Learning Fine-Grained Knowledge about Contingent Relations between Everyday Events
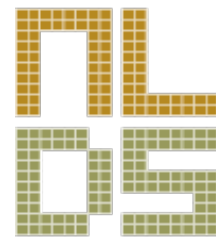
Elahe Rahimtoroghi, Ernesto Hernandez and Marilyn A Walker

Natural Language and Dialogue Systems Lab

Department of Computer Science

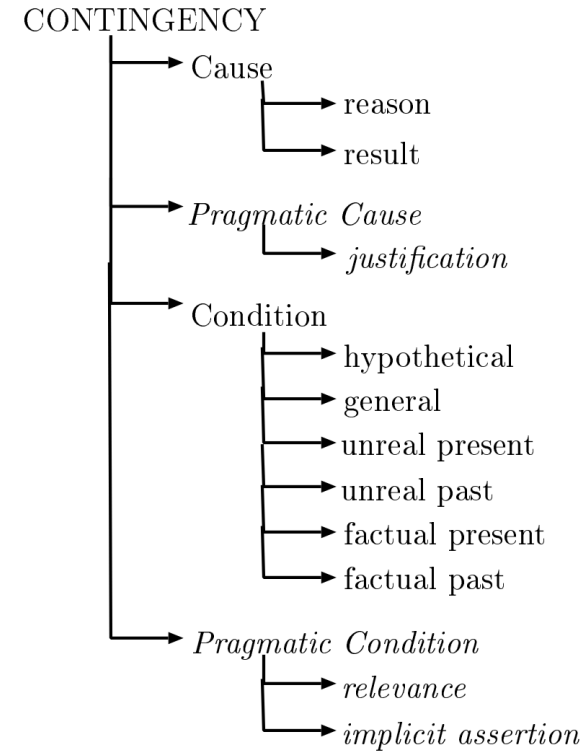University of California Santa Cruz
Santa Cruz, CA 95064, USA

Introduction

Goal

▶ Capture common-sense knowledge about the fine-grained events of everyday experience

  ▶ *opening a fridge* enabling *preparing food*

  ▶ *getting out of bed* being triggered by *an alarm going off*

  Contingency relation between events (Cause and Condition)

PDTB

CONTINGENCY
- Cause
  - reason
  - result
- *Pragmatic Cause*
  - *justification*
- Condition
  - hypothetical
  - general
  - unreal present
  - unreal past
  - factual present
  - factual past
- *Pragmatic Condition*
  - *relevance*
  - *implicit assertion*

Much of the user-generated content on social media is provided by ordinary people telling stories about their daily lives

Camping Trip

We packed all our things on the night before Thu (24 Jul) except for frozen food. We brought a lot of things along. **We woke up** early on Thu and JS started packing the frozen marinatinated food inside the small cooler... In the end, we decided the best place to set up the tent was the squarish ground that's located on the right. Prior to setting up our tent, **we placed a tarp on the ground**. In this way, the underneaths of the tent would be kept clean. After that, **we set the tent up**.

Storm

I don't know if I would've been as calm as I was without the radio, as **the hurricane made landfall** in Galveston at 2:10AM on Saturday. As **the wind blew**, branches thudded on the roof or trees snapped, it was helpful to pinpoint the place... **A tree fell** on the garage roof, but it's minor damage compared to what could've happened. We then **started cleaning up**, despite Sugar Land implementing a curfew until 2pm; I didn't see any policemen enforcing this. Luckily my dad has a gas saw (as opposed to electric), so **we helped cut up** three of our neighbors' trees. **I did a lot of raking**, and there's so much debris in the garbage.

- Rich with common-sense knowledge about contingent relations between events
  - *placing a tarp, setting up a tent*
  - *the hurricane made landfall, the wind blew, a tree fell*
  - *started cleaning up, cut up the trees, raking*

This fine-grained knowledge is simply not found in previous work on narrative event collections

A Brief Look at Previous Work

▶ Much of the previous work is not focused on a particular relation between events (Chambers and Jurafsky, 2008; Chambers and Jurafsky, 2009; Manshadi et al., 2008; Nguyen et al., 2015; Balasubramanian et al., 2013; Pichotta and Mooney, 2014)   ⟶   **Contingency**

▶ Main focus is on newswire   ⟶   **Personal stories**

▶ Evaluation criteria: narrative cloze test   ⟶   **New evaluation method as well as previous work**

Challenge: Personal stories provide both advantages and disadvantages

▶ Told in chronological order

▶ Temporal order between events is a strong cue to contingency

▶ Their structure is more similar to oral narrative (Labov and Waletzky, 1967; Labov, 1997) than to newswire

▶ Only about a third of the sentences in a personal narrative describe actions (Rahimtoroghi et al., 2014; Swanson et al., 2014)

▶ Novel methods are needed to find useful relationships between events

#	Category	Story Clause
1	Orientation	Now, on with this week's story...
2	Orientation	The last month has been hectic.
3	Orientation	Turbo charged.
4	Orientation	Lot's of work because I was learning from Tim, my partner in crime.
5	Orientation	This hasn't been helped by the intense pressure in town due to the political transition coming to an end.
6	Orientation	This week things started alright and on schedule.
7	Action	But I managed to get myself arrested by the traffic police (rouleage) early last Wednesday.
8	Action	After yelling excessively at their outright corrupted methods
9	Action	and asking incessantly for what law I actually broke,
10	Action	they managed to bring me in at the police HQ.
11	Action	I was drawing too much of a curious crowd for the authorities.
12	Action	In about half an hour at police HQ I had charmed every one around.
13	Action	I had prepared my "gift" as they wished.
14	Evaluation	Decision witheld, they decided that I neednt to bother,
15	Evaluation	they liked me too much.
16	Evaluation	I should go free.

Event Representation and Extraction

Event: Verb Lemma (subj:Subject Lemma, dobj:Direct Object Lemma, prt:Particle)
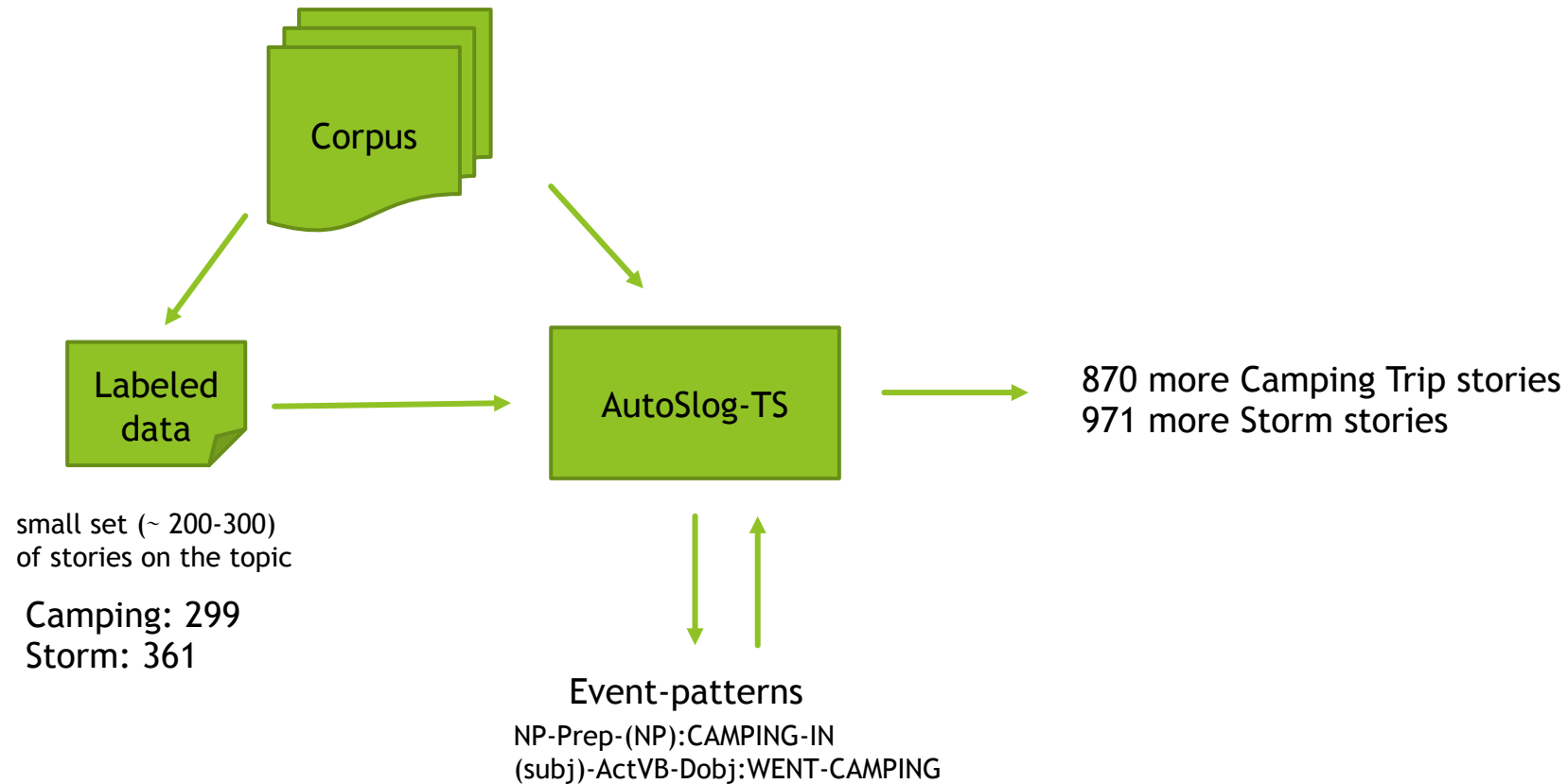
#	Sentence → Event Representation
1	but it wasn't at all frustrating *putting up the tent* and setting up the first night → put (dobj:tent, prt:up)
2	The next day *we had oatmeal* for breakfast → have (subj:PERSON, dobj:oatmeal)
3	by the time *we reached the Lost River Valley Campground*, it was already past 1 pm → reach (subj:PERSON, dobj:LOCATION)
4	then *JS set up a shelter* above the picnic table → set (subj:PERSON, dobj:shelter, prt:up)
5	once the rain stopped, *we built a campfire* using the firewoods → build (subj:PERSON, dobj:campfire)

- ▶ Multi-argument representation is richer, capable of capturing interactions between multiple events (Pichotta and Mooney, 2014)

- ▶ Event extraction
  - ▶ Stanford dependency parser
  - ▶ Stanford NER

Contributions

▶ Generate topic-sorted personal stories using bootstrapping

▶ Direct comparison of topic-specific data vs. general-domain stories

  ▶ Learn more fine-grained and richer knowledge from topic-specific corpus

  ▶ Even with less amount of data

▶ Two sets of experiments

  ▶ Directly compare to previous work

  ▶ Introduce new evaluation methods

Semi-Supervised Algorithm for Generating Topic-Specific Dataset

Corpus

Labeled data

AutoSlog-TS

870 more Camping Trip stories
971 more Storm stories

small set (~ 200-300)
of stories on the topic

Camping: 299
Storm: 361

Event-patterns

NP-Prep-(NP):CAMPING-IN
(subj)-ActVB-Dobj:WENT-CAMPING

Causal Potential (Beamer and Girju 2009)

- Unsupervised distributional measure

- Tendency of an event pair to encode a causal relation

- Probability of occurring in a causal context

$$CP(e_1, e_2) = log\frac{P(e_2|e_1)}{P(e_2)} + log\frac{P(e_1 \rightarrow e_2)}{P(e_2 \rightarrow e_1)}$$

Temporal order

- Calculate CP for every pair of adjacent events
  - Skip-2 bigram model
  - Two related events may often be separated by a non-event sentences

Evaluations

- Narrative cloze test
  - Sequence of narrative events in a document from which one event has been removed
  - Predict the missing event

- Unigram model results nearly as good as other complicated models (Pichotta and Mooney, 2014)

Automatic Two-Choice Test

▶ Automatically generated set of two-choice questions with the answers

  ▶ Modeled after the COPA task (An Evaluation of Commonsense Causal Reasoning, Roemmele et al., 2011)

  ▶ From held-out test sets for each dataset

▶ Each question consists of one event and two choices

Question event: arrange (dobj:outdoor)

Choice 1: help (dobj:trip)
Choice 2: call (subj:PERSON)

▶ Predict which of the two choices is more likely to have a contingency relation with the event in the question

Comparison to Previous Work:
Rel-gram Tuples (Balasubramanian et al., 2013)

- *Rel-grams:* Generate pairs of relational tuples of events

  - Use co-occurrence statistics based on Symmetric Conditional Probability

  - Publicly available through an online search interface

  - Outperform the previous work

$$SCP(e_1, e_2) = P(e_2|e_1) \times P(e_1|e_2)$$

- Two experiments:

  - Content of the learned event knowledge

  - Method: one of the baselines on our data

Baselines

- **Event-Unigram**
  - Produce a distribution of normalized frequencies for events

- **Event-Bigram**
  - Bigram probability of every pair of adjacent events using skip-2 bigram model

- **Event-SCP**
  - Symmetric Conditional Probability between event tuples (Balasubramanian et al., 2013)

Datasets

▶ General-domain dataset

  ▶ Train (4,000 stories)

  ▶ Held-out test (200 stories)

▶ Topic-specific dataset

Topic	Dataset	# Docs
Camping Trip	Hand-labeled held-out test	107
	Hand-labeled train (Train-HL)	192
	Train-HL + Bootstrap (Train-HL-BS)	1,062
Storm	Hand-labeled held-out test	98
	Hand-labeled train (Train-HL)	263
	Train-HL + Bootstrap (Train-HL-BS)	1,234

Results

Topic	Model	Train Dataset	Accuracy
Camping Trip	Event-Unigram	Train-HL-BS	0.507
	Event-Bigram	Train-HL-BS	0.510
	Event-SCP	Train-HL-BS	0.508
	Causal Potential	Train-HL	0.631
	Causal Potential	Train-HL-BS	0.685
Storm	Event-Unigram	Train-HL-BS	0.510
	Event-Bigram	Train-HL-BS	0.523
	Event-SCP	Train-HL-BS	0.516
	Causal Potential	Train-HL	0.711
	Causal Potential	Train-HL-BS	0.887

Model	Accuracy
Event-Unigram	0.478
Event-Bigram	0.481
Event-SCP (Rel-gram)	0.477
Causal Potential	0.510

General-Domain Stories

▶ CP results stronger than all the baselines

▶ Results on topic-specific dataset is significantly stronger than general-domain narratives

▶ More training data collected by bootstrapping improves the accuracy

Compare Camping Trip Event Pairs against the Rel-gram tuples

▶ Find tuples relevant to Camping Trip

  ▶ Used our top 10 indicative event-patterns, generated and ranked during the bootstrapping

  ▶ Apply filtering and ranking

  ▶ Evaluate top N = 100

go (dobj: camping)

[person]	go to	camp	[<<]	[person]	work with	[person]
[person]	go to	camp	[<<]	[person]	go with	[organization]
[person]	go to	camp	[<<]	[person]	be director of	[organization]
[person]	go to	camp	[<<]	[person]	lose	[person]
X:[person]	go to	camp	[<<]	X:[person]	go with	[person]
X:he	go to	camp	[<<]	X:he	go in	[time_unit]
[person]	go to	camp	[<<]	[person]	leave	[person]

Evaluation on Mechanical Turk

▶ New method for evaluating topic-specific contingent event pairs

▶ Rate each pair

   0: The events are not contingent
   1: The events are contingent but not relevant to the specified topic
   2: The events are contingent and somewhat relevant to the specified topic
   3: The events are contingent and strongly relevant to the specified topic

▶ More readable representation for annotators:

Subject – Verb Particle – Direct Object
pack (subj:PERSON, dobj:car, prt: up) → person – pack up - car

Rel-gram Evaluation Results

Label	Rel-gram Tuples
Contingent & Strongly Relevant	7 %
Contingent & Somewhat Relevant	0 %
Contingent & Not Relevant	35 %
Total Contingent	42 %

Label >2: Contingent & strongly topic-relevant

Label = 2: Contingent & somewhat topic-relevant

1 ≤ Label < 2: Contingent & not topic-relevant

Label < 1: Not contingent

Topic-Specific Contingent Event Pairs

- Two filtering methods
  - Selected the frequent pairs for each topic and removed the ones that occur less than 5 times
  - Used the indicative event-patterns for each topic and extracted the pairs that at least included one of these patterns

- Rank by Causal Potential scores to identify the highly contingent ones
  - Evaluated the top N = 100 pairs on Mechanical Turk task

Topic-Specific Pairs Evaluation Results

Label	Camping	Storm
Contingent & Strongly Relevant	44 %	33 %
Contingent & Somewhat Relevant	8 %	20 %
Contingent & Not Relevant	30 %	24 %
Total Contingent	82 %	77 %

▶ Inter-annotator reliability

  ▶ average kappa = 0.73 (substantial agreement)

Examples of Event Pairs

Topic-Specific Dataset

climb → person - find - rock

person - pack up - car → head out

wind - blow - transformer → power - go out

tree - fall - eave → crush

hit - location → evacuate - person

General-Domain Dataset

person - go → go down - trail

person - find - fellow → go back

person - see - gun → see - police

person - go → person - walk down

Conclusions

- Learned new type of knowledge
  - Common-sense knowledge about everyday events focused on contingency relation
- Data collection
  - Semi-supervised bootstrapping approach create topic-sorted dataset
- New evaluation methods
  - Two-choice test and Mechanical Turk task
- Results
  - On topic-specific dataset is significantly stronger than general-domain
  - Method used on the news genre do not work as well on personal stories
  - Fine-grained relations we learn are not found in existing event collections

Thank you!