

The Extraordinary Failure of Complement Coercion Crowdsourcing

Yanai Elazar, Victoria Basmov, Shauli Ravfogel,
Yoav Goldberg and Reut Tsarfaty

Crowdsourcing for Data Collection

- Crowdsourcing has become extremely popular in recent years
- It is used to collect both simple and complex linguistic annotations
- E.g. Sentiment, NLI, QA-SRL, ...



Crowdsourcing Complement Coercion

- In this work, we aim to crowdsource data for the complement coercion phenomenon
- ... and achieve low agreement scores
- Why does this happen?

Complement Coercion

- Constructions that involve an implied action. E.g.:
 - “I **started** __ a new book.”
 - ... **started** *reading/writing/...* (but not *eating*)
 - “I **started** __ a new book I bought last week.”
 - ... **started** *reading* (*but not writing*)

Complement Coercion

- Constructions that involve an implied action. E.g.:
 - “I **started** __ a new book.”



Complement Coercion

- Constructions that involve an implied action. E.g.:

- “I **started** __ a new book.”



- “I **started** __ a new book I bought last week.”



RTE/NLI As Evaluation Framework

The PASCAL Recognising Textual Entailment Challenge

Ido Dagan, Oren Glickman

Computer Science Department

Bar Ilan

Ramat G

{dagan, glikman}

Bernardo Magnini

ITC-irst

A large annotated corpus for learning natural language inference

Samuel R. Bowman^{*†}

sbowman@stanford.edu

Gabor Angeli^{†‡}

angeli@stanford.edu

Christopher Potts^{*}

cgpotts@stanford.edu

Christopher D. Manning^{*†‡}

manning@stanford.edu

^{*}Stanford Linguistics [†]Stanford NLP Group [‡]Stanford Computer Science

RTE/NLI As Evaluation Framework

Premise/Hypothesis	Label
The man is holding a saxophone The man is holding an electric guitar	contradiction ¹
A little girl is very sad. A little girl is very unhappy.	entailment
A couple drinking wine A couple drinking champagne	neutral

*Glockner et al., 2018

Complement Coercion - NLI*

- “I **started** a new book.”
 - ↗ “I **started** *reading* a new book” *Neutral*
 - ↗ “I **started** *eating* a new book” **Contradict**
- “I **started** a new book I bought last week.”
 - ↗ “I **started** *reading* a new book I bought last week.” *Entail*
 - ↗ “I **started** *writing* a new book I bought last week.” **Contradict**

*The NLI Task: can the hypothesis be inferred from the premise?

Complement Coercion Data

- We collect sentences that contain *aspectual verbs* such as: *start*, *begin*, *continue* and *finish*
- ... where the aspectual verb is attached to another verb
- “*it was pike’s idea to start playing these games.*” *Hypothesis*

Complement Coercion Data

- We collect sentences that contain *aspectual verbs* such as: *start*, *begin*, *continue* and *finish*
- ... where the aspectual verb is attached to another verb
- “it was pike’s idea to start playing these games.” *Hypothesis*
↗ “it was pike’s idea to start ~~playing~~ these games.” *Premise*

NLI Crowdsourcing

- We collect labels using Amazon Mechanical Turk
- We follow **standard methodologies**
- We got **0.24** Kappa



NLI Crowdsourcing

- Many work

NLI

A la

ce

¹D

ience



imgflip.com

NLI Crowdsourcing

- Many works

A SICK cure for

How v

M. Marelli¹,

t alexis_ross@o



NLI

Challenge Corpus for
through Inference

gia²
yu.edu

Samuel R. Bowman^{1,2,3}
bowman@nyu.edu

Science
iversity

³Department of Computer Science
New York University

Disagreements

- "... as the young man **finished** his case."
- ↗ "... as the young man **finished** *pleading* his case."



Contradict

Disagreements

- "... as the young man **finished** his case."
- ↗ "... as the young man **finished** *pleading* his case."



Entail

Disagreements

- "... as the young man **finished** his case."
- ↗ "... as the young man **finished** *pleading* his case."



Neutral

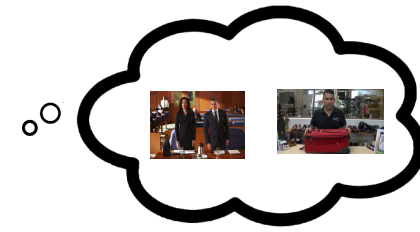
Construals



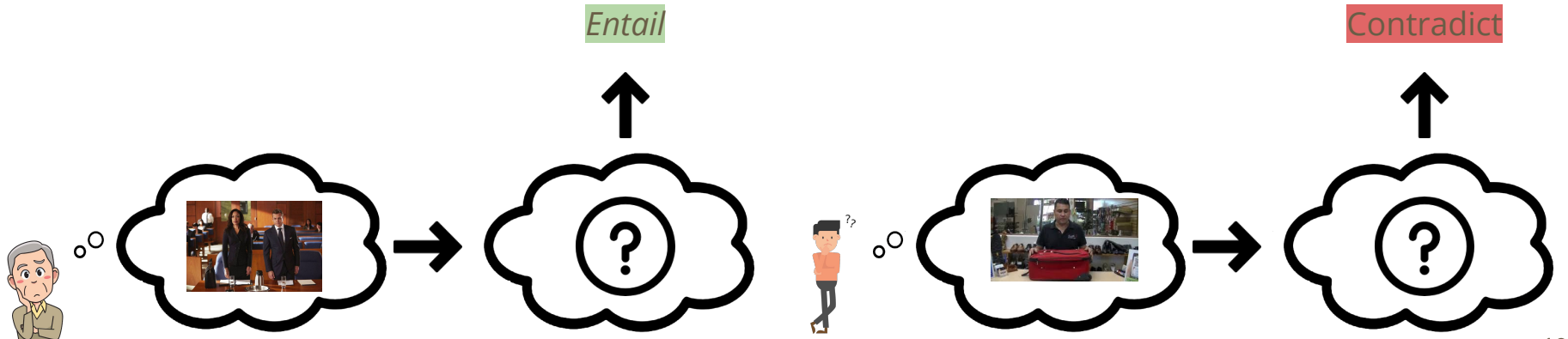
Trott et al. (2020) define *construals*:

- *“The dynamic process of meaning construction, in which speakers and hearers encode and decode, respectively, some intended meaning in a given communicative context”*

Construals



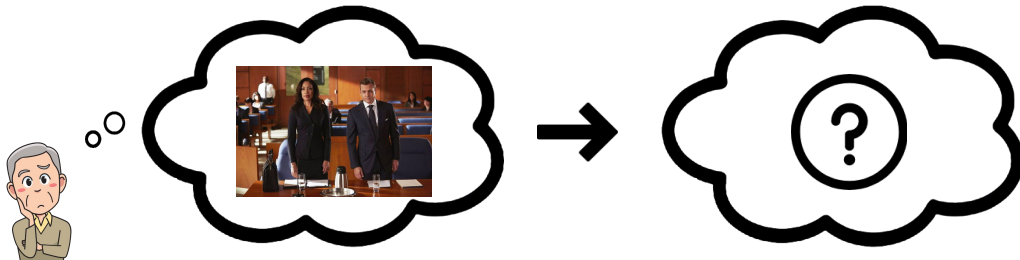
- The complement coercion Instances allow for multiple *construals*
- Once fixating on a scenario, it is hard to imagine others



Construals



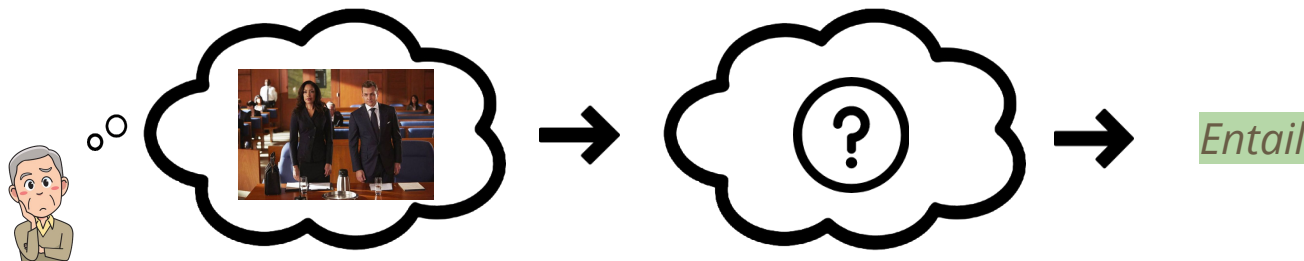
- The complement coercion Instances allow for multiple *construals*
- Once fixating on a scenario, it is hard to imagine others



Construals



- The complement coercion Instances allow for multiple *construals*
- Once fixating on a scenario, it is hard to imagine others
- Which results in different labels, based on the *construal* the reader holds



Construals

- Multiple construals is an under-explored scenario in today's NLP
- We observe it in the *complement coercion* phenomenon
- This is an interesting future research direction

Pitfalls in NLI Data Collection

- Short and loose task definition

Task Definition

- Carefully read the two given sentences
- The two sentences are identical apart from a verb that was dropped
- You will be asked to answer three questions:
Do the sentences describe the same event?
Does the new sentence add new information to the original sentence?
Is the first sentence incorrect/ungrammatical?

Make sure to read all the examples and their explanations to get a notion of the different cases

Pitfalls in NLI Data Collection

- Short and loose task definition
- Minimal annotators training and reliance on common-sense

Task Definition

- Carefully read the two given sentences
- The two sentences are identical apart from a verb that was dropped
- You will be asked to answer three questions:
Do the sentences describe the same event?
Does the new sentence add new information to the original sentence?
Is the first sentence incorrect/ungrammatical?

Make sure to read all the examples and their explanations to get a notion of the different cases



Pitfalls in NLI Data Collection

- Short and loose task definition
- Minimal annotators training and reliance on common-sense
- AMT as a platform does not allow discussion between annotators

Task Definition

- Carefully read the two given sentences
- The two sentences are identical apart from a verb that was dropped
- You will be asked to answer three questions:
Do the sentences describe the same event?
Does the new sentence add new information to the original sentence?
Is the first sentence incorrect/ungrammatical?

Make sure to read all the examples and their explanations to get a notion of the different cases



Conclusions

- Certain phenomena may highlight difficulties, which are not revealed in large-scale “general” data collection
- The **common** methodology for NLI crowdsourcing is suboptimal
- Multiple *construals* is an under-explored scenario, which will require tailored solutions, also in data collection

Thanks

(For reaching this point)



See you at the Q&A Session!

Complement Coercion - Disagreements

- Pavlick and Kwiatowski, 2019 found a similar phenomenon of disagreements in NLI
- However, these are due to a different phenomena, which has to do with people's situation interpretation

p: Someone confessed that a particular thing happened.
h: That thing happened.

