

BioEvent Extraction and Hedge Classification

Sabine Bergler

CLaC Lab
Department of Computer Science and Software Engineering
Concordia University

CLaC Lab - S

BioNLP event recognition and hedging Halil Kilicoglu

PP attachment in Proteomics Jona Schuman

causal relations in BioNLP Michelle Khalife

IR from biotext Jona Schuman, Stephen Morse

sentiment analysis Alina Andreevskaia

emotion annotation Elena Frantova

social media navigation Julien Dubuc

BioNLP

- The goal of BioNLP is to assist biomedical scientists with their informational needs
 - Manage information overload
 - Keep abreast of recent advances reported in research articles
 - Allow generation of new knowledge
- Tasks include:
 - Information Retrieval (IR) (bag of words)
 - Event extraction (*protein-protein interactions, regulatory events*) (predicate argument structure, grammatical relations)
 - Deeper semantic interpretation of scientific texts (*uncertainty, speculation, negation*) (heuristics over parse information)

Events

Definition An *event* is a process or a state with a certain number of engaged participants and a temporal (and possibly spatial) location

Example *GrpL*, a *Grb2-related adaptor protein*, *interacts* with *SLP-76* ...

Notation *interacts*(e_1 , *GrpL*, *SLP-76*)

	Predicate		<i>interacts</i>
Predicate-Argument Structure	Theme1		<i>GrpL</i>
	Theme2		<i>SLP-76</i>

Realization Verbs, nominalizations, and copular constructions

Higher Order Predicates

Definition A *higher order predicate* takes another predicate (possibly higher order) as an argument

Example ... *Nmi enhances the association of CBP/p300 coactivator proteins with Stat1* ...

Notation *associates*(e_1 , *CBP/p300 coactivator proteins*, *Stat1*)
enhances(e_2 , e_1 , *Nmi*)

Realization Verbs, nominalizations, copular constructions, adverbs

Embedding Constructions

reported speech $\text{SAY}(X, \phi)$

sentiment analysis $+/-(\phi)$

emotion $\text{SAD}(\text{speaker}, \phi)$

factivity/modality $\text{BELIEVE}(X, \phi), \text{POSS}(\phi)$

Speculative Language

Simple factive statements are rare in biotext, speculated events abound.

- ... *raising the possibility* that endogenous *PtdIns* *may modulate* cellular responses to *LPS* ...
- ... *Egr-2 is likely to play a direct role* in aberrant *fasL up-regulation* ...

Linguistic Treatment of Higher Order Predications

- Negation
 - It reverses the truth value of a proposition
 - Multiple negations do not simply cancel out
 - *He is not unhappy.*
- Factuality, speculation, hedging, evidentiality, belief
 - They communicate information about the *status* of a proposition
 - Associated with the reliability of a proposition
 - *I doubt that he is learning.*
- Causation: Semantic analysis of event causation focuses on sentences like the following:
 - *Mary closed the door.*
 - Mary DO CAUSE BECOME the door be closed.

Authorities believe that outbreak may not have caused the deaths.

Linguistic Treatment of Higher Order Predications

- Negation
 - It reverses the truth value of a proposition
 - Multiple negations do not simply cancel out
 - *He is not unhappy.*
- Factuality, speculation, hedging, evidentiality, belief
 - They communicate information about the *status* of a proposition
 - Associated with the reliability of a proposition
 - *I doubt that he is learning.*
- Causation: Semantic analysis of event causation focuses on sentences like the following:
 - *Mary closed the door.*
 - Mary DO CAUSE BECOME the door be closed.

Authorities believe that outbreak may not have caused the deaths.

Linguistic Treatment of Higher Order Predications

- Negation
 - It reverses the truth value of a proposition
 - Multiple negations do not simply cancel out
 - *He is not unhappy.*
- Factuality, speculation, hedging, evidentiality, belief
 - They communicate information about the *status* of a proposition
 - Associated with the reliability of a proposition
 - *I doubt that he is learning.*
- Causation: Semantic analysis of event causation focuses on sentences like the following:
 - *Mary closed the door.*
 - *Mary DO CAUSE BECOME the door be closed.*

Authorities believe that outbreak may not have caused the deaths.

Linguistic Treatment of Higher Order Predications

- Negation
 - It reverses the truth value of a proposition
 - Multiple negations do not simply cancel out
 - *He is not unhappy.*
- Factuality, speculation, hedging, evidentiality, belief
 - They communicate information about the *status* of a proposition
 - Associated with the reliability of a proposition
 - *I doubt that he is learning.*
- Causation: Semantic analysis of event causation focuses on sentences like the following:
 - *Mary closed the door.*
 - Mary DO CAUSE BECOME the door be closed.

Authorities believe that outbreak may not have caused the deaths.

Simple Event vs. Higher Order Predication

Multidrug-resistant HL-60 cells respond to TPA ...

respond(e_1 , *HL-60 cells*, *TPA*)

Simple Event vs. Higher Order Predication

The results *demonstrate* that multidrug-resistant *HL-60 cells* *fail* to *respond* to *TPA* ...

respond(e_1 , *HL-60 cells*, *TPA*)

NEGATION: *fail*(m_1 , e_1)

EVIDENTIAL: *demonstrate*(m_2 , *The results*, m_1)

Higher Order Dimensions

- Negation
 - *In the **absence** of TCR-mediated **activation**, ...*
- Factuality
 - subsumes speculation, hedging, belief, evidentiality dimensions
 - *Our results **suggest** that IL-12 had a direct **effect** on IRF-1 ...*
- Causation
 - *... I kappa B alpha is **degraded**, **resulting in** NF-kappa B **translocation** into the nucleus.*

Higher Order Dimensions

- Negation
 - *In the **absence** of TCR-mediated **activation**, ...*
- Factuality
 - subsumes speculation, hedging, belief, evidentiality dimensions
 - *Our results **suggest** that IL-12 had a direct **effect** on IRF-1 ...*
- Causation
 - *... I kappa B alpha is **degraded**, **resulting in** NF-kappa B **translocation** into the nucleus.*

Unified approach?

BioNLP Shared Task 2009

- BioNLP'09 Event Extraction Shared Task participation
 - Task 1: Core event extraction (3rd place)
 - Task 3: Speculation and negation recognition (1st place)
- Rule-based methodology
 - Statistical compilation of a dictionary of event triggers with limited expansion
 - Grammar for participant identification from dependency parse
 - Heuristics-based speculation and negation detection

Related Work

- Speculation recognition
 - Speculative sentences vs. non-speculative sentences
 - Supervised machine learning techniques [Light et al., 2004; Medlock and Briscoe, 2007; Szarvas, 2008]
 - Lexico-syntactic patterns [Kilicoglu and Bergler, 2008]
- Negation detection
 - Identification of negated terms or concepts
 - Machine learning techniques [Goldin and Chapman, 2003; Averbuch et al., 2004]
 - Pattern-based approaches [Mutalik et al., 2001; Chapman et al., 2001]
- Recent interest in negation and speculation scope identification
 - Memory-based classification [Morante et al., 2008]

Preprocessing

- Create an enriched XML representation of documents
 - Sentences, entities and their offset positions
 - Word information: tokens, POS tags, lemmas and indexes
 - Dependency parses of sentences
- Stanford Lexicalized Parser for word information and dependency parsing

Core Event Extraction Pipeline

- 1 Determine event triggers and their event types
 - Dictionary lookup and “goodness score”
- 2 Identify potential arguments
 - Participant identification grammar
- 3 Recursive sub-event identification for regulatory events
- 4 Post-processing rules for event trigger limitations

Event Triggers

- A dictionary of event triggers drawn from training data
- Event triggers: verbs, nouns and adjectives (predicates)
- Limited term expansion and filtering
 - Both hyphenated and non-hyphenated forms of prefixed triggers (e.g., *co-*, *down-*, *up-*, *trans-*)
 - Derivational forms (e.g., *dimerization* after *dimerize*)
 - Weak event trigger elimination (e.g., *absence* for `Negative_regulation`)
 - No multi-word event triggers
- Event trigger/event type “goodness score”
 - Maximum likelihood estimation
 - Used as a threshold

Event Participant Identification

- Based on rules using “collapsed” Stanford dependency relations
- Extracted and ranked the dependency paths between event triggers and participants
 - Frequency of occurrence

Dependency	POS	Role	Event	Trigger
<i>dobj</i> ¹	VB	Theme	*	*
<i>nsubj</i> ²	JJ	Cause	[Regulatory]	[<i>responsible,sufficient,...</i>]
<i>prep_to</i> ³	VB	Theme	Binding	*
<i>prep_to</i>	VB	Theme	Pos.regulation	[<i>lead,contribute</i>]
<i>prep_between</i> ⁴	NN*	Theme	Binding	[<i>association,interaction</i>]

¹direct object

²nominal subject

³prepositional phrase headed by *to*

⁴prepositional phrase headed by *between*

Special Considerations for Regulatory Events

- Events as event participants
 - Dependency path between the event trigger and the sub-event trigger
 - *We have examined the **effect** of LTB4 on the **expression** of ...*
- Hyphenated participle modifiers
 - Additional rule without dependency relations
 - *... **LPS-mediated** TF expression ...*
- Reversal of semantic roles for *require* and *involve*
- “Corrected” dependency paths for PP attachment errors
 - For certain trigger words (e.g., *effect*, *influence*, *role*)

$\begin{array}{l}
 \text{doj}(\text{examined}, \text{effect}) \\
 \text{prep_on}(\text{examined}, \text{expression})
 \end{array}
 \Rightarrow \text{prep_on}(\text{effect}, \text{expression})$

Coordination and Apposition

■ Coordination

- Both event trigger and participant coordination
- Derived from dependency relations (*conj_**)
- Additional rules for better resolution of coordinated entities
 - Separated by comma or semi-colon
 - Separated by a coordinating conjunction
 - Separated by parenthetical expression
 - Combination of above
- ... *interleukin-2 (IL-2) and IL-4 gene transcription* ...

■ Apposition

- An entity and a word in an apposition construction are considered equivalent
- Derived from dependency relations (e.g., *appos*, *abbrev*)
- ... *regulation of inflammatory cytokine genes including TNF*,
prep_of(regulation,genes)
prep_including(genes,TNF)

Postprocessing Rules

- Loosen the strict assumptions about event triggers
- Multi-word event triggers
 - (*positive* OR *negative*) + nominal Regulation trigger
OR
(*positively* OR *negatively*) + verbal Regulation trigger
⇒ `Positive_regulation` OR `Negative_regulation`
- “One trigger-one event” limitation
 - *overexpression*, *transfect*, etc.

Core Event Extraction Results

Event Class	Recall	Precision	F-score	Rank
Simple events	43.10	73.47	54.33	5
Regulatory events	27.47	49.89	35.43	2
TOTAL	34.98	61.59	44.62	3

- Our system favors precision over recall
- The results at empirically determined “goodness score” threshold of .08
- No special treatment of simple events
- Special focus on regulatory events
 - Led to better ranking

Core Event Extraction Errors

- Naive view of event triggers
 - “Once a trigger-always a trigger” (precision errors) (e.g., *oligomerization*)
 - Previously unseen triggers (recall errors)
- Errors in dependency relations
- Pattern incompleteness
- Anaphoric expressions
- Events spanning multiple sentences

Speculation Recognition

- Refinement of a speculation cue dictionary developed in prior work [Kilicoglu and Bergler, 2008]
 - Ignore some cue classes completely (modal verbs, epistemic adverbs)
 - Introduce a new speculative verb class (active cognition verbs)
 - *examine, evaluate, analyze, study, investigate*
 - Consider the nominal forms as well
- Speculation scope
 - A dependency path between the speculation cue and the event trigger
 - ... *these data suggest that ETS1 may be involved in mediating the increased GM-CSF production ...*
ccomp(suggest,involved)

Negation Detection

- Lexical cues
- Syntactic dependency rules between lexical cues and event triggers

Cue	Dependency
<i>lack, absence</i>	<i>prep_of</i>
<i>inability, failure</i>	<i>infmod</i>
<i>no, not, cannot</i>	<i>det</i>

- Certain dependencies involving the event trigger (e.g., *neg*, *conj_negcc*)
- A negation cue word preceding the event trigger or an event participant (*no*, *not*, *cannot*)

Speculation and Negation Recognition Results

	Recall	Precision	F-score	Rank
Speculation	16.83	50.72	25.27	1
Negation	14.98	50.75	23.13	1
TOTAL	15.86	50.74	24.17	1

- Misidentified or missed base events
- Errors due to speculation and negation detection module
 - Speculation: 4 false positives (out of 39), 7 false negatives (out of 95)
 - Negation: 5 false positives (out of 31), 5 false negatives (out of 107)

Speculation and Negation Detection Errors

- False positive errors
 - Controversial false positive cases
 - *An unidentified Ets family protein binds to ... and **appears** to **negatively regulate** the human IL-2R alpha promoter.*
 - Difficult to annotate correctly and consistently
 - The negation pattern involving negation cues in the token preceding the event trigger
 - Introduced to increase recall
- False negative errors
 - Complex and infrequent patterns
 - *... **suggest** a molecular mechanism for the **inhibition** of ...*
 - *Galectin-3 is ... and is expressed in many leukocytes, **with the notable exception** of B and T lymphocytes*

Conclusions

Linguistically inspired heuristics based on dependency relations

- perform well
- port well to other data sets
- are extensible

Work has to be done on

- consistent annotation guidelines for speculation and other higher order predicates
- anaphora resolution
- events that span several sentences
- subcategorization information for event triggers