplain can also apply to other social contents such as personal profile, photo sharing, or social bookmarking to acquire different contextual knowledge from human. New challenges for crowdsourcing contextual knowledge presented by SocialExplain is "How can we pose the questions that best frame human's reasoning process for contextualizing the given observations?" The question templates used in SocialExplain provide the workers a constrained search space for identifying relevant contextual cues but preserve enough flexibility for variations in associations between concepts. As a result, we can estimate the scope of the search space and use the *verify* operation to get contextual knowledge in our experiment.

Conclusion

This paper introduced SocialExplain, a human computation algorithm for soliciting the contextual knowledge people use while reading social content. By leveraging social content as a source of observations of the world, we are able to acquire contextual knowledge of different domains. Experiments have been conducted to collect contextual knowledge using Twitter updates. The results showed that 9 domains are identified in the generated commonsense semantic network to extend the OMCS ConceptNet. In addition, the collected concepts are not easy to find in the original social content and ConceptNet (hit rate = 41.67%), and the precision of collected sentences are improved with SocialExplain (precision = 92.45%). Future work would try to use machine computation techniques to filter possible explanations, initialize probabilities of edges, and etc. The proposed approach can then be integrated with other social contents to collect a wide variety of contextual knowledge.

Acknowledgments

The authors would like to thank the anonymous reviewers for their valuable comments. This research was supported by a grant from the National Science Council of Taiwan, NSC 99-2221-E-002-139-MY3.

References

Blei, D.; Ng, A.; and Jordan, M. 2003. Latent dirichlet allocation. *The Journal of Machine Learning Research* 3:993–1022.

Carlson, A.; Betteridge, J.; Wang, R. C.; Hruschka Jr., E. R.; and Mitchell, T. M. 2010. Coupled semi-supervised learning for information extraction. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining*.

Chang, T.-H.; Kuo, Y.-L.; and Hsu, J. Y.-j. 2011. Actraversal: Graph traversal for ranking commonsense assertions and certifications. In *Proceedings of the 14th International Conference on Principles and Practice of Multi-Agent Systems (PRIMA-2011)*.

Chklovski, T., and Gil, Y. 2005. An analysis of knowledge collected from volunteer contributors. In *Proceedings of the Twentieth National Conference on Artificial Intelligence (AAAI-05)*.

Etzioni, O.; Cafarella, M.; Downey, D.; Popescu, A.-M.; Shaked, T.; Soderland, S.; Weld, D. S.; and Yates, A. 2004. Methods for domain-independent information extraction from the web: an experimental comparison. In *Proceedings of the 19th national conference on Artifical intelligence*, AAAI'04, 391–398.

Gupta, R., and Kochenderfer, M. J. 2004. Common sense data acquisition for indoor mobile robots. In *Proceedings of the 19th national conference on Artifical intelligence*, AAAI'04, 605–610. AAAI Press.

Hutton, Amanda, L.-A. M. C. 2012. Crowdsourcing evaluations of classifier interpretability. In *In Proceedings of the 2012 AAAI Spring Symposium on Wisdom of the Crowd*.

Kittur, A.; Smus, B.; Khamkar, S.; and Kraut, R. E. 2011. Crowdforge: crowdsourcing complex work. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, UIST '11, 43–52. ACM.

Kuo, Y.-L., and Hsu, J. Y.-j. 2010. Goal-oriented knowledge collection. In *2010 AAAI Fall Symposium on Common Sense Knowledge (FSS10)*.

Kuo, Y. L.; Lee, J. C.; Chiang, K. Y.; Wang, R.; Shen, E.; Chan, C. W.; and Hsu, J. Y.-j. 2009. Community-based game design: experiments on social games for commonsense data collection. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*.

Law, E., and Zhang, H. 2011. Towards large-scale collaborative planning: Answering high-level search queries using human computation. In *Proceedings of AAAI-2011*. AAAI Press.

Lee, S.-c.; Lin, G.-y.; Jih, W.-r.; Huang, C.-C.; and Hsu, J. Y.-j. 2012. Energy-aware agents for detecting nonessential appliances. *Springer Lecture Notes in Computer Science: Principles and Practice of Multi-Agent Systems* 7057:475 – 486. The 13th International Conference on Principles and Practice of Multi-Agent Systems (PRIMA-2010).

Lenat, D. B. 1995. CYC: A large-scale investment in knowledge infrastructure. *Communications of the ACM* 38(11):33–38.

Liu, H., and Singh, P. 2004. ConceptNet: A practical commonsense reasoning toolkit. *BT Technology Journal* 22(4):211–226.

Maher, M. L., and Fisher, D. H. 2012. The role of ai in wisdom of the crowds for the social knowledge on sustainability. In *In Proceedings of the 2012 AAAI Spring Symposium on Wisdom of the Crowd*.

McGarty, C.; Yzerbyt, V.; and Spears, R. 2002. *Stereotypes as explanations: the formation of meaningful beliefs about social groups*. Cambridge University Press.

Miller, G. A. 1995. WordNet: A lexical database for english. *Communications of the ACM* 38(11):39–41.

Mulholland, P.; Collins, T.; and Zdrahal, Z. 2004. Story fountain: intelligent support for story research and exploration. In *Proceedings of the 9th international conference on Intelligent user interfaces*, IUI '04, 62–69. New York, NY, USA: ACM.

Parameswaran, A.; Sarma, A. D.; Garcia-Molina, H.; Polyzotis, N.; and Widom, J. 2011. Human-assisted graph search: it's okay to ask questions. *Proceeding of the VLDB Endowment* 4(5):267–278.

Patalano, A.; Chin-Parker, S.; and Ross, B. 2006. The importance of being coherent: Category coherence, cross-classification, and reasoning. *Journal of memory and language* 54(3):407–424.

Schubert, L. 2002. Can we derive general world knowledge from texts? In *Proceedings of the 2nd International Conference on Human Language Technology Research (HLT02)*.

Singh, P.; Lin, T.; Mueller, E. T.; Lim, G.; Perkins, T.; and Zhu, W. L. 2002. Open mind common sense: Knowledge acquisition from the general public. In *On the Move to Meaningful Internet Systems, 2002 - DOA/Coop IS/ODBASE 2002 Confederated International Conferences DOA, CoopIS and ODBASE 2002*.

Tania, and Lombrozo. 2006. The structure and function of explanations. *Trends in Cognitive Sciences* 10(10):464 – 470.

Thagard, P. 2000. Probabilistic networks and explanatory coherence. *Cognitive Science Quarterly* 1:91–114.

von Ahn, L.; Kedia, M.; and Blum, M. 2006. Verbosity: A game for collecting common-sense knowledge. In *ACM Conference on Human Factors in Computing Systems (CHI Notes)*, 75–78.

Willett, W.; Heer, J.; and Agrawala, M. 2012. Strategies for crowdsourcing social data analysis. In *To Appear: ACM Conference on Human Factors in Computing Systems*.