The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

Towards Interpretation as Natural Logic Abduction

Naoya Inoue[*1] Pontus Stenetorp[*2] Sebastian Riedel[*2] Kentaro Inui[*1*3]

[*1] Tohoku University [*2] University College London
[*3] RIKEN Center for Advanced Intelligence Project

This paper studies an abductive reasoning framework on natural language representations. We show how to incorporate Natural Logic (McCartney et al. 2009) into Interpretation as Abduction (Hobbs et al. 1993). The resulting framework avoids the difficulty of natural language-formal representation conversion, while handling important linguistic expressions such as negations and quantifiers. For proof-of-concept, we demonstrate that missing, implicit premises of arguments can be recovered by the proposed framework by a manual example walkthrough.

1. Introduction

Abduction is inference to the best explanation. This paper explores abduction for natural language understanding. Given a textual observation, the task is to generate the best, textual explanation. For example, the task of finding a missing assumption of an argument [Habernal 17] is a special form of abduction. Suppose the following argument is given[*1]:

(1) *Claim: Immigration is really a problem.*
Reason: Illegal immigrants have overrun large portions of California.

Our task is to generate an additional, unstated assumption that is needed for this argument to be valid, for example, the following statement:

(2) *These portions of California are doing far worse than anywhere else in the state.*

Argument completion is useful for many applications where logic-based evaluation of texts is needed, such as automated essay scoring [Taghipour 16], and automated essay feedback. Additionally, an abductive framework in general has been considered as a promising, unified framework for text understanding [Charniak 91, Hobbs 93b].

Conventionally, abduction has been studied in the context of formal knowledge representation such as propositional logic [Santos 94], or first-order logic [Stickel 91, Hobbs 93b, Inoue 12a]. The advantage is that reasoning on formal knowledge representation provides an ability to handle information in a structured way by using special operators such as quantifiers, negations, etc. However, converting a natural language text into a formal representation, a.k.a. semantic parsing, itself is a difficult task to date [Krishnamurthy 16], which hampers applying abduction to real-world problems.

This paper explores an abductive inference framework that directly works on natural language texts, which is orthogonal to the conventional formal representation-based approaches. More specifically, we study *Natural Logic* [Maccartney 09, Bowman 13, Angeli 14] in the context of abductive reasoning. Natural Logic is a framework originally developed for entailment prediction. The advantage is that (i) it works on natural language expression directly, and (ii) it supports important linguistic expressions such as negations and quantifiers.

We show how Natural Logic-based inference can be integrated with Interpretation as Abduction [Hobbs 93a], one of the promising, abductive natural language understanding framework on formal representations (Sec. 3.). The key idea is that following the Natural Logic edit rules, we rewrite an input sentence (i.e. a textual observation) to generate its explanation, instead of applying logical inference on formal representations. We then demonstrate how this proposed framework solves an argument completion problem with a manual example walkthrough (Sec. 4.).

2. Preliminaries

2.1 Natural Logic

Given a premise p and a hypothesis h (in natural language representation), Natural Logic [Maccartney 09] provides a mechanism to determine whether p entails h or not.

Natural Logic first finds a sequence of atomic edits that transforms p into h. An atomic edit can be one of the following: (i) insertion (INS), (ii) deletion (DEL), or (iii) substitution (SUB). For example, given p = "Stimpy is a cat." and h = "Stimpy is not a dog.", the edit sequence will be the following: {SUB(cat, dog), INS(not)}. The corresponding mutated sentences are: p' = "Stimpy is a dog.", and p'' = "Stimpy is not a dog.", respectively.

To determine the relation between p and h, Natural Logic assigns a semantic operator to each edit. For example, the semantic operator of SUB(cat, dog), the first edit, is "alternation" ($|$) because cat and dog are mutually exclusive concepts. The semantic operator of INS(not) is "negation" ($\hat{}$). Finally, a semantic relation between p and each mutated sentence (e.g. p', p'') is estimated by these local semantic operators. For example, we first have $p|p'$ (by initialization). We then apply the second semantic operator (i.e. "negation") to this semantic relation, yielding "for-

Contact: Naoya Inoue, Graduate School of Information Sciences, Tohoku University, naoya-i@ecei.tohoku.ac.jp

*1 Taken from the Reasoning Comprehension Task: https://github.com/UKPLab/argument-reasoning-comprehension-task ward entailment" (i.e. $p \sqsubset p''$), according to the Natural Logic theory. The result of semantic operator application depends on the context (e.g. whether an edited word is in a negation scope or not), but it is omitted here for the brevity. See the original paper[Maccartney 09] for further details.

2.2 Interpretation as Abduction

Hobbs et al. propose *Interpretation as Abduction* (IA), the framework of text understanding based on the idea that interpreting sentences is to prove the logical form of the sentence [Hobbs 93b]. They demonstrated that a process of natural language understanding, such as word sense disambiguation or reference resolution, can be described in the single framework based on abduction.

In IA, observations are given with costs, and background axioms are given with weights. It then performs backward-reasoning on each observation, propagates its cost to the assumed literals according to the weights on the applied axioms, and merges redundancies where possible (henceforth, *factoring*). A cost of interpretation is then the sum of all the costs on elemental hypotheses in the interpretation. Finally, it chooses the lowest cost interpretation as the best interpretation.

Formally, the task of IA can be described as follows:

Given: Background knowledge B, and observations O, where B is a set of first-order logical formulae, and O is a set of literals or substitutions,

Find: $\arg\min_{H \in \mathcal{H}} Cost(H)$, where $\mathcal{H}$ is a set of hypotheses H such that $H \cup B \models O$, $H \cup B \not\models \perp$,

where H is a set of literals or substitutions and $Cost(H)$ is a cost function of hypotheses.

Let us suppose $O = r(a)^{\$20} \wedge b(a)^{\$10}$. We use backward chaining to generate new hypotheses, increasing the cost of a literal back-chained on. For example, when $\forall x(p(x)^{0.3} \wedge b(x)^{0.9} \rightarrow r(x))$ is applied, we get $p(a)^{\$6} \wedge b(a)^{\$18} \wedge b(a)^{\$10}$. The cost of this hypothesis is $\$6+\$18+\$10 = \$34$. Because we have two literals with the same content, i.e. $b(a)$, we can apply factoring. This yield new hypothesis: $p(a)^{\$6} \wedge b(a)^{\$10}$, taking the least cost among those of two literals. This factoring mechanism encourages IA to find a small explanation that explains input observations.

3. Interpretation as Natural Logic Abduction

We now describe how Natural Logic can be applied to Interpretation as Abduction (IA). The basic idea is that we replace backward chaining in IA with atomic edits in Natural Logic to perform abductive reasoning on natural language representation. This allows us to perform abductive reasoning without the difficulty of natural language-formal representation conversion, while handling important linguistic expressions such as negations and quantifiers properly.

Analogously to IA, we define observations to be cost-assigned textual observations (e.g. $O = \{$"Immigration is really a problem."$^{\$10}$, "Illegal immigrants have overrun large portions of California."$^{\$10}\}$).

To generate new hypotheses, we first initialize H to be O. Formally, let $H = \{s_1^{\$c_1}, s_2^{\$c_2}, ..., s_n^{\$c_n}\}$. We then apply the following procedure:

1. Choose a target sequence ID $i \in \{1, 2, ..., n\}$.
2. Mutate s_i by applying one of Natural Logic edits described in Sec. 2.1 (i.e. substitution, deletion, or insertion). Let the mutated sentence be s'_i.
3. Let $H' = \{..., s_{i-1}, s'_i, s_{i+1}, ...\}$.
4. Estimate a semantic relation between s_i and s'_i according to the Natural Logic theory. If $s_i \sqsubset s'_i$, let $c(s'_i) = 1.2 \cdot c(s_i)$; otherwise $c(s'_i) = c(s_i)$. Intuitively, this means that abductive reasoning has to pay a cost, but deductive reasoning does not.
5. Factor H' if possible.

For substitution in Step 2, one can use a knowledge base of lexical relations such as hypernym/hyponym, cause-effect relations (e.g. WordNet [Fellbaum 98], ConceptNet [Speer 12]). For insertion in Step 2, one can add a modifier (e.g. adjective or adverb) to a noun phrase or verb phrase in s_i according to a language model-like knowledge (e.g. World Knowledge Proposition (WKP) [Clark 09, etc.]), or adding special linguistic operators such as *every*, *not*.

By applying this procedure repeatedly, we will have a set of hypotheses (i.e. $\mathcal{H}$ in Sec. 2.2). Finally, we pick a hypothesis with the minimum cost from $\mathcal{H}$ as an output.

4. Proof-of-concept

In this section, we show a walkthrough demonstrating how Natural Logic-based IA can recover an additional, implicit assumption of an argument.

We use Example (1) for a walkthrough. We assume the following knowledge base for Natural Logic atomic edits:

- Value judgement: problem $\sqsubset$ BAD (V1)
- Agentive: immigration $\rightsquigarrow$ illegal immigrants (A1)
- Ontology: exists in $\sqsupset$ overrun (O1)
- Generic: X is bad $\sqsupset$ {X exists in Y, Y is in a bad situation} (G1)

With this knowledge base, we perform the following Natural Logic-based abductive reasoning:

1. Initialization: {"Immigration is really a problem."$^{\$10}$, "Illegal immigrants have overrun large portions of California."$^{\$10}$}
2. By V1: {"Immigration is really BAD."$^{\$10}$, "Illegal immigrants have overrun large portions of California."$^{\$10}$}

3. By A1: {"Illegal immigrants are really BAD."$^{\$10}$, "Illegal immigrants have overrun large portions of California."$^{\$10}$}.

Since substitution with an agentive lexical relation (substituting Immigration with Illegal immigration in Step 3) is not defined in the original Natural Logic theory [Maccartney 09], the semantic relation between its original sentence and the mutated sentence is not obvious. The question here is: does a statement S_1 about the thing X entail a statement S_2 about the thing Y that is created by X? S_1 is likely to entail S_2 because S_1 implicitly states that Y is involved in the statement (e.g. "*I love the guitar.*" is likely to entail "*I love music.*"). Hence, we assume that this reasoning is deductive; we will study reasoning with lexical relations not defined in the Natural Logic theory in future work.

The cost of the hypotheses unchanged in Step 1-3 because the reasoning is deductive. The reasoning continues.

4. By G1: {"Illegal immigrants exist in X."$^{\$6}$, "X is in a bad situation."$^{\$6}$, "Illegal immigrants have overrun large portions of California."$^{\$10}$}

The cost increased because the reasoning is abductive. Note that we distributed the increased cost \$12 to two statements.

5. By O1: {"Illegal immigrants overrun X."$^{\$7.2}$, "X is in a bad situation."$^{\$6}$, "Illegal immigrants have overrun large portions of California."$^{\$10}$}

6. By factoring with X = large portions of California: {"Large portions of California is in a bad situation."$^{\$6}$, "Illegal immigrants have overrun large portions of California."$^{\$7.2}$}

The final result contains "Large portions of California is in a bad situation." with the cost \$6, which means that this statement is an implicit assumption of the original input argument.

5. Related work

In Natural Language Processing, commonsense reasoning has received much attention in recent years. One such example is the field of knowledge graph completion. Given an existing knowledge graph encoding entity relationships (e.g. Trump-isPresidentOf-US), the goal is to infer a missing relationship between entities in the graph. There are a wide variety of computational models for this task; see [Wang 17] for a good overview. The second example is the study on narrative schema [Chambers 08, Modi 14, Granroth-wilding 16, Pichotta 16, Modi 16, Weber 17]. Given an event mention (e.g. *John commited a crime*), the goal is to predict an event that will most likely happen next to the input (e.g. *Police arrest John*). These are not studied in the context of abduction; however, these are closedly related to substitution, one of the Natural Logic atomic edits. In future work, we will borrow these well-developed technologies for making abductive reasoning more robust.

In the machine learning community, the interpretability of models have recieved much attention. There are a wide variety of studies for explaining predictions of black box machine learning models (e.g. neural network) [Zhang 18]. On the other hand, our work aims at creating a transparent, white box model instead of finding an explanation to its own prediction.

6. Conclusions and future work

We have presented an abductive natural language understanding framework, integrating Interpretation as Abduction [Hobbs 93b] with Natural Logic [Maccartney 09]. For proof-of-concept, we have demonstrated that this framework explicates an implicit assumption underlying an argument on one concrete example.

Our future work includes several research directions. First, we plan to extract lexical knowledge from an existing, rich knowledge base such as WordNet [Fellbaum 98], FrameNet [Baker 98], or ConceptNet [Speer 12]. Second, we will extend our framework to support a wider variety of commonsense reasoning. We develop new Natural Logic semantic operators for lexical relations such as cause-effect relations, which are frequently used in natural language understanding. In order to make inference more robust, we also plan to use a knowledge graph embedding models [Wang 17], sequence-to-sequence models [Sutskever 14] for substituion, and distributed representation-based similarity calculation for factoring. We also try to learn the cost function of weighted abduction in a supervised manner, extending [Inoue 12b]. Finally, we will develop an efficient search mechanism for finding the least-cost explanation.

Acknowledgements

This work is supported by JSPS KAKENHI 16H06614, and JST CREST (JPMJCR1513).

References

[Angeli 14] Angeli, G. and Manning, C. D.: NaturalLI : Natural Logic Inference for Common Sense Reasoning, *Proceedings of EMNLP*, pp. 534–545 (2014)

[Baker 98] Baker, C. F., Fillmore, C. J., and Lowe, J. B.: The berkeley framenet project, in *Proc. of COLING*, pp. 86–90Association for Computational Linguistics (1998)

[Bowman 13] Bowman, S.: Can recursive neural tensor networks learn logical reasoning?, *arXiv preprint arXiv:1312.6192*, pp. 1–10 (2013)

[Chambers 08] Chambers, N. and Jurafsky, D.: Unsupervised Learning of Narrative Event Chains., in *ACL*, pp. 789–797 (2008)

[Charniak 91] Charniak, E. and Goldman, R.: A Probabilistic Model of Plan Recognition, in *Proceedings of the*

Ninth National Conference on Artificial Intelligence, pp. 160–165 (1991)

[Clark 09] Clark, P. and Harrison, P.: Large-scale extraction and use of knowledge from text, *Proceedings of the fifth international conference on Knowledge capture - K-CAP '09*, pp. 153–160 (2009)

[Fellbaum 98] Fellbaum, C.: WordNet: an electronic lexical database (1998)

[Granroth-wilding 16] Granroth-wilding, M. and Clark, S.: What Happens Next ? Event Prediction Using a Compositional Neural Network Model, in *AAAI* (2016)

[Habernal 17] Habernal, I., Wachsmuth, H., Gurevych, I., and Stein, B.: The Argument Reasoning Comprehension Task (2017)

[Hobbs 93a] Hobbs, J. R., Stickel, M. E., Appelt, D. E., and Martin, P.: Interpretation as abduction, *Artificial Intelligence*, Vol. 63, No. 1-2, pp. 69–142 (1993)

[Hobbs 93b] Hobbs, J. R., Stickel, M., Martin, P., and Edwards, D.: Interpretation as Abduction, *Artificial Intelligence*, Vol. 63, pp. 69–142 (1993)

[Inoue 12a] Inoue, N., Ovchinnikova, E., Inui, K., and Hobbs, J.: Coreference Resolution with ILP-based Weighted Abduction, in *COLING*, pp. 1291–1308 (2012)

[Inoue 12b] Inoue, N., Yamamoto, K., Watanabe, Y., Okazaki, N., and Inui, K.: Online Large-margin Weight Learning for First-order Logic-based Abduction (2012)

[Krishnamurthy 16] Krishnamurthy, J. and Tafjord, O.: Semantic Parsing to Probabilistic Programs for Situated Question Answering, pp. 160–170 (2016)

[Maccartney 09] Maccartney, B. and Manning, C. D.: An extended model of natural logic, *Proceedings of the Eighth International Conference on Computational Semantics IWCS8 09*, No. January, pp. 1–12 (2009)

[Modi 14] Modi, A. and Titov, I.: Inducing neural models of script knowledge, in *CoNLL-2014*, p. 49 (2014)

[Modi 16] Modi, A., Anikina, T., Ostermann, S., and Pinkal, M.: Inscript: Narrative Texts Annotated with Script Information, *Proceedings of the 10th Language Resources and Evaluation Conference (LREC 2016)*, pp. 3485–3493 (2016)

[Pichotta 16] Pichotta, K. and Mooney, R. J.: Using Sentence-Level LSTM Language Models for Script Inference, *Arxiv*, pp. 279–289 (2016)

[Santos 94] Santos, E.: A linear constraint satisfaction approach to cost-based abduction, *Artificial Intelligence*, Vol. 65, No. 1, pp. 1–27 (1994)

[Speer 12] Speer, R. and Havasi, C.: Representing General Relational Knowledge in ConceptNet 5, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, pp. 3679—-3686 (2012)

[Stickel 91] Stickel, M. E.: A prolog-like inference system for computing minimum-cost abductive explanations in natural-language interpretation, *Annals of Mathematics and Artificial Intelligence*, Vol. 4, No. 1-2, pp. 89–105 (1991)

[Sutskever 14] Sutskever, I., Vinyals, O., and Le, Q. V.: Sequence to sequence learning with neural networks, in *Advances in neural information processing systems*, pp. 3104–3112 (2014)

[Taghipour 16] Taghipour, K. and Ng, H. T.: A Neural Approach to Automated Essay Scoring, *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP-16)*, pp. 1882–1891 (2016)

[Wang 17] Wang, Q., Mao, Z., Wang, B., and Guo, L.: Knowledge graph embedding: A survey of approaches and applications, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 29, No. 12, pp. 2724–2743 (2017)

[Weber 17] Weber, N., Balasubramanian, N., and Chambers, N.: Event Representations with Tensor-based Compositions (2017)

[Zhang 18] Zhang, Q. and Zhu, S.-C.: Visual Interpretability for Deep Learning: a Survey (2018)