Evaluation and Performance Improvement of the BioExcom System for the Automatic Detection of Speculation in Biomedical Texts

Olfa Makkaoui, Julien Desclés and Jean-Pierre Desclés

BioTxtM -May 26  2012

LaLIC-STIH/Paris-Sorbonne
Extracting speculation from biological literature is important for researchers who can be especially interested in finding factual statements in the text.

- Information: certain or speculative.

- Biologists can be also interested in extracting speculation linked for example to a particular entity.

- Authors are not sure about the results and speculation they provide can be a starting point for new experiments.
The SWAN project: aims to collect hypothetical information about the Alzheimer disease in order to use it as discussion subjects between researchers (Ciccarese et al., 2008).

Speculation according to BioExcom:
• Speculation in the biomedical literature is a proposal about a biological issue that is explicitly presented as not certain in the paper.
The BioExcom system

BioExcom aim:

- Automatic annotation of speculative sentences in biological papers.
The BioExcom system

BioExcom aim:

• Automatic annotation of speculation in biological papers.

• Categorization of speculative sentences into «new» and «prior» speculation.
Categorisation of speculation

- Speculation
  - Prior speculation
  - New speculation
The BioExcom system

Categorisation of speculation

Prior speculation:
Speculative sentences cited in the paper, but presented as having been proposed previously.

Aim:
- Detection of the emergence of an idea, which is taken into consideration by the scientific community
- Anticipation of some trends and maybe future results
Categorisation of speculation

Speculation

- Prior speculation
- New speculation
Categorisation of speculation

New speculation:
Speculative sentences presented for the first time in the paper or not explicitly presented as prior speculation.

Aim:
- Reveal some of the real new output of a paper
The BioExcom system

Implementation:

• The Contextual Exploration processing (Desclés et al., 2006) that is based on the search for linguistic markers: indicator (strong markers) and clues (complementary markers) to annotate sentences.

• **Linguistic markers for detection of speculation:**
  - **Verbs:** to suppose, to suggest, to hypothesis, to propose, to assume,…
  - **Nouns:** suggestion, hypothesis, speculation, …
  - **Modality verbs:** may, might, could, …
  - **Adjectives:** convincing, probable, possible, conceivable, …
  - **Adverbs:** possibly, probably, perhaps,…
  - **Conjunction:** if, whether, or
The BioExcom system

Implementation:

Linguistic markers for the categorisation of speculation:

Bibliographic citation: Presence (prior speculation)/absence (new speculation).

Presence of specific words (previous works, recently, in this study,…).
Aim of the paper:

• To use the BioScope corpus (Szarvas and al., 2008) to evaluate BioExcom concerning the detection of speculative sentences task.

• To improve the BioExcom performance using the evaluation results basing on the comparison between BioScope and BioExcom annotations.
BioScope 14500 segments

- Removing tags

2576 annotated segments

Comparison

- Automatic annotation

1830 annotated segments

BioExcom

Divergence annotations

(A) Sentences annotated either by BioScope or BioExcom

(A1) Sentences annotated by BioExcom and not by BioScope

97 spe

17 Not spe

5 Uncharac

397 spe

437 Not spe

(A2) Sentences annotated by BioScope and not by BioExcom

1711 Sentences annotated by BioScope and by BioExcom

12652 Sentences not annotated by BioScope and BioExcom

Converging annotations

(B) Sentences annotated and not annotated by both BioScope and BioExcom

BioExcom performance improvement
Aim of the paper:

• To present a new corpus BioSpe for the annotation of speculative sentences according to the BioExcom definition of speculation
• To present a user interface for the automatic annotation of speculation and their categorization into «new» and «prior» speculation.  www.bio-excom.net
A role for phosphorylation in gating the pore of three distinct aquaporins (PvTIP3;1, GmNod26, SoPIP2;1) has been proposed, based on similar functional analyses in Xenopus oocytes [53], [76] and [78]. In particular, pharmacological alteration of oocyte protein phosphatase and/or kinase activities was able to modulate the water transport activity of these aquaporins and punctual mutations of the putative phosphorylation sites prevented these effects. These studies have established that the loop B and the N- and C-terminal tails of aquaporins can be important domains for water channel regulation.

4.2. Molecular mechanisms of gating

Water transport assays in purified plasma membrane vesicles have indicated that protons and divalent cations can act from the cytosolic face to regulate the water channel activity of PIPs [81] and [82]. These properties, the regulation of PIPs by multiple phosphorylations, and the delivery of SoPIP2;1 atomic structure have definitely established PIPs as a remarkable model for investigating the gating mechanisms of aquaporins and more generally membrane channels [46].

Initial structure–function analyses have indicated that a perfectly conserved His residue in the loop D of PIPs plays a central role in pH sensing and possibly gating [83].
References


Thank you