Event Extraction as Dependency Parsing (in BioNLP 2011)

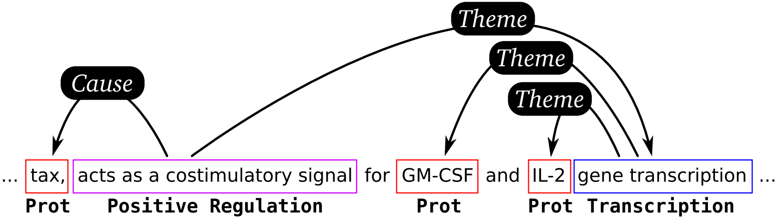
David McClosky

Stanford University

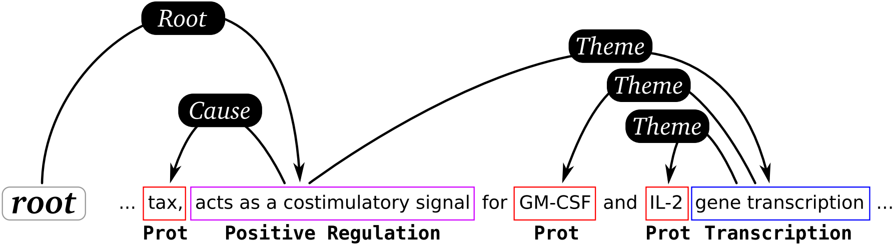
6.24.2011

Joint work with **Mihai Surdeanu** and **Christopher D. Manning**

Our approach in two slides...



Our approach in two slides...



Full details in [McClosky, Surdeanu, and Manning, ACL 2011]

Outline

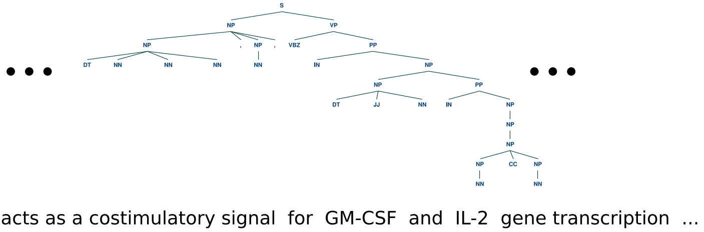
1. Event Parsing
2. Adapting to BioNLP 2011
3. Experiments
4. Conclusion

Approach

... tax, acts as a costimulatory signal for GM-CSF and IL-2 gene transcription ...

Preprocessing: Segmentation, tokenization

Approach

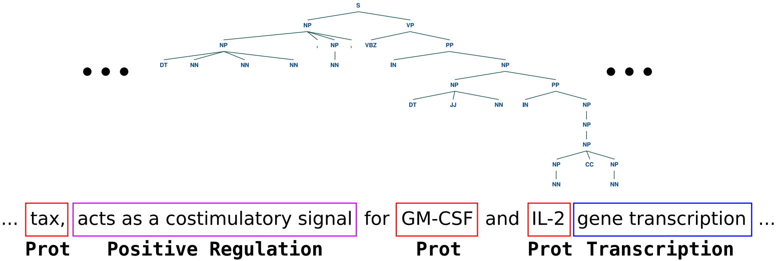


… tax, acts as a costimulatory signal for GM-CSF and IL-2 gene transcription …

Preprocessing: Segmentation, tokenization, syntactic parsing

Self-trained biomedical parser: [McClosky, 2010]

Approach



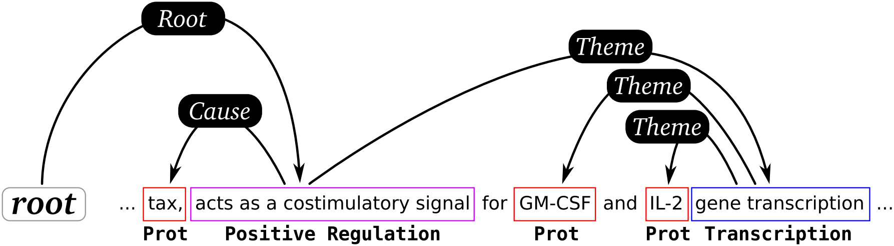
... tax, acts as a costimulatory signal for GM-CSF and IL-2 gene transcription ...

Prot **Positive Regulation** **Prot** **Prot** **Transcription**

Anchor classification: Token classification for event anchors

(similar to [Björne *et al.*, BioNLP 2009])

Approach



Event parsing: Parse anchors and proteins using reranking parser

Maximum-spanning tree based parsing

Why a dependency parser?

- Event structures are non-projective (non-planar)

Maximum-spanning tree based parsing

Why a dependency parser?

- Event structures are non-projective (non-planar)

Why MSTParser? [McDonald *et al.*, EMNLP 2005]

- Handles non-projective trees naturally

Maximum-spanning tree based parsing

Why a dependency parser?

- Event structures are non-projective (non-planar)

Why MSTParser? [McDonald *et al.*, EMNLP 2005]

- Handles non-projective trees naturally
- Easy to extend feature extractor

Maximum-spanning tree based parsing

Why a dependency parser?

- Event structures are non-projective (non-planar)

Why MSTParser? [McDonald *et al.*, EMNLP 2005]

- Handles non-projective trees naturally
- Easy to extend feature extractor
- Support for *n*-best parsing

Adapting to BioNLP 2011

- General improvements
  - Distributional similarity features in anchor detection

Adapting to BioNLP 2011

- General improvements
  - Distributional similarity features in anchor detection
  - Improved head percolation rules for multiword anchors

Adapting to BioNLP 2011

- General improvements
  - Distributional similarity features in anchor detection
  - Improved head percolation rules for multiword anchors
  - Using lemmas (along with word forms) during event parsing

Adapting to BioNLP 2011

- General improvements
  - Distributional similarity features in anchor detection
  - Improved head percolation rules for multiword anchors
  - Using lemmas (along with word forms) during event parsing
- Domain-specific customization
  - Update event type information (EPI, ID)

Adapting to BioNLP 2011

- General improvements
  - Distributional similarity features in anchor detection
  - Improved head percolation rules for multiword anchors
  - Using lemmas (along with word forms) during event parsing
- Domain-specific customization
  - Update event type information (EPI, ID)
  - Combine ID training data with GENIA (ID)

Adapting to BioNLP 2011

- General improvements
    - Distributional similarity features in anchor detection
    - Improved head percolation rules for multiword anchors
    - Using lemmas (along with word forms) during event parsing
- Domain-specific customization
    - Update event type information (EPI, ID)
    - Combine ID training data with GENIA (ID)
    - Removing nested entities (ID)

Results on Genia development

Decoder(s)	Parser	Reranker
1P	49.0	49.4
2P	49.5	50.5
1N	**49.9**	50.2
2N	46.5	47.9
All	—	**50.7**

Results on Epigenetics development

Decoder(s)	Parser	Reranker
1P	62.3	63.3
2P	62.2	63.3
1N	**62.9**	**64.6**
2N	60.8	63.8
All	—	64.1

(note: issues with our internal evaluator implementation)

Domain adaptation for Infectious Diseases

Model	Precision	Recall	*f*-score
ID	**59.3**	38.0	46.3
ID (×1) + GE	52.0	40.2	45.3
ID (×2) + GE	52.4	41.7	46.4
ID (×3) + GE	54.8	**45.0**	**49.4**
ID (×4) + GE	55.2	43.8	48.9
ID (×5) + GE	55.1	44.7	**49.4**

(parser only with 2N decoder)

Results on Infectious Diseases development

Decoder(s)	Parser	Reranker
1P	46.0	48.5
2P	47.8	49.8
1N	48.5	49.4
2N	**49.4**	48.8
All	—	**50.2**

Summary

- New approach to event extraction
    - Parsing can be used for event extraction
    - Reranker further improves performance
- Minimal changes to adapt to new BioNLP domains
- Component in the FAUST system (stay tuned!)
- Code coming soon!
  `http://nlp.stanford.edu/software/eventparsing.shtml`

Summary

- New approach to event extraction
  - Parsing can be used for event extraction
  - Reranker further improves performance
- Minimal changes to adapt to new BioNLP domains
- Component in the FAUST system (stay tuned!)
- Code coming soon!
  `http://nlp.stanford.edu/software/eventparsing.shtml`

Questions?