Discovering contradicting protein-protein interactions in text

Olivia Sanchez-Graillet
Univ. of Essex, Wivenhoe Park, Colchester CO4 3SQ, U.K.
osanch@essex.ac.uk

Massimo Poesio
Univ. of Essex, Wivenhoe Park, Colchester CO4 3SQ, U.K.
DIT and Center for Mind/Brain Sciences, Univ. of Trento, Via Sommarive 14 I-38050 POVO (TN) - Italy
poesio@essex.ac.uk

1 Introduction

In biomedical texts, contradictions about protein-protein interactions (PPIs) occur when an author reports observing a given PPI whereas another author argues that very same interaction does not take place: e.g., when author X argues that "protein A interacts with protein B" whereas author Y claims that "protein A does not interact with B". Of course, merely discovering a potential contradiction does not mean the argument is closed as other factors may have caused the proteins to behave in different ways. We present preliminary work towards the automatic detection of potential contradictions between PPIs from text and an agreement experimental evaluation of our method.

2 Method

Our method consists of the following steps: i) extract positive and negative cases of PPIs and map them to a semantic structure; ii) compare the pairs of PPIs structures that contain similar canonical protein names iii) apply an inference method to the selected pair of PPIs.

We extract positive and negative cases of PPIs by applying our system (Sanchez & Poesio, submitted). Our system considers proteins only as well as events where only one protein participates (e.g. "*PI-3K activity*"). The system produces the semantic interpretation shown in Table 1. We manually corrected some of the information extracted in order to compare exclusively our inference method with human annotators.

The decision to determine if a C-PPI holds is given by the context. This context is formed by the combination of semantic components such as PPI polarity, verb direction, and manner polarity.

P1	Canonical name of the first participant protein
P2	Canonical name of the second participant protein.
Cue-word	Word (verbs or their nominalizations) expressing a PPI (e.g. interact, interaction, activate, activation, etc.).
Semantic Relation	Categories in which cue-words are grouped according to their similar effect in an interaction. (See Table 2).
Polarity	Whether the PPI is positive or negative
Direction	Direction of a relation according to the effect that a protein causes on other molecules in the interaction. (See Table 3)
Manner	Modality expressed by adverbs or adjectives (e.g. directly, weakly, strong, etc.)
Manner Polarity	Polarity assigned to manner according to the influence they have on the cue-word (see Table 4)

Table 1. Semantic structure of a PPI

Semantic Relation	*Verbs/nouns examples*
Activate	Activat (e, ed,es,or,ion), transactivat (e,ed,es,ion)
Inactivate	decreas (e,ed,es), down-regulat(e,ed,es,ion)

Table 2. Example of semantic verb relations

+	-	*Neutral*
Activate, Attach	Inactivate	Substitute, React
Create bond	Break bond	Modify, Cause
Generate	Release	Signal, Associate

Table 3. Directions of semantic relations

Polarity	*Word*
(+) 1	strong(ly), direct(ly), potential(y), rapid(ly)
(-) 0	hardly, indirect(ly), negative(e,ly)

Table 4. Example of manner polarity

Manner polarity is neutral (2) if the manner word is not included in the manner polarity table or if no manner word affects the cue-word.

The method first obtains what we call "PPI state" of each PPI. The PPI state is obtained in two steps that follow decision tables[1]: a) the values for

[1] Some decision tables are omitted due to space reasons.

BioNLP 2007: Biological, translational, and clinical language processing, pages 195–196,
Prague, June 2007.

the combination of the verb direction and the manner polarity (DM) of each PPI; b) then, the DM value and the polarity of the corresponding PPI are evaluated.

Second, the method compares the PPI states of both PPIs as shown in Table 5.

State1	*Sstate2*	*Result*	*State1*	*State2*	*Result*
0	0	NC	3	3	U
0	1	C	0	4	C
0	3	U	1	4	C
1	1	NC	3	4	C
1	3	U			

Table 5. Decision table for results[2]

The following example illustrates our method. The table below shows two sentences taken from different documents.

Document 1	*Document 2*
Cells treated with hyperosmolar stress, UV-C, IR, or a cell-permeable form of ceramide, *C2 ceramide*, rapidly down-regulated *PI(3)K* activity to 10%-30% of the activity found in serum-stimulated control cells…	And fourth, *C2-ceramide* did not affect the amount of *PI 3-kinase* activity in anti-IRS-1 precipitates.

The semantic structures corresponding to these sentences are shown in the next table.

	DocA	*DocB*
P1	C2-ceramide	C2-ceramide
P2	PI-3K	PI-3K
Cue	down-regulate	affect
Semantic relation	Inactivate	Cause
Polarity	positive	negative
Direction	negative	neutral
Manner	rapidly	--
Manner polarity	positive	neutral

The decision tables produced for this example are the following[3].

PPI	*Direction*	*Manner*	*DM*
A	- (0)	+ (1)	- (0)
B	N (2)	N (2)	U (3)

PPI	*Polarity*	*DM*	*State*
A	+ (1)	- (0)	- (0)
B	- (0)	U(3)	NN (4)

PPIA state	*PPIB state*	*Result*
-(0)	NN (4)	**Contradiction**

The result obtained is "Contradiction".

3 Agreement experiment

As a way of evaluation, we compared agreement between our method and human annotators by using the kappa measure (Siegel and Castellan, 1998). We elaborated a test containing only of 31 pairs of sentences (*JBC* articles) since this task can be tiring for human annotators.

The test consisted on classifying the pairs of sentences into three categories: contradiction (C), no contradiction (NC) and unsure (U). The values of kappa obtained are presented in the following table.

Groups	*Kappa*
Biologists only	0.37
Biologists and our method	0.37
Non-biologists only	0.22
Non-biologists and our method	0.19

Table 6 Agreement values

Biologists mainly justified their answers based on biological knowledge (e.g. methodology, organisms, etc.) while non-biologists based their answers on syntax.

4 Conclusions

We have presented a simple method to detect potential contradictions of PPIs by using context expressed by semantics and linguistics constituents (e.g. modals, verbs, adverbs, etc). Our method showed to perform similarly to biologists and better than non-biologists. Interestingly, biologists concluded that C-PPIs are rarely found; nevertheless, the cases found may be highly significant.

Continuing with our work, we will try our system in a larger set of data.

References

Sanchez,O and Poesio,M. (Submitted). Negation of protein-protein interactions: analysis and extraction.

Siegel, S. and Castellan, N.J. (1998). Nonparametric statistics for the behavioral sciences. 2nd. edition, McGraw-Hill.

[2] Result values: contradiction (C), no contradiction (NC) and unsure (U).

[3] The values included in the tables are: positive=1, negative=0, neutral=2, unsure=3, and negative-neutral=4.