

Recognizing Mentions of Adverse Drug Reaction in Social Media

Gabriel Stanovsky, Daniel Gruhl, Pablo N. Mendes

Bar-Ilan University, IBM Research, Lattice Data Inc.

April 2017

In this talk

1. **Problem:** Identifying adverse drug reactions in social media
 - ▶ *“I stopped taking **Ambien** after three weeks, it gave me a **terrible headache**”*

In this talk

1. **Problem:** Identifying adverse drug reactions in social media
 - ▶ “I stopped taking **Ambien** after three weeks, it gave me a **terrible headache**”
2. **Approach**
 - ▶ LSTM transducer for BIO tagging
 - ▶ + Signal from knowledge graph embeddings

In this talk

1. **Problem:** Identifying adverse drug reactions in social media
 - ▶ “I stopped taking **Ambien** after three weeks, it gave me a **terrible headache**”
2. **Approach**
 - ▶ LSTM transducer for BIO tagging
 - ▶ + Signal from knowledge graph embeddings
3. **Active learning**
 - ▶ Simulates a low resource scenario

Task Definition

Adverse Drug Reaction (ADR)

Unwanted reaction clearly associated with the intake of a drug

- ▶ We focus on automatic ADR identification on social media

Motivation - ADR on Social Media

1. Associate unknown side-effects with a given drug
2. Monitor drug reactions over time
3. Respond to patients' complaints

ADR annotation in forum posts (Ask-A-Patient)

- ▶ Train: **5723 sentences**
- ▶ Test: **1874 sentences**

Drug Ratings for AMBIEN

Average Rating: 3.2 (1408 Ratings)

RATING	REASON	SIDE EFFECTS FOR AMBIEN
1	insomnia due to MS	Sleep was disturbed by <u>waking and vivid dreams</u> . Day after side effects are horrible- <u>dizziness, nausea, diarrhea,</u> headache, severe depression.
1	insomnia	<u>Woke up off and on all night</u> headaches <u>vivid disturbing dreams, heightened senses</u> too much so change in mood aggressiveness

Challenges

Challenges

- ▶ **Context dependent**

 - “Ambien gave me a **terrible headache**”

 - “Ambien made my **headache** go away”

Challenges

- ▶ **Context dependent**

 - “Ambien gave me a **terrible headache**”

 - “Ambien made my **headache** go away”

- ▶ **Colloquial**

 - “hard time getting some **Z's**”

Challenges

- ▶ **Context dependent**

*“Ambien gave me a **terrible headache**”*

*“Ambien made my **headache** go away”*

- ▶ **Colloquial**

“hard time getting some Z’s”

- ▶ **Non-grammatical**

“Short term more loss”

Challenges

- ▶ **Context dependent**

*“Ambien gave me a **terrible headache**”*

*“Ambien made my **headache** go away”*

- ▶ **Colloquial**

*“hard time getting some **Z's**”*

- ▶ **Non-grammatical**

*“Short term more **loss**”*

- ▶ **Coordination**

*“abdominal **gas, cramps and pain**”*

Approach:

LSTM with knowledge graph embeddings

Task Formulation

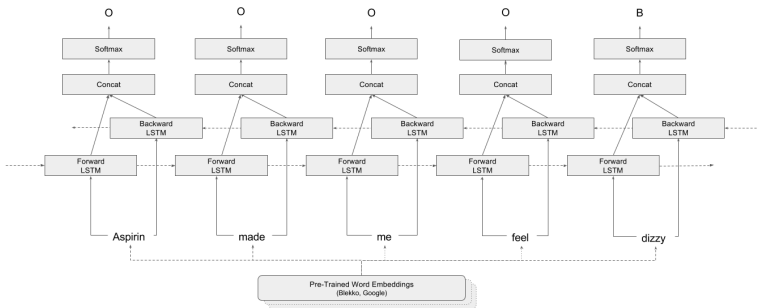
Assign a Beginning, Inside, or Outside label for each word

Example

“*[I]*_O *[stopped]*_O *[taking]*_O *[Ambien]*_O *[after]*_O *[three]*_O *[weeks]*_O –
*[it]*_O *[gave]*_O *[me]*_O *[a]*_O ***[terrible]***_{ADR-B} ***[headache]***_{ADR-I}”

Model

- ▶ bi-RNN transducer model
 - ▶ Outputs a BIO tag for each word
 - ▶ Takes into account context from both past and future words



Integrating External Knowledge

- ▶ DBPedia: Knowledge graph based on Wikipedia
 - ▶ (**Ambien**, *type*, **Drug**)
 - ▶ (**Ambien**, *contains*, **hydroxypropyl**)

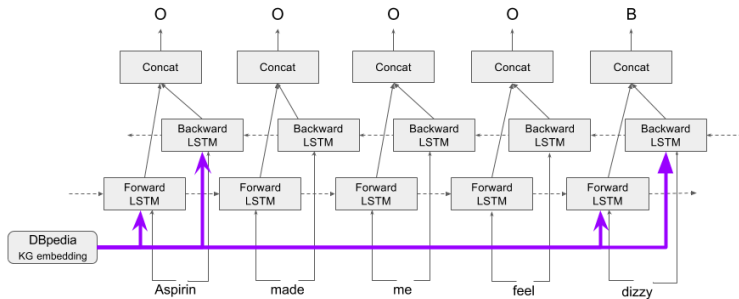
Integrating External Knowledge

- ▶ DBPedia: Knowledge graph based on Wikipedia
 - ▶ (**Ambien**, *type*, **Drug**)
 - ▶ (**Ambien**, *contains*, **hydroxypropyl**)
- ▶ Knowledge graph embedding
 - ▶ Dense representation of entities
 - ▶ Desirably:
Related entities in DBPedia \iff Closer in KB-embedding

Integrating External Knowledge

- ▶ DBPedia: Knowledge graph based on Wikipedia
 - ▶ (**Ambien**, *type*, **Drug**)
 - ▶ (**Ambien**, *contains*, **hydroxypropyl**)
- ▶ Knowledge graph embedding
 - ▶ Dense representation of entities
 - ▶ Desirably:
Related entities in DBPedia \iff Closer in KB-embedding
- ▶ We experiment with a simple approach:
 - ▶ Add verbatim *concept* embeddings to word feats

Prediction Example



Evaluation

	P	R	F1
ADR Oracle	55.2	100	71.1

- ▶ ADR Oracle - Marks gold ADR's regardless of context
 - ▶ Context matters → Oracle errs on 45% of cases

Evaluation

	Emb.	% OOV	P	R	F1
ADR Oracle			55.2	100	71.1
LSTM	Random		69.6	74.6	71.9
LSTM	Google	12.5	85.3	86.2	85.7
LSTM	Blekko	7.0	90.5	90.1	90.3

- ▶ ADR Oracle - Marks gold ADR's regardless of context
 - ▶ Context matters → Oracle errs on 45% of cases
- ▶ External knowledge improves performance:
 - ▶ Blekko > Google > Random Init.

Evaluation

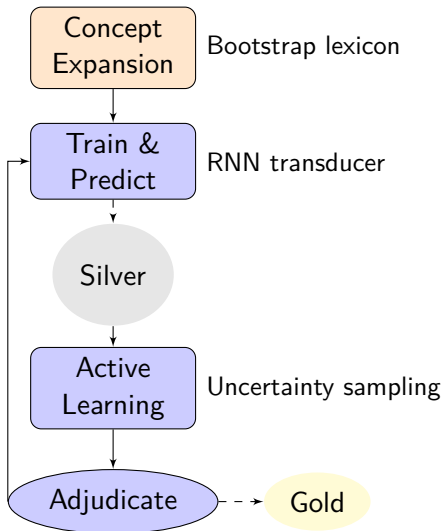
	Emb.	% OOV	P	R	F1
ADR Oracle			55.2	100	71.1
LSTM	Random		69.6	74.6	71.9
LSTM	Google	12.5	85.3	86.2	85.7
LSTM	Blekkko	7.0	90.5	90.1	90.3
LSTM + DBPedia	Blekkko	7.0	92.2	94.5	93.4

- ▶ ADR Oracle - Marks gold ADR's regardless of context
 - ▶ Context matters → Oracle errs on 45% of cases
- ▶ External knowledge improves performance:
 - ▶ Blekkko > Google > Random Init.
 - ▶ DBPedia provides embeddings for 232 (4%) of the words

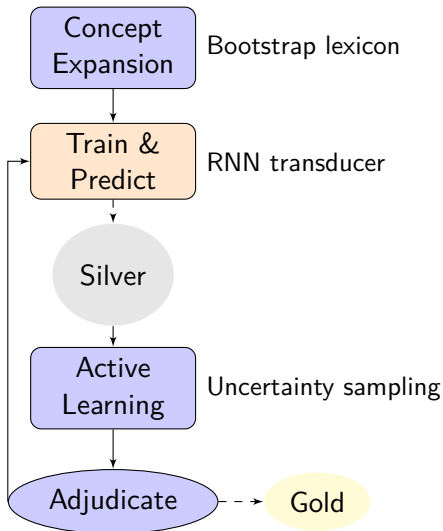
Active Learning:

Concept identification for low-resource tasks

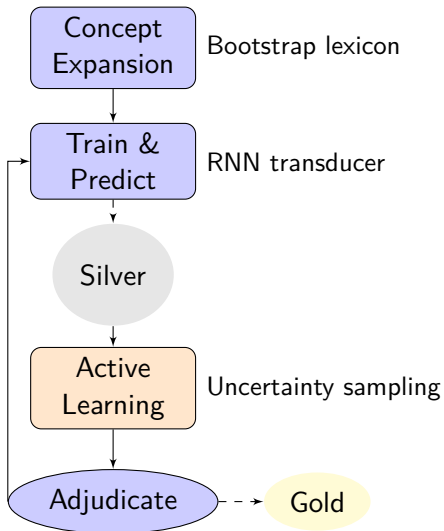
Annotation Flow



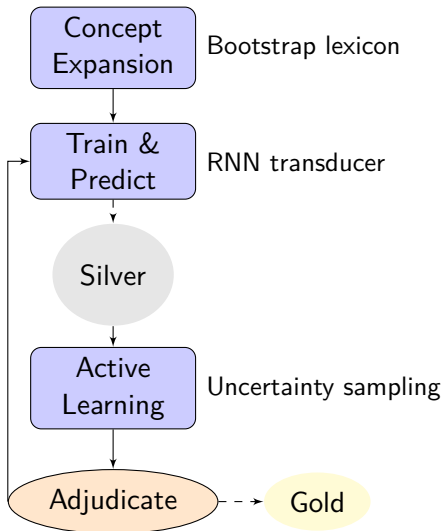
Annotation Flow



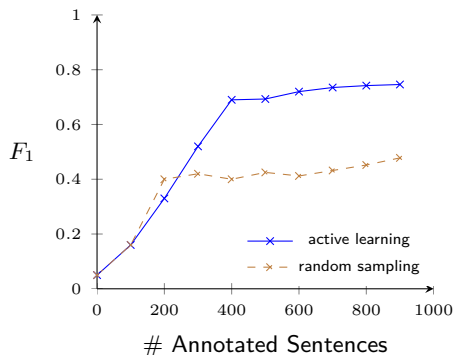
Annotation Flow



Annotation Flow



Training from Rascal



- ▶ Performance after 1hr annotation: **74.2 F1 (88.8 P, 63.8 R)**
- ▶ Uncertainty sampling boosts improvement rate

Wrap-Up

Future Work

- ▶ Use more annotations from CADEC
 - ▶ E.g., symptoms and drugs
- ▶ Use coreference / entity linking to find DBPedia concepts

Conclusions

- ▶ LSTMs can predict ADR on social media
- ▶ Novel use of knowledge base embeddings with LSTMs
- ▶ Active learning can help ADR identification in low-resource domains

Conclusions

- ▶ LSTMs can predict ADR on social media
- ▶ Novel use of knowledge base embeddings with LSTMs
- ▶ Active learning can help ADR identification in low-resource domains

Thanks for listening!
Questions?