Toward Plan Recognition in Discourse Using Large-Scale Lexical Resources

Naoya Inoue Kentaro Inui

Tohoku University
{naoya-i, inui}@ecei.tohoku.ac.jp

1 Introduction

An agent's belief and intention to achieve a goal is called a *plan*. We call an and-or tree that encodes the structure of intention-action relations with their belief as a *plan tree*. For example, in Example 1 we recognize a writer's plan as illustrated in Figure 1.

(1) *I sleep with my baby daughter on a bed. I wash the sheets everyday because I'm afraid that we will be bitten by fleas. I want to clean the mattress as well, but it's too big to wash. I don't want to use chemicals. Will a tumble dryer work? How should I clean the mattress?*

Our task, *plan recognition in discourse*, is thus to infer an agents' plan from utterances in discourse, as a plan tree. Recognizing plans in discourse is essential to natural language processing (NLP) tasks (e.g., anaphora resolution) as well as to acquire richer world knowledge.

Computational models for plan recognition have been studied extensively in the 80s and 90s [1, 5, 6, 18, 7, etc.]. Yet, the models have not been tested with open-data since the researchers suffered from a shortage of world knowledge, and hence it has not been demonstrated that they are robust. In the several decades since, however, a number of methods for large-scale knowledge acquisition have been proposed, and the products of their efforts are made available to the public. Now is the time to tackle the problem of plan recognition *in an open-domain*. However, the following two issues are still open for plan recognition with large-scale lexical resources:

Issue 1: The sufficiency of knowledge base *Are existing large-scale knowledge bases enough to perform plan inference in an open-domain? If they are insufficient, what kind of knowledge is necessary?*

Issue 2: Inference mechanism *How should the inference system utilize large-scale knowledge bases?*

We focus on *issue 1* in this paper. As for *issue 2*, we have assumed the framework of Hobbs et al. [12]'s weighted abduction as our model's first step.

The rest of this paper is structured as follows. In the next section, we review previous work about plan recognition. In Section 3, we describe our plan recognition model and methods for constructing knowledge bases. In Section 4, we apply the knowledge bases to web texts and give analyses of the results. Finally, the conclusion is presented along with possibility for further study.

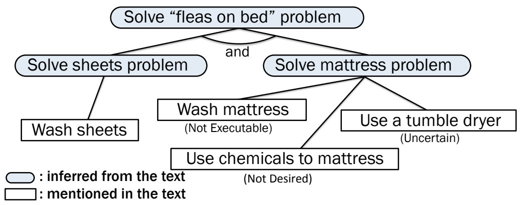

Figure 1: The plan tree for example 1

2 Related work

Many researchers have studied plan recognition extensively as a central field of natural language comprehension literature from the 80s to 90s [1, 5, 6, 18, 7, etc.]. Since the task of plan recognition can be viewed as finding the best explanation (i.e., a plan) to observations (i.e., utterances), most of the proposed methods have been based on *abduction*, the inference process of generating hypotheses to explain observations using background knowledge. For NLP, the most promising approach based on abduction is Hobbs et al. [12]'s framework. They demonstrated that a process of natural language understanding, such as word sense disambiguation or reference resolution, can be described in a single framework based on abduction. In nature, abductive inference generates many possible explanations. An extended framework that chooses the best explanation according to costs of the possible explanations is called *weighted abduction*.

It is critical for the abductive inference model to have broad-coverage knowledge bases in order to build proper explanations. Nevertheless, in the 80s and 90s, the researchers manually created them or simply presumed the existence of them, since methods for automatic acquisition of large-scale lexical knowledge had not been proposed. However, the situation has changed over the past decade. A wide variety of approaches to learn lexical knowledge from large corpora have been proposed, and we now have large-scale lexical resources publicly available. This has led us to an exciting stage to challenge problems of natural language understanding in the 80s or 90s, such as discourse understanding based on Schank and Abelson [23]'s script, Lehnert [16]'s plot unit or plan, but for an open domain.

Moreover, a few researchers have studied probabilistic abductive inference models [3, 10, etc.] and some applied them to the task of plan recognition [13, 21, 3, etc.]. Ovchinnikova et al. [19] applied weighted abduction using WordNet [9] and FrameNet [2] as the knowledge base to discourse processing and evaluated their model through recognizing textual entailment (RTE). They showed the performance of the model as promising when compared to the other existing RTE systems.

However, to the best of our knowledge, plan recognition models with large-scale lexical resources have not been explored. Toward the large-scale plan recognition, we have assumed Hobbs et al. [12]'s weighted abduction as our initial step as detailed in Section 3.1. In weighted abduction, we have two substantial tasks:

- *interpretation generation*: to generate possible explanations for observations, and;
- *interpretation selection*: to evaluate the explanations based on their costs, and choose the best explanation among them.

Our attempt in this paper addresses the first task *interpretation generation*. The second task *interpretation selection* will be explored in future work.

3 Plan recognition in discourse

3.1 Weighted abduction for plan recognition

We use Hobbs et al. [12]'s weighted abduction as our foundation of a large-scale plan recognition model. As mentioned before, weighted abduction generates possible explanations and then evaluates them. Given observations with costs and axioms with weights, it performs backward-reasoning on each observation, propagates its cost to the hypothesis according to the weights on the applied axioms, and merges redundancies where possible. A cost of explanation is then the sum of all the costs on hypotheses in the explanation. Finally, it chooses the lowest cost explanation as the best *interpretation*.[1] Since the task of plan inference can be viewed as finding the best explanation to given observations, the abductive framework is reasonable to capture the nature of the plan recognition problems.

Our plan recognition model consists of three procedures. The first step is to convert an input text to logical forms (LFs). We assume that predicate argument structures for intra-sentential arguments are provided, and directly translate them to Davidsonian-style quasi-LFs. In addition, we introduced the two special predicates: (i) $rel(x, y)$ to encode an association between entities x and y connected by *on* or *of* (e.g., *fleas on a bed*), and (ii) $lexLink(e1, e2, c)$ to capture a conjunction c between events $e1$ and $e2$ (e.g., *John made a pancake because he was hungry*). In the second step, we perform weighted abduction taking the LFs as observations.[2] Knowledge bases used in the

[1] Following [12], we use the term *interpretation* as "explanation".

[2] We assigned the cost \$20 to all the literals. This heuristics will be taken to further consideration.

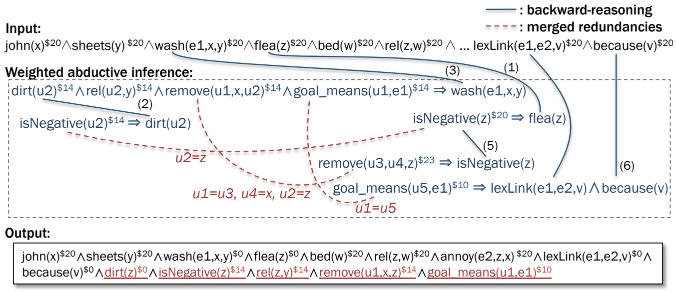

Figure 2: The possible interpretation for the sentence *John washed his sheets because he was annoyed by fleas on his bed.*

abduction consist of four types: (i) relations between predicates, (ii) relations between nouns, (iii) sentiment polarity information and (iv) general inference rules. The inference derives possible interpretations by bridging lexical gaps and merging with similar entities, then chooses the best interpretation. The final step is to extract a plan tree from the interpretation. As described in Section 3.2, since we create axioms of predicates together with its relation (e.g., $goal\_means(e1, e2)$), we just traverse the arguments of $goal\_means$ predicates to create a plan tree.

Now, let us illustrate how our model works for plan recognition taking the sentence *John washed his sheets because he was annoyed by fleas on his bed* as an example. Suppose we have the following axioms in our knowledge base:

(1) $negative(x)^{1.0} \Rightarrow flea(x)$
(2) $negative(x)^{1.0} \Rightarrow dirt(x)$
(3) $dirt(z)^{0.7} \wedge rel(z, y)^{0.7} \wedge remove(e2, x, z)^{0.7} \wedge goal\_means(e2, e1)^{0.7} \Rightarrow wash(e1, x, y)$
(4) $pleasant\_aroma(z)^{0.7} \wedge rel(z, y)^{0.7} \wedge smell(e2, x, z)^{0.7} \wedge goal\_means(e2, e1)^{0.7} \Rightarrow wash(e1, x, y)$
(5) $remove(e, y, x)^{1.3} \Rightarrow isNegative(x)$
(6) $goal\_means(e3, e1)^{0.25} \Rightarrow lexLink(e1, e2, x) \wedge because(x)$

First, we convert the input sentence into the LFs: $john(x)^{\$20} \wedge sheets(y)^{\$20} \wedge wash(e1, x, y)^{\$20} \wedge flea(z)^{\$20} \wedge bed(w)^{\$20} \wedge rel(z, w)^{\$20} \wedge annoy(e2, z, x)^{\$20} \wedge lexLink(e1, e2, v) \wedge because(v)$. Then we perform weighted abductive inference on these LFs and generate its interpretations. A possible interpretation is shown in Figure 2 with the applied axioms. Suppose this interpretation is selected as the lowest cost interpretation among the generated interpretations. We identify the plan "*the goal of washing sheets is to remove the fleas on the sheets*" from $goal\_means(u1, e1)$ predicate in this interpretation. Note that our model resolves the bridging reference [8] of the *fleas* as well as identifying the plan.

3.2 Knowledge base

In this section, we describe how the existing lexical resources are converted to axioms for weighted abduction.[3] The knowledge bases used in our model are summarized in Table 1.

[3] Due to spatial limitations, we describe a part of the resources here.

Table 1: Knowledge bases used in abductive inference

Type of knowledge	Axiom example	# of axioms	Lexical resources
Synonym	$find(x)^{1.0} \Rightarrow discover(x)$, $warmer(x)^{1.0} \Rightarrow heater(x)$	1,419,948	JWN, RPAS
Hypernym-Hyponym	$human(x)^{1.3} \Rightarrow mammal(x)$, $mammal(x)^{1.0} \Rightarrow human(x)$	1,871,984	JWN, RPAS
Words of similar context	$bed(y)^{1.3} \Rightarrow futon(x)$, $pillow(y)^{1.3} \Rightarrow futon(x)$	4,998,620	WSC
Relations between events with roles	$dirt(z)^{0.7} \wedge rel(z,y)^{0.7} \wedge remove(e2,x,z)^{0.7} \wedge goal\_means(e2,e1)^{0.7} \Rightarrow wash(e1,x,y)$	12,033	RPAS
Sentiment Polarity	$isPositive(x)^{1.0} \Rightarrow bonus(x)$, $isNegative(x)^{1.0} \Rightarrow flea(x)$	33,755	JSPL, TRB
Meta-knowledge	$remove(e,y,x)^{1.3} \Rightarrow isNegative(x)$, $\langle verb\rangle(e2,x,z)^{1.3} \wedge goal\_means(e2,e1) \Rightarrow \langle verb\rangle er(y) \wedge use(e1,x,y,z)$	23	Handcoded

JWN: Japanese WordNet [4], WSC: A database of words of similar context [14], RPAS: A database of relations between predicate argument structures [17], JSPL: Japanese sentiment polarity lexicon [11, 15], TRB: A database of trouble expression [22]

We have extracted hypernym-hyponym relations for nouns, verbs and adjectives from Japanese WordNet [4], and converted them into bi-directional axioms but assigned different weights. The assignments of weights reflect our intuition about the hypothesizing of hypernym and hyponym: inference of hypernym from hyponym is more reliable than inference of hyponym from hypernym. Therefore, we assigned lower weights to axioms hypothesizing hypernym than to the other.

Another resource to extract relations between nouns is a database of words of similar contexts [14]. We have extracted the top-5 words from each entry and converted them into axioms so that we need to pay a larger cost than the original word, since these hypotheses are to be used for bridging the lexical gap between axioms rather than for explaining something by these alone. In addition, we used different variables for each entity because words of similar contexts tend to introduce the other objects.

We have converted Matsuyoshi et al. [17]'s database of relations between predicate argument structures into axioms. The database covers major event relations (e.g., synonym, goal-means, hyponym) between events with role information. We have assigned lower weights to axioms that give explanations of a goal than the others since our task is to infer goals of actions.

Lastly, we have manually encoded 23 general inferences of human beings as meta-knowledge. The meta-knowledge captures information coming from the words in discourse themselves (e.g., *if two events are connected by "because", the rst event may have a goal*), and our general intuition to positive or negative (e.g., *if there is something negative, we remove it*).

4 Experiments

4.1 Experimental setup

In weighted abduction, we choose the best interpretation from a set of interpretations induced from background knowledge base. Therefore, we require the best interpretation to be in the evaluated candidate interpretations. In this section, we test the capability of our knowledge bases to generate the best interpretation, and give error analyses. More speci cally, we performed backward-chaining on each input text using our background knowledge base described in Section 3.2, and checked whether an acceptable interpretation about goals could be generated for each observed ac-

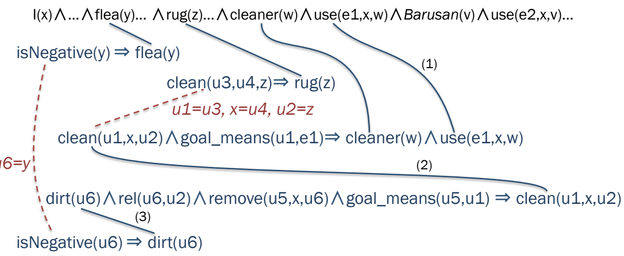

Figure 3: The interpretation for example 2 generated by backward-chaining.

tions. An interpretation is acceptable if goals in the interpretation are judged as coherent with its input text by human.[4] The 30 texts were randomly extracted from housekeeping category in *Yahoo! Chiebukuro*.[5] We manually converted the texts into LFs supplying intra-sentential ellipses and making co-referential entities to the same instance. Recall that we will tackle a problem how we select the best interpretation among the candidate interpretations in future work.

4.2 Results and discussion

Among 62 actions included in the test set, we could generate acceptable interpretations for 48 actions, whereas we could not for 14 actions. One of the interpretations including correctly inferred goals is illustrated in Figure 3. This interpretation was generated from the following input text:

(2) *My daughter and I were bitten by fleas. We have a rug in my room. I use a cleaner$_{(a)}$ every two or three days. If I use the cleaner everyday, will the fleas go away? I don't want to use Barusan$_{(b)}$ as I'm worried about our health. Tell me the solution.*

We successfully generated the goal of *use a cleaner*$_{(a)}$ as "to remove the fleas on the rug" by applying meta-knowledge, relations between events, and sentiment polarity (see (1), (2) and (3) in Figure 3). Note that the knowledge bases we used have not been tuned to the test set. As to *issue 1* mentioned in Section 1, this result preliminarily indicates that *large-scale knowledge bases appear to be moderately sufficient to break into the next step: inference mechanism.*

Yet, the results also showed that we need to ac-

[4]However, re exive explanations are not acceptable: "*John washed his sheets to wash the sheets*" is coherent but not considered to be acceptable. A more strict de nition of the acceptability of an interpretation will be explored in future work.

[5]We considered only texts (i) which have 4 or more than 4 sentences, and (ii) from which we can recognize its plan.

quire further extensive world knowledge about named entities. For instance, we need the knowledge about *Barusan* as well as *use* in order to infer the goal of $\underline{\textit{use Barusan}}_{(b)}$ in example 2: *Barusan* is product name of insecticide, and *use* triggers its purposeful action. In Pustejovsky [20]'s Generative Lexicon (GL), this kind of knowledge is called *telic role*. Some researchers have proposed methods for automatic construction of GL-like knowledge; we can adopt such techniques as proposed in [26, 25, etc.] to acquire the telic roles of named entities. Another choice is to combine the knowledge about the hypernym-hyponym relations of named entities [24] and the telic roles of common nouns. We plan to extract the telic roles by using the explanations of words de ned in a dictionary. For example, we obtain $insects(z) \wedge kill(e2, x, z) \wedge goal\_means(e2, e1) \Rightarrow \wedge insecticide(y) \wedge use(e1, x, y)$ from the explanation of *insecticide*, *a substance used for killing insects*.

Other notable observations give two future directions for *issue 2*. First, the experiment reveals that the meta-knowledge plays a crucial role in generating interpretations as the acceptable interpretations of the 27 actions out of the 48 actions could not be generated without the meta-knowledge. Note that the meta-knowledge considered here can be regarded as generic knowledge about how to use or combine speci c object-level knowledge for interpretation generation. Our next step will be to start addressing *issue 2* by further exploring a wider range of meta-knowledge for both generating and selecting interpretations. Second, in our preliminary experiment, we observed that the inference became intractable very quickly as the number of the input axioms goes beyond one hundred or so. We used a state-of-the-art package for weighted abduction;[6] yet, we suffered severely from its limited scalability in computation. It is crucial to develop a further efficient method for inference to receive the full bene ts of using large-scale knowledge bases.

5 Conclusions

We have discussed the possibility of open-domain plan recognition by exploring the coverage of some of presently available large-scale knowledge bases. We have demonstrated that existing knowledge bases are moderately adequate to perform open-domain plan inference as long as we acquire the additional world knowledge about named entities. In this investigation, we have assumed the framework of Hobbs et al. [12]'s weighted abduction as the basis for the inference mechanism. Our preliminary experiment showed that we need to explore a wider range of meta-knowledge and develop a further efficient method for inference.

Acknowledgements We would like to thank Yahoo Japan Corporation and The National Institute of Informatics which provide *Yahoo! Chiebukuro* data. We also would like to thank The National Institute of Information and Communications which organizes ALAGIN[7], the portal site for large-scale language resources. This work was supported by Grant-in-Aid for JSPS Fellows (22-9719).

[6] http://www.isi.edu/~rutu/.

[7] http://alaginrc.nict.go.jp/

References

[1] J. F. Allen and C. R. Perrault. Analyzing intention in utterances. *Artificial Intelligence*, 15(3):143–178, 1980.

[2] C. F. Baker, C. J. Fillmore, and J. B. Lowe. The Berkeley FrameNet project. In *COLING*, 1998.

[3] J. Blythe, J. R. Hobbs, P. Domingos, R. J. Kate, and R. J. Mooney. Implementing Weighted Abduction in Markov Logic. In *IWCS*, 2011.

[4] F. Bond, H. Isahara, K. Kanzaki, and K. Uchimoto. Bootstrapping a WordNet using multiple existing WordNets. In *LREC*, 2008.

[5] S. Carberry. *Plan Recognition in Natural Language Dialogue*. MIT Press, 1990.

[6] E. Charniak and R. P. Goldman. A Probabilistic Model of Plan Recognition. In *AAAI*, pages 160–165, 1991.

[7] E. Charniak and S. E. Shimony. Cost-based abduction and map explanation. *Artificial Intelligence*, 66(2):345–374, 1994.

[8] H. H. Clark. Bridging. In *Thinking: Readings in Cognitive Science*, 1977.

[9] C. Fellbaum. *WordNet: An Electronic Lexical Database*. MIT Press, 1998.

[10] D. Garrette, K. Erk, and R. Mooney. Integrating Logical Representations with Probabilistic Information using Markov Logic. In *IWCS*, 2011.

[11] M. Higashiyama, K. Inui, and Y. Matsumoto. Acquiring noun polarity knowledge using selectional preferences (in Japanese). In *NLP*, pages 584–587, 2008.

[12] J. R. Hobbs, M. Stickel, D. Appelt, and P. Martin. Interpretation as Abduction. *Artificial Intelligence*, 63:69–142, 1993.

[13] R. J. Kate and R. J. Mooney. Probabilistic Abduction using Markov Logic Networks. In *PAIRS*, 2009.

[14] J. Kazama, S. De Saeger, K. Kuroda, M. Murata, and K. Torisawa. A Bayesian Method for Robust Estimation of Distributional Similarities. In *ACL*, pages 247–256, 2010.

[15] N. Kobayashi, K. Inui, Y. Matsumoto, K. Tateishi, and S. Fukushima. Collecting Evaluative Expressions for Opinion Extraction (in Japanese). *Journal of Natural Language Processing*, 12(2):203–222, 2005.

[16] W. G. Lehnert. Plot Units and Narrative Summarization. *Cognitive Science*, 5(4):293–331, 1981.

[17] S. Matsuyoshi, K. Murakami, Y. Matsumoto, and K. Inui. A Database of Relations between Predicate Argument Structures for Recognizing Textual Entailment and Contradiction. In *ISUC*, 2008.

[18] H. T. Ng and R. J. Mooney. Abductive Plan Recognition and Diagnosis: A Comprehensive Empirical Evaluation. In *KR*, pages 499–508, 1992.

[19] E. Ovchinnikova, N. Montazeri, T. Alexandrov, J. R. Hobbs, M. C. McCord, and R. Mulkar-Mehta. Abductive Reasoning with a Large Knowledge Base for Discourse Processing. In *IWCS*, 2011.

[20] J. Pustejovsky. The generative lexicon. *Computational Linguistics*, 17:409–441, December 1991.

[21] S. Raghavan and R. J. Mooney. Bayesian Abductive Logic Programs. In *STARAI*, pages 82–87, 2010.

[22] S. De Saeger, K. Torisawa, and J. Kazama. Looking for trouble. In *COLING*, pages 185–192, 2008.

[23] R. C. Schank and R. P. Abelson. *Scripts, plans, goals and understanding*. Lawrence Erlbaum, 1977.

[24] A. Sumida, N. Yoshinaga, and K. Torisawa. Boosting precision and recall of hyponymy relation acquisition from hierarchical layouts in wikipedia. In *LREC*, 2008.

[25] K. Torisawa. Automatic Acquisition of Expressions Representing Preparation and Utilization of an Object. In *RANLP*, pages 556–560, 2005.

[26] I. Yamada, T. Baldwin, H. Sumiyoshi, M. Shibata, and N. Yagi. Automatic Acquisition of Qualia Structure from Corpus Data. *IEICE - Transactions on Information and Systems*, E90-D: 1534–1541, October 2007.