On the contribution of specific entity detection and comparative construction to automatic spin detection in biomedical scientific publications

Anna Koroleva, Patrick Paroubek

LIMSI, CNRS, Université Paris-Saclay, France

koroleva@limsi .fr, pap@limsi.fr

The Second Workshop on Processing Emotions, Decisions and Opinions (EDO 2017)
17 Nov. 2017, Poznań, Poland

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 676207

Background

- Spin: **beautifying** the observed research results.
- Spin in **clinical trials**: claiming that the studied treatment had a **positive** effect **greater** than the trial showed.

Abstract with inappropriate claims	Abstract rewritten without spin
Treatment A **may be useful** in controlling cancer-related fatigue **in patients who present with severe fatigue**.	Treatment A **was not more effective** than placebo in controlling cancer-related fatigue.
This study demonstrated **improved PFS and response** for the treatment A compared with comparator B alone, although this did not result in improved survival.	The treatment A was **not more effective** than comparator B on overall survival in patients with metastatic breast cancer previously treated with anthracycline and taxanes.

Background

Impact of spin:

- clinicians **overestimate** efficacy of the treatment

  → impact on **clinical decisions**;
- more positive presentation of a treatment in **health news**;
- change in perception of the treatment by **patients**.

Our aim:

- to help scientific authors, reviewers and editors to identify probable occurences of spin with the use of Natural Language Processing and Machine Learning

Our focus:

- abstracts of articles on Randomized Controlled Trials (RCTs)

A model of Spin

Categories of spin

1. Inappropriate **presentation** of research results (7 subtypes):

- **not evaluating** some of the **results** (e.g. the primary outcome), focus on secondary results or particular **subgroups** of patients;
- ...

2. Inappropriate **interpretation** of research results (3 subtypes):

- a claim that the studied treatment has a **positive** effect / effect **equivalent** to the standard treatment in spite of **non-significant** results;
- ...

3. Inappropriate **extrapolation** (2 subtypes):

- The conclusions are **inappropriate** for **clinical practice** (e.g. an advice to use the treatment not supported by evidence).
- ...

A model of Spin

NLP functionalities for spin identification

- **classification** of articles according to the trial type (RCTs vs. other types);
- document **structure** analysis;
- **evaluation** analysis;
- **entity** extraction (outcomes, population, etc.);
- **relation** extraction (e.g. an outcome and its statistical significance);
- **paraphrase identification** for comparing entity mentions from the abstract with those from the body;
- **syntactic analysis**:

This study demonstrates improved PFS and response for the treatment A compared with comparator B, ***although*** *this did not result in improved survival.*

Entity extraction: previous works

- Focus on RCTs
- Focus on abstracts
- Two-step approach: sentence classifier; entity extraction
- Combination of Machine Learning and symbolic rule based algorithms
- Use of systems to match the article contents with thesauri (e.g. UMLS)
- The definition of the exact text boundaries of an entity mention remains a difficult task.
- The outcome is the most difficult entity to identify.

Entity extraction: our experiments

- **Goal**: baseline, corpus pre-annotation
- **Corpus**: 3,938 articles on RCTs from PubMed Central (PMC)
- **Approach**: manual exploration of the corpus + finite state automata (Unitex environment)
- Outcome identification example:

The <PROL>primary outcome was</PROL> <OUT type=PRIM>the remission of depressive symptoms at the 2-month follow up visit</OUT>, defined as a HDRS score of 7 or less.

<PROL>Secondary outcome parameters are</PROL> <OUT type=SEC>overall mortality, severity of BPD, number of days on the ventilator, number of treatment failures, ventilation-induced lung injury and pulmonary hypertension</OUT> according to clinical parameters.

Comparatives: previous works

- 3 types of constructions: morpho-lexical, syntactic, semantic.
- Components of a comparison: a comparative word, compared entities, compared features[1]
- Works on opinion mining from comparative sentences[2]
- Subjective vs. objective sentences:

*The experimental treatment is **better** than the control treatment.*

*The treatment significantly **improved** survival.*

[1] Gupta, Samir, A. S. M. Ashique Mahmood, Karen E. Ross, Cathy H. Wu, and K. Vijay-Shanker, 2017. Identifying comparative structures in biomedical text. In Proceedings of the BioNLP 2017 workshop.

[2] Ganapathibhotla, Murthy and Bing Liu, 2008. Mining opinions in comparative sentences. In Proceedings of the 22nd International Conference on Computational Linguistics (Coling2008). Coling 2008 Organizing Committee.

Comparatives: relevance for spin detection

Main goal of RCTs: **comparing** two or more **treatments** with respect to a number of **outcomes**

Positive or **negative evaluation** of treatment:

- statements of **superiority** of one treatment over the other treatment:

  *Patients receiving treatment A performed **better** than those receiving treatment B.*

  *Adverse events rate was **higher** in the experimental group than in the control group.*

- **similarity** between the two treatments:

  *Treatment A was **as effective as** treatment B.*

  *Treatment A was **similar** to treatment B in terms of observed adverse events.*

- some **changes** under the experimental treatment:

  *Treatment A **increased** overall survival.*

  *Treatment A **increased** mortality rate.*

Comparatives: relevance for spin detection

Entity extraction from comparatives:

- compared entities: compared treatments; patient groups that received the treatments; value of an outcome before and after a treatment;
- compared feature: an outcome.

Treatment A *was better that* ***treatment B*** *in terms of* ***efficacy****.*

The group receiving treatment A *showed better* ***response rate*** *than* ***the group receiving treatment B****.*

PANSS score *improved with* ***treatment A****.*

Comparatives: our experiments

- **Aim**: to extract the compared entities and features and to detect their type (treatment, patient group, outcome).
- **Corpus**: 3934 abstracts of articles from PMC + 5005 abstracts of articles from Cochrane Schizophrenia group database.
- **2 steps**:

  1) collecting concordances for words and constructions conveying comparative meaning;

  2) identifying morphological, lexical, morphosyntactic and syntactic features of different types of components.

- Outcomes have less specific lexical and morphological features compared to patients and treatments.

Comparatives: our experiments

Type of components	Pattern	N of occurrences / percentage	Examples
Outcome	Verb.change. prtcp + OUT	2306 / 34,6%	Three sessions of education led to significantly increased <Out>insight</Out>.
Outcome	noun.change + PP(in/on/of + OUT)	1991 / 29,9%	Reductions in <Out>plasma estrogen levels</Out> and increases in <Out>bone-resorption markers</Out> were comparable in both groups.
Outcome	OUT.subj + verb.change. active/passive	1336 / 20%	<Out>HRQL</Out> improves after successful treatment.

Future work

- Corpus annotation with the use of current baseline algorithms as pre-annotation
- Employing more advanced linguistic features
- Use of state-of-the-art approaches (word embeddings, neural networks)

Thank you for attention!