

Recognizing Mentions of Adverse Drug Reaction in Social Media

Gabriel Stanovsky, Daniel Gruhl, Pablo N. Mendes
Bar-Ilan University, IBM Research, Lattice Data Inc.

Adverse Drug Reaction in Social Media

- ▶ An **unwanted reaction** clearly associated with the intake of a specific drug
- ▶ We focus on automatic ADR identification on social media
- ▶ Poses many challenges:

- ▶ **Context dependent**
"Ambien gave me a **terrible headache**"
"Ambien made my **headache go away**"
- ▶ **Multi-sentence**
"Started having tension headaches."
"Did not relate to Ambien."
- ▶ **Colloquial**
"hard time getting some Z's"
- ▶ **Presupposition**
"I had the usual problems as most people."
"Driving, online shopping, cooking."
- ▶ **Coordination**
"abdominal gas, cramps and pain"
- ▶ **Non-grammatical**
"Short term more loss"

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 waking and vivid dreams. Day after side effects are horrible: <u>dizziness, nausea, diarrhea, headache, severe depression.</u>
1	insomnia	<u>Woke up off and on</u> all night headaches vivid disturbing dreams, <u>heightened senses</u> too much so change in mood aggressiveness

Useful in many scenarios:

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

(1) Task Formulation

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

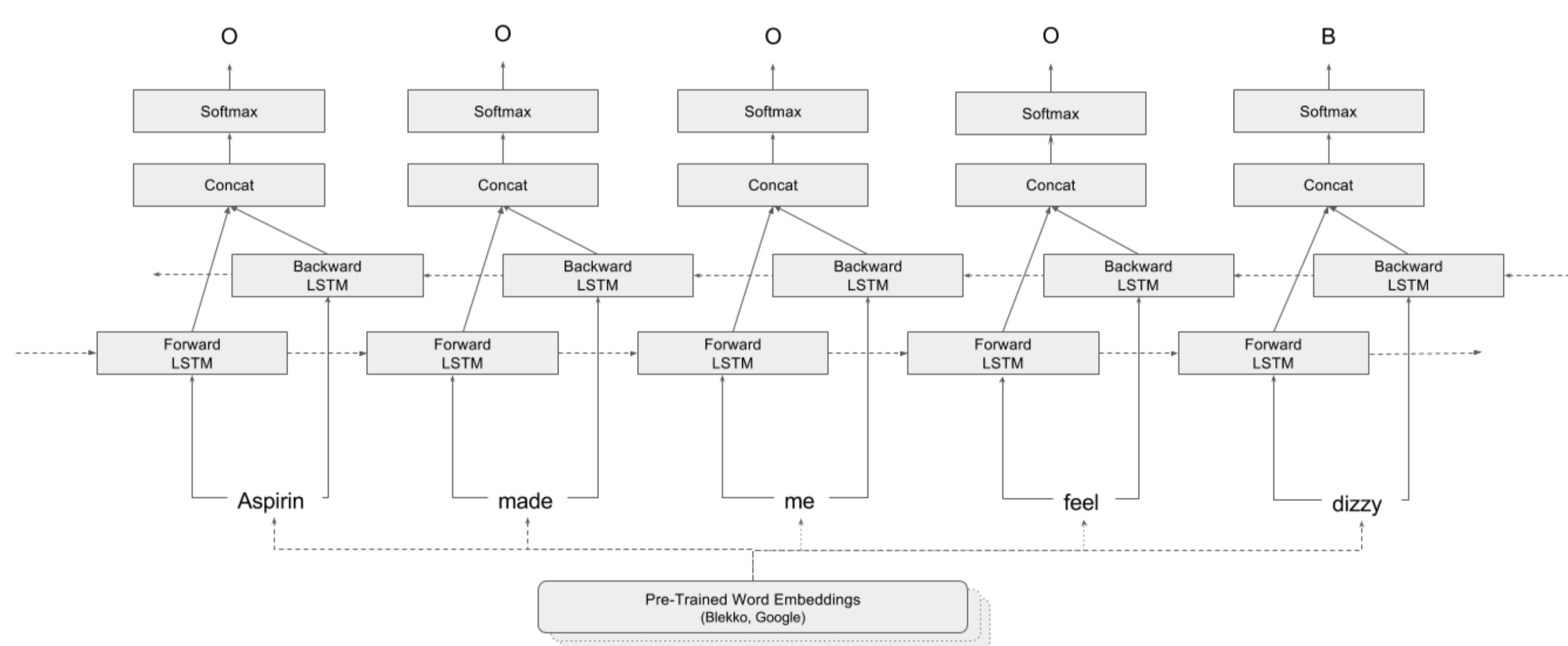
"I₀ stopped₀ taking₀ Ambien₀ after₀ three₀ weeks₀ - it₀ gave₀ me₀ a₀ **terrible**_{ADR-B} **headache**_{ADR-I}"

Similar to the approach taken in other sequence labeling tasks:

- ▶ Named Entity Recognition
- ▶ Noun phrase chunking

(2) Model

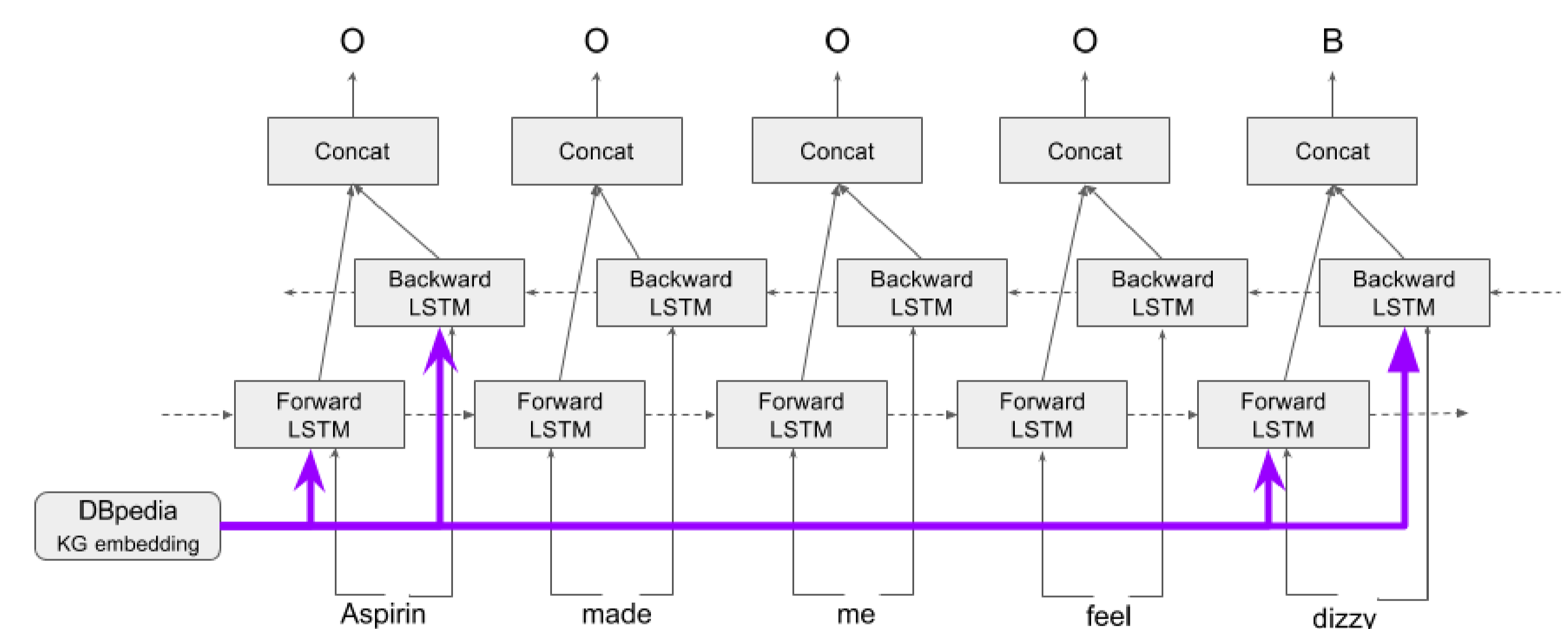
- ▶ bi-RNN transducer model
 - ▶ Outputs a BIO tag for each word
 - ▶ Takes into account context from both past and future words



(3) 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

(4) Knowledge-Infused Model



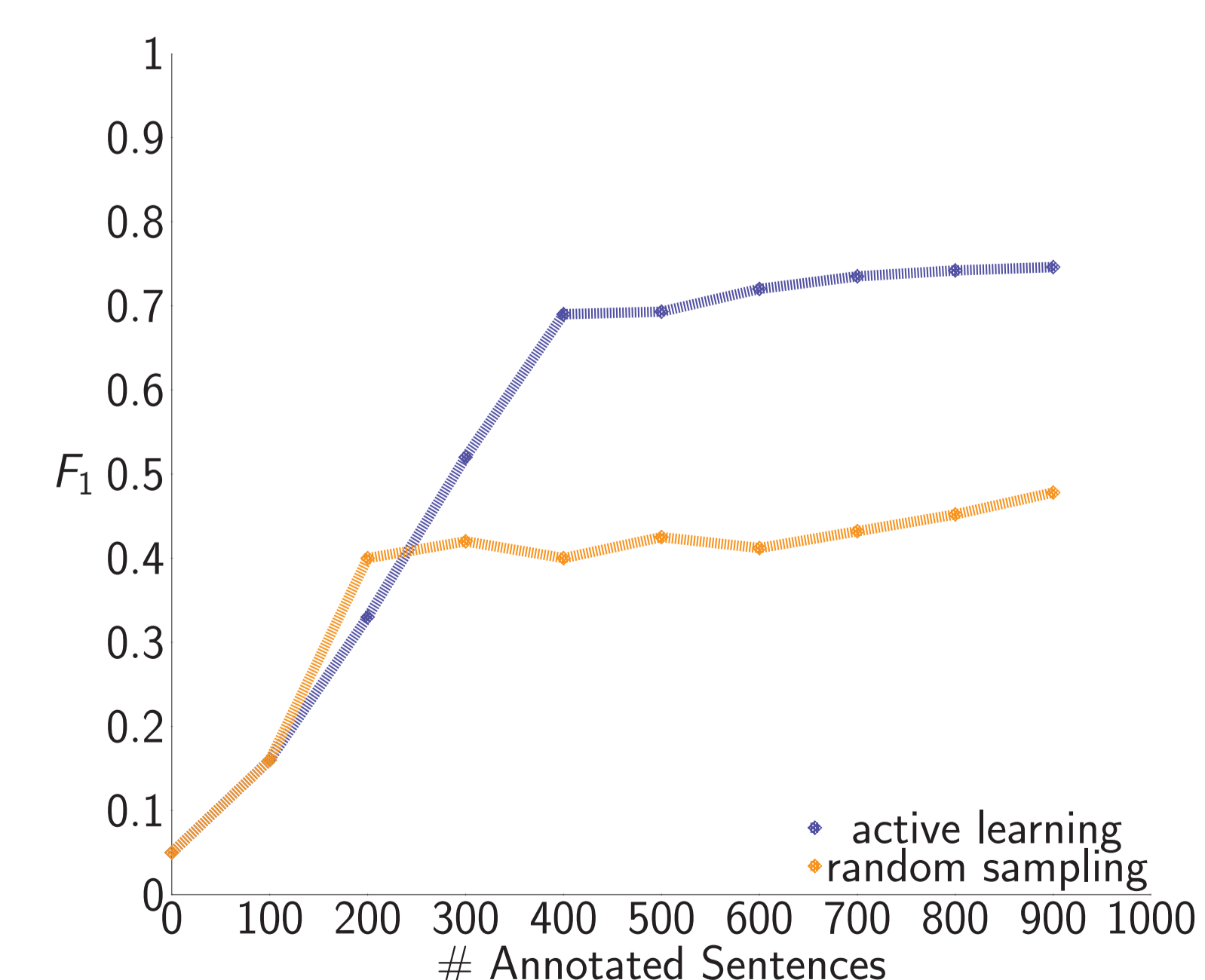
Evaluation

Fully-Supervised Setting

	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
LSTM + DBPedia	Blekko	7.0	92.2	94.5	93.4

- ▶ ADR Oracle - Marks gold ADR's regardless of context
- ▶ Context matters \rightarrow Oracle errs on 45% of cases
- ▶ Blekko > Google > Random Init.
- ▶ DBPedia provides embeddings for 232 (4%) of the words

Active Learning



- ▶ Simulating a domain with low resources
- ▶ Performance after 1hr annotation: **74.2 F1 (88.8 P, 63.8 R)**
- ▶ Uncertainty sampling boosts improvement rate