

Commonsense Knowledge in Machine Intelligence

Niket Tandon
Allen Institute for Artificial
Intelligence
Seattle, WA
nikett@allenai.org

Aparna S. Varde
Department of CS
Montclair State University
Montclair, NJ
vardea@montclair.edu

Gerard de Melo
Department of CS
Rutgers University
Piscataway, NJ
gdm@demelo.org

ABSTRACT

There is growing conviction that the future of computing depends on our ability to exploit big data on the Web to enhance intelligent systems. This includes encyclopedic knowledge for factual details, common sense for human-like reasoning and natural language generation for smarter communication. With recent chatbots conceivably at the verge of passing the Turing Test, there are calls for more common sense oriented alternatives, e.g., the Winograd Schema Challenge. The Aristo QA system demonstrates the lack of common sense in current systems in answering fourth-grade science exam questions. On the language generation front, despite the progress in deep learning, current models are easily confused by subtle distinctions that may require linguistic common sense, e.g. *quick food* vs. *fast food*. These issues bear on tasks such as machine translation and should be addressed using common sense acquired from text. Mining common sense from massive amounts of data and applying it in intelligent systems, in several respects, appears to be the next frontier in computing. Our brief overview of the state of Commonsense Knowledge (CSK) in Machine Intelligence provides insights into CSK acquisition, CSK in natural language, applications of CSK and discussion of open issues. This paper provides a report of a tutorial at a recent conference with a brief survey of topics.

1. INTRODUCTION

Commonsense knowledge (CSK) is inherent in human cognition and behavior, yet is often too subtle for machines to acquire and use. It differs from encyclopedic knowledge, which is more factual and explicit. Clearly, modern intelligent systems can far surpass humans with respect to encyclopedic knowledge such as knowledge of named entities. For example, if we query the name of a sufficiently prominent person on a modern search engine, it will return specific details about this person, including their date and place of birth, occupation, education, significant achievements, awards, controversies, and so forth. A regular human being would

find it hard to memorize such trivia about millions of people. Still, intelligent machines lag behind humans in performing simple tasks such as distinguishing between a truck and an overpass, as observed in an incident with a semi-automated automobile in 2016. The Tesla vehicle confused the truck with an overpass due to the truck's height, leading to the loss of a human life. While human drivers may suffer from fatigue and other challenges, a responsible human driver is easily able to draw on common sense to differentiate between the two. Thus, it is important to endow machines with commonsense knowledge.

Our recent tutorial on this topic has centered on precisely this challenge. It has been presented at the ACM Conference on Information and Knowledge Management (CIKM) in Singapore in November 2017.

We start from text often found in sources such as the Web. We survey the literature on extracting commonsense knowledge from text and using the acquired knowledge to provide textual outputs useful in intelligent machines. Hence we go from text to knowledge and knowledge to text. A related issue is common sense in natural language processing. For instance, a machine translation engine should not emit *quick food* as the translation when the input was in fact referring to *fast food*. We survey such high-potential areas: CSK mining methods; CSK for smarter natural language processing; and applications towards smart computing.

The remainder of this paper is laid out as follows. Section 2 provides an overview of commonsense knowledge bases and CSK acquisition. Section 3 focuses on CSK for natural language processing. In Section 4, we discuss CSK applications and open issues in various domains, including smart cities.

2. CSK ACQUISITION

2.1 Introduction to Common Sense

Commonsense knowledge differs from encyclopedic knowledge in that it deals with general knowledge rather than the details of specific entities. Most regular knowl-

edge bases (KBs) contribute millions of facts about entities such as people or geopolitical entities, but fail to provide fundamental knowledge such as the notion that a toddler is likely too young to have a doctoral degree in physics. The challenges in acquiring CSK include its elusiveness and context-dependence. Common sense is elusive because it is scarcely and often only implicitly expressed, it is affected by reporting bias [13], and it may require considering multiple modalities. Context plays an important role for common sense in defining its correctness, and this must be accounted for while acquiring it. We partition common sense into three dimensions [21],

- (i) Common sense of objects in the environment, including properties, theories (such as physics), and associated emotions;

- (ii) Common sense of object relationships, including taxonomic, spatial and structural relationships among the objects;

- (iii) Common sense of object interactions, including actions, processes, and procedural knowledge.

Well-known projects in the commonsense KB space include hand-crafted resources such as WordNet [9] and Cyc [16], ConceptNet [14], WebChild [22], and visual KBs such as Visual Genome [8].

2.2 CSK Representation

To represent and reason over such commonsense knowledge, there are a wide range of representations that we partition across two dimensions,

- (i) **Discrete or continuous:** Discrete representation of common sense in the form of structured frames, micro-theories [16, 20] and unstructured natural language representation [1, 14] have been very popular. Recently, continuous representation based on factorization and other deep learning methods [2, 26] learns representations from large amounts of Web data [12] and generalizes better than discrete representation.

- (ii) **Multimodal:** Embedding based representation that account for textual and visual knowledge [10] can combine words and images in the same space and enable similarity computations as well as analogical reasoning.

Note that some assumptions typical in continuous representations for encyclopedic KBs may not hold for commonsense KBs. For instance, typical methods of generating negative training data for continuous representation learning do not apply equally well to CSK.

2.3 Acquiring CSK

The next level is to obtain more advanced commonsense facts, both from text and from video and other multimodal Web data. We characterize CSK acquisition across the following dimensions:

- (i) **Level of supervision:** High quality manual, text-

based commonsense KBs (CKBs) include Cyc and WordNet. Such KBs have been used extensively in various applications due to their high accuracy, but they remain costly to create and extend. Common sense acquisition through crowd-sourcing has been a well-motivated technique because commonsense games are easy for humans. ConceptNet is among the well-known crowd-sourced acquisitions, while Verbosity [24] uses visual data to drive the game experience. The main challenge facing these approaches is user engagement, because humans do not find much challenge in simple common sense based games. Automated systems have attempted to mine big data on the Web to overcome the limitations of manual and semi-automated systems. However, noisy data is a central challenge here and thus, robustness is an important dimension in the machinery. WebChild [22] is a semantically refined commonsense knowledge base mined from Web-scale textual data.

- (ii) **Modality:** NEIL [3] generates a small-scale commonsense knowledge base exclusively from visual orientation and visual features from images, by starting with seed images for a phrase and refining the senses and classifiers by clustering images discovered for the phrase. In addition to individual facts, we can also mine entailments for commonsense reasoning [1]. This aspect has also been touched upon in the tutorial.

2.4 Evaluating CSK

Being less factual in nature, evaluation of CSK is a formidable challenge. Fortunately, many different techniques have evolved to address this challenge, which can be partitioned across two dimensions.

- (i) **Intrinsic or extrinsic evaluation:** We argue that an intrinsic evaluation is most practical when judging “what usually holds” as opposed to “what can hold”. While measuring recall is typically not feasible, recent efforts have designed some proxies towards this direction [7]. More recently, intrinsic evaluation of commonsense knowledge has been automated by visual verification and detecting inconsistencies. Extrinsic evaluation indirectly measures the correctness through performance gains on an external tasks [22].

- (ii) **Manual or automated evaluation:** A number of disparate large-scale annotated challenge sets exist for measuring the impact of commonsense knowledge. These challenge datasets are either text based or visuals based, and are inference easy or hard. This includes Winograd stories [15], Aristo QA [4], and VQA [25].

Finally, we consider physical and social common sense as interesting future directions. Multimodal mining to acquire commonsense knowledge is a scalable method that overcomes the limitations of the elusiveness of CSK and visual verification and jointly leveraging disparate information sources can help overcome reporting bias.

Salient and concise KBs are helpful for quality control in KBs. In this regard, the evaluation of common-sense knowledge needs to be standardized with extrinsic datasets, to continuously track progress.

3. CSK FOR NATURAL LANGUAGE

As an example for the use of CSK in natural language, we consider the task of detecting and avoiding inappropriate collocations. A correctly collocated expression is one that a native speaker of a language such as English would typically use in good communication, e.g., *strong tea* or *red tape*. Conversely, erroneous or odd collocations include expressions that are not typically used in correct communication, e.g., *mighty tea* or *powerful tea* (instead of *strong tea*), and *crimson tape* or *scarlet tape* (instead of *red tape*). These are referred to in the literature as *collocation errors*.

Incorrectly collocated expressions can be encountered when a literal translation is conducted from a source to a target expression. If an expression does not get adequately translated, this can adversely affect communication in intelligent systems. It is thus important to fix odd collocations based on common sense. For instance, consider the expression *powerful tea* entered as a Web query. One finds that search engine results for this query contain the words *powerful* or *tea* or both. However, the user probably means to search for the availability of *strong tea*, which could further be used in an online shopping context. Upon entering the correct collocation *strong tea*, it is observed that we obtain significantly better results, including appropriate images and websites.

Linguistic classification of collocation errors: Previous work [11] has proposed a method of identifying collocation errors using association measures over syntactic patterns via a frequency based approach. CSK is captured through the writings of native speakers in KBs that serve as sources of ground truth evidence of correct collocations. Further research [6] has suggested a method of using the native or source language, i.e., the L1 language to classify collocation errors. They use annotated texts written by second language learners, incorporating corrections by professional English instructors. These serve as their sources for CSK with correctly collocated expressions. Such works address CSK-based collocations mainly from a linguistic classification perspective. As an added plus, they tangentially point towards corrective measures.

Detection and correction: Different types of collocations are addressed by Park et al. [18]. They categorize collocation errors into insertion, deletion, substitution, and transposition types. For example, substitution errors occur when a non-preferred word is used in place of a more commonly used word, e.g., *pure sky* instead of *clear sky*. Transposition errors occur when words

are used in an order different from the intended meaning, e.g., *make friendships close* instead of *make close friendships*. They develop a tool called AwkChecker to detect and correct such errors in documents.

CollOrder [23] outputs ordered responses to odd collocations by relying on semantic similarity, ranking techniques and ensemble learning. This entails error detection with POS tagging and search for matches (odd collocations) followed by error correction by searching for precise collocations, ranking, filtering and frequency ordering. Large repositories such as the British National Corpus serve as sources of CSK in the form of correct collocations. Classical rule induction [5] in the context of ensemble learning is found useful to learn similarity measures for collocation error detection and correction.

Broader impacts: Incorporating CSK into natural language processing helps us develop smarter systems in machine intelligence by providing better responses and better machine translation. Open research issues such as the challenge of sparse data (as opposed to frequent data) and literary allusion are relevant to the enhancement of CSK-based approaches such as collocation error correction. Sparsity is a challenge because many approaches in the literature rely on the frequency of expressions to assess their appropriateness. As these challenges are addressed, CSK-based natural language processing will improve and second language learners and more generally users of intelligent systems will benefit.

4. FURTHER APPLICATIONS

Apart from natural language generation, further applications using CSK include sentiment analysis, set expansion and computer vision. There are challenges in reasoning with CSK that have possible solutions and present some open issues.

Many use cases for CSK will stem from the evolution towards smart cities, e.g., [17] consider smart environment, smart mobility, smart government, smart people, smart economy and smart living as key ingredients.

For each characteristic, there are important current and future applications of CSK. For example, deployment of CSK in autonomous vehicles helps them make more well-informed decisions and hence drive better [19], thus avoiding potential accidents, e.g., Tesla crashing into a truck by confusing it with an overpass. This affects the smart mobility characteristic. Developments in CSK-based natural language processing have the potential to benefit 21st century education, thus enhancing the smart people characteristic. There are significant open issues calling for further research. For instance, the use of CSK can enhance systems such as canal lights in Amsterdam that brighten and dim based on pedestrian usage [17], to promote a cleaner environment. This calls for further research on the specifics of harnessing CSK

from given repositories to improve such systems, so as to enhance the smart environment characteristic.

5. CONCLUSIONS

We have briefly surveyed CSK acquisition and the use of repositories such as WebChild, CSK in natural language for addressing collocation issues based on linguistic classification as well as detecting and correcting collocation errors, and CSK applications in domains including smart cities. We emphasize that commonsense knowledge has made people smarter, is making machines smarter and will make smart cities smarter.

The tutorial we presented at ACM CIKM on these topics was particularly well-received. The slides for this tutorial can be found at:

<http://allenai.org/tutorials/csk>.

6. REFERENCES

- [1] Gabor Angeli and Christopher D. Manning. Naturalli: Natural logic inference for common sense reasoning. In *EMNLP*, 2014.
- [2] Antoine Bordes and Evgeniy Gabrilovich. Constructing and Mining Web-scale Knowledge Graphs. In *KDD Tutorials*, 2014.
- [3] Xinlei Chen, Abhinav Shrivastava, and Abhinav Gupta. Neil: Extracting visual knowledge from web data. In *ICCV*, 2013.
- [4] Peter Clark, Niranjan Balasubramanian, Sumithra Bhakthavatsalam, Kevin Humphreys, Jesse Kinkead, Ashish Sabharwal, and Oyvind Tafjord. Automatic construction of inference-supporting knowledge bases. In *AKBC '14*, 2014.
- [5] William Cohen. Fast effective rule induction. In *ICML*, 1995.
- [6] Daniel Dahlmeier and Hwee Ton Ng. Correcting semantic collocation errors with l1-induced paraphrases. In *EMNLP*, 2011.
- [7] Bhavana Dalvi, Niket Tandon, and Peter Clark. Domain-targeted, high precision knowledge extraction. *TACL*, 2017.
- [8] Ranjay Krishna et. al. Visual genome: Connecting language and vision using crowdsourced dense image annotations. *IJCV*, 2016.
- [9] Christiane Fellbaum, editor. *WordNet: An Electronic Lexical Database*. MIT Press, 1998.
- [10] Andrea Frome, Gregory S. Corrado, Jonathon Shlens, Samy Bengio, Jeffrey Dean, Marc'Aurelio Ranzato, and Tomas Mikolov. Devise: A deep visual-semantic embedding model. In *NIPS*, 2013.
- [11] Yoko Futagi, Paul Deane, Martin Chodorow, and Joel Tetreault. A computational approach to detecting collocation errors in the writing of non-native speakers of english. *Computer Assisted Language Learning*, 2008.
- [12] Yoshua Bengio Geoffrey Hinton. Cifar ncap - summer school 2014.
- [13] Jonathan Gordon and Benjamin Van Durme. Reporting bias and knowledge acquisition. In *AKBC*, 2013.
- [14] Hugo Liu and Pushpak Singh. ConceptNet: a practical commonsense reasoning toolkit. *BT Technology Journal*, 2004.
- [15] Quan Liu, Hui Jiang, Zhen-Hua Ling, Xiao-Dan Zhu, Si Wei, and Yu Hu. Combing context and commonsense knowledge through neural networks for solving winograd schema problems. *CoRR*, 2016.
- [16] C. Matuszek, M. Witbrock, R.C. Kahlert, J. Cabral, D. Schneider, P. Shah, and D. Lenat. Searching for common sense: Populating Cyc from the Web. In *AAAI*, 2005.
- [17] Technical University of Vienna. European smart cities: Technical report, August 2015.
- [18] Taehyun Park, Edward Lank, Pascal Poupart, and Michael Terry. Is the sky pure today - awkchecker: An assistive tool for detecting and correcting collocation errors. In *Proceedings of the ACM Symposium on User Interface Software and Technology*, 2008.
- [19] Priya Persaud, Aparna Varde, and Stefan Robila. Enhancing autonomous vehicles with commonsense: Smart mobility in smart cities. In *IEEE ICTAI workshop on Smart Cities*, 2017.
- [20] R. Schank and R. Abelson. *Scripts, plans, goals and understanding: An inquiry into human knowledge structures*. Lawrence Erlbaum Associates, Hillsdale, NJ., 1977.
- [21] Niket Tandon. Commonsense knowledge acquisition and applications. In *Doctoral dissertation*, 2016.
- [22] Niket Tandon, Gerard de Melo, Fabian Suchanek, and Gerhard Weikum. Webchild: Harvesting and organizing commonsense knowledge from the web. In *WSDM*, 2014.
- [23] Alan Varghese, Aparna Varde, Jing Peng, and Eileen Fitzpatrick. A framework for collocation error correction in web pages and text documents. *ACM SIGKDD Explorations*, 2015.
- [24] Luis von Ahn, Mihir Kedia, and Manuel Blum. Verbosity: a game for collecting common-sense facts. In *CHI*, 2006.
- [25] Peng Wang, Qi Wu, Chunhua Shen, Anton van den Hengel, and Anthony R. Dick. Fvqa: Fact-based visual question answering. *TPAMI '17*.
- [26] Bishan Yang and Tom Mitchell. Leveraging Knowledge Bases in LSTMs for Improving Machine Reading. In *ACL*, 2017.