Accelerating Chemical Assessments : A Case Study in Automatic Evidence Extraction from Text

Catherine Blake[1] and Jodi Flaws[2]

[1] School of Information Sciences and Department of Computer Science, [2] Department of Comparative Biosciences, College of Veterinary Medicine, University of Illinois at Urbana-Champaign.

Motivation

The manual processes used to extract mechanistic evidence from studies is one of the most time consuming steps when conducting a chemical assessment. Our goal is to automate evidence extraction in order to reduce the time to conduct a review and/or increase the scope of a review.

Method

Explicit claims (Blake, 2010) were identified from abstracts (n=3078) collected in a previous study (Korhonen et all 2012). Prior mode of action annotations were used to identify keywords using Shannon's measure of entropy and expert review. The number of supporting claims, where the MOA has increased (e.g. improve, extend), where there was some effect but the claim is neutral (e.g. change, effect), and where the MOA has decreased (e.g. reduce, inhibited) are reported.

Abstracts from Medline were pre-processed to identify sections and sentences. A dependency parse was generated (Manning, 2014) for each sentence in order to attend to elliptical coordinated compound noun phrases using the method described in (Blake and Rindflesch, 2017).

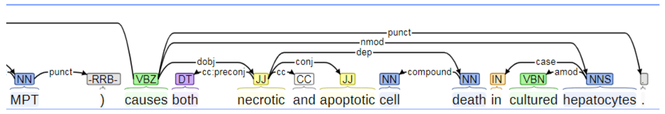


Figure 1 – Partial dependency parse for the sentence from PMID 12700412 "Onset of the mitochondrial permeability transition (MPT) causes both necrotic and apoptotic cell death in cultured hepatocytes.". The coordinated noun phrases would be resolved to produce "apoptotic cell death" and "necrotic cell death"

Explicit claims involve an agent, the nature of the change and an object (Blake, 2010). Explicit claims that include a keyword associated with cell proliferation and cell death were identified automatically. Claims were characterized as supporting if they increased the MOA, neutral if no directionality was provided and refuting if they decreased the MOA. Negation was detected.

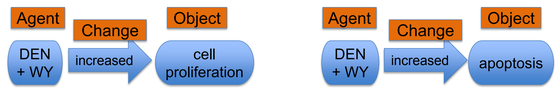


Figure 2 – Explicit claim with MOA from "DEN + WY increased both cell proliferation and apoptosis in both the wild-type and p50 +/- mice; DEN treatment alone has no effect.."

Abstracts were collected from an earlier study (Korhonen et al 2012) that characterized evidence and modes of action for 7 chemicals: 4-aminobiphenyl, asbestos, ethylene oxide, formaldehyde, genistein, methylene chloride, pyridine. This pilot study considers two of the non-genotoxic modes of action reported - cell proliferation and cell death.

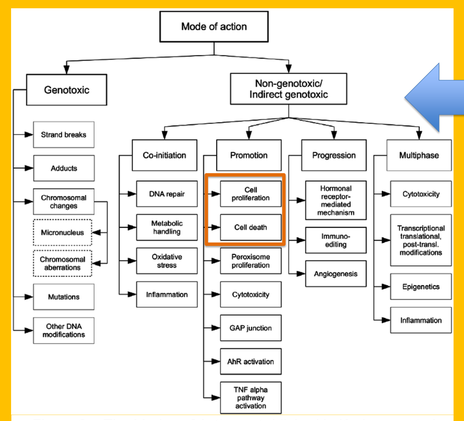
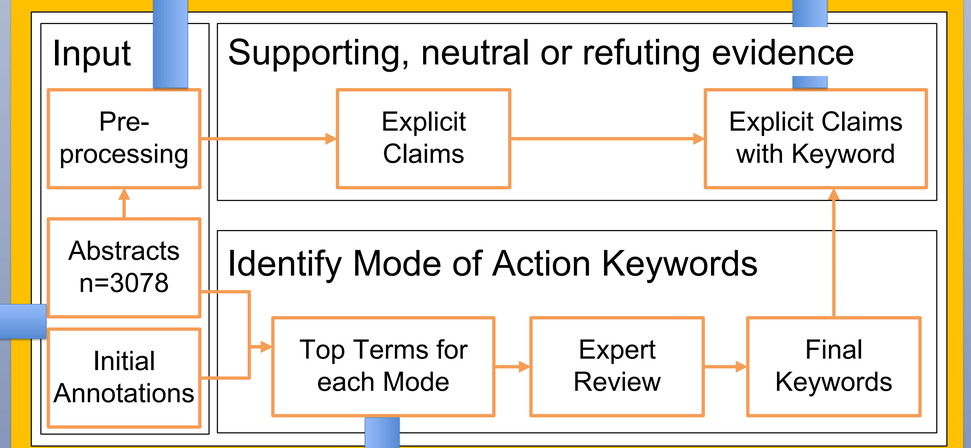


Figure 3 – The Modes of Action established in Korhonen et al (2012). Modes explored in this pilot study are shown in red.



Shannon entropy was used with the initial set of abstracts (e.g. cell proliferation versus not cell proliferation) to establish keywords for each mode of action, which were then reviewed by an expert. Noun phrases that include a keyword were identified in all abstracts from the collection.

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log p(x_i).$$

Results

Refuting evidence
CaN inhibitor cyclosporine A (CsA) reduced cell growth in these cell lines.

Negated supporting evidence
We observed no treatment-related increases in cellular proliferation.

Cell Proliferation

Supporting		Neutral		Refuting	
Neg		Neg		Neg	
8	168	5	53	2	91

Neutral evidence
Binary mixtures of the compounds produced effects on cell proliferation and on each of the responsive protein ions that were fully consistent with concentration additivity.

Of the 327 cell proliferation claims, 99 (30.3%) directly contradicted the premise that cell proliferation increased because the evidence was either negated or reported a decrease.

Just over half (168, 51.4%) of the evidence supports the premise that cell proliferation actually increased.

Negated refuting evidence
Application of PCB 52 in calcium-free medium reduced the calcium accumulation, but did not reduce cell death.

Supporting evidence
Activation of protein kinase G is sufficient to induce apoptosis and inhibit cell migration in colon cancer cells.

Negated neutral evidence
The HDAC inhibitor-induced apoptosis appears to be p53 independent, because no change in apoptotic cell death was observed in H1299 cells that expressed exogenous wild-type p53 (H1299 cells express no endogenous p53 protein).

Cell Death

Supporting		Neutral		Refuting	
Neg		Neg		Neg	
5	416	8	74	3	153

Of the 659 cell death claims, 158 (24.0%) directly contradicted the premise that cell death had increased because the evidence was either negated or reported a decrease in cell death.

About two-thirds (416, 63.1%) of the evidence supports the premise that cell death increased.

Conclusions and Future Work

Results show that simply reporting a mode of action should not be interpreted as evidence that the MOA has increased. Explicit claims from the Claim Framework provide the granularity necessary to differentiate between supporting, neutral, and refuting claims for a given MOA. Further work is required to differentiate between claims made as background knowledge and the results from current experiments.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1535167.

Works Cited

Blake, C., Beyond genes, proteins, and abstracts: Identifying scientific claims from full-text biomedical articles. *Journal of Biomedical Informatics*, 2010. 43(2): p. 173-189.

Blake, C. and T. Rindflesch, Leveraging syntax to better capture the semantics of elliptical coordinated compound noun phrases. *Journal of Biomedical Informatics*, 2017. 72(120-131).

Manning, C.D., et al., The Stanford CoreNLP Natural Language Processing Toolkit in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 2014, Association of Computational Linguistics: Baltimore, Maryland, USA. p. 55-60.

Korhonen, A., et al., Text mining for literature review and knowledge discovery in cancer risk assessment and research. *PLoS One*, 2012. 7(4): p. e33427.

ILLINOIS