Tracking Semantic Change in Medical Information

PI: Ritwik Banerjee
Stony Brook University

http://www3.cs.stonybrook.edu/~rbanerjee/project-pages/eager-satc-2018/tracking.html

Follow the information. Play the Telephone game.

My vision is to combine aspects of knowledge representation and syntactic stylometry from computational linguistics to capture intrinsic properties of how information morphs as it propagates. There are critical gaps in

a) the study of gradual distortion over time of natural language information over multiple documents,

b) domains other than political news,

c) computational models of misinformation beyond true/false classification, and

d) computational models that do not rely on human judgments of true/false.

News is not just true or false – information distorts due to semantic changes in the narrative.

- *Particularly* true for medical findings, which are nuanced and highly specific, but media coverage often glosses over the details and distorts the original message.

Consequences

- Deviation from scientific findings related to health.
- Quality of information suffers, and may corrupt any data-driven field that relies on available information to create datasets, test new algorithms, and learn patterns.
- A large population with unfounded expectations about health-related issues, and/or harmful lifestyle choices.

Approach

A story-focused dataset:

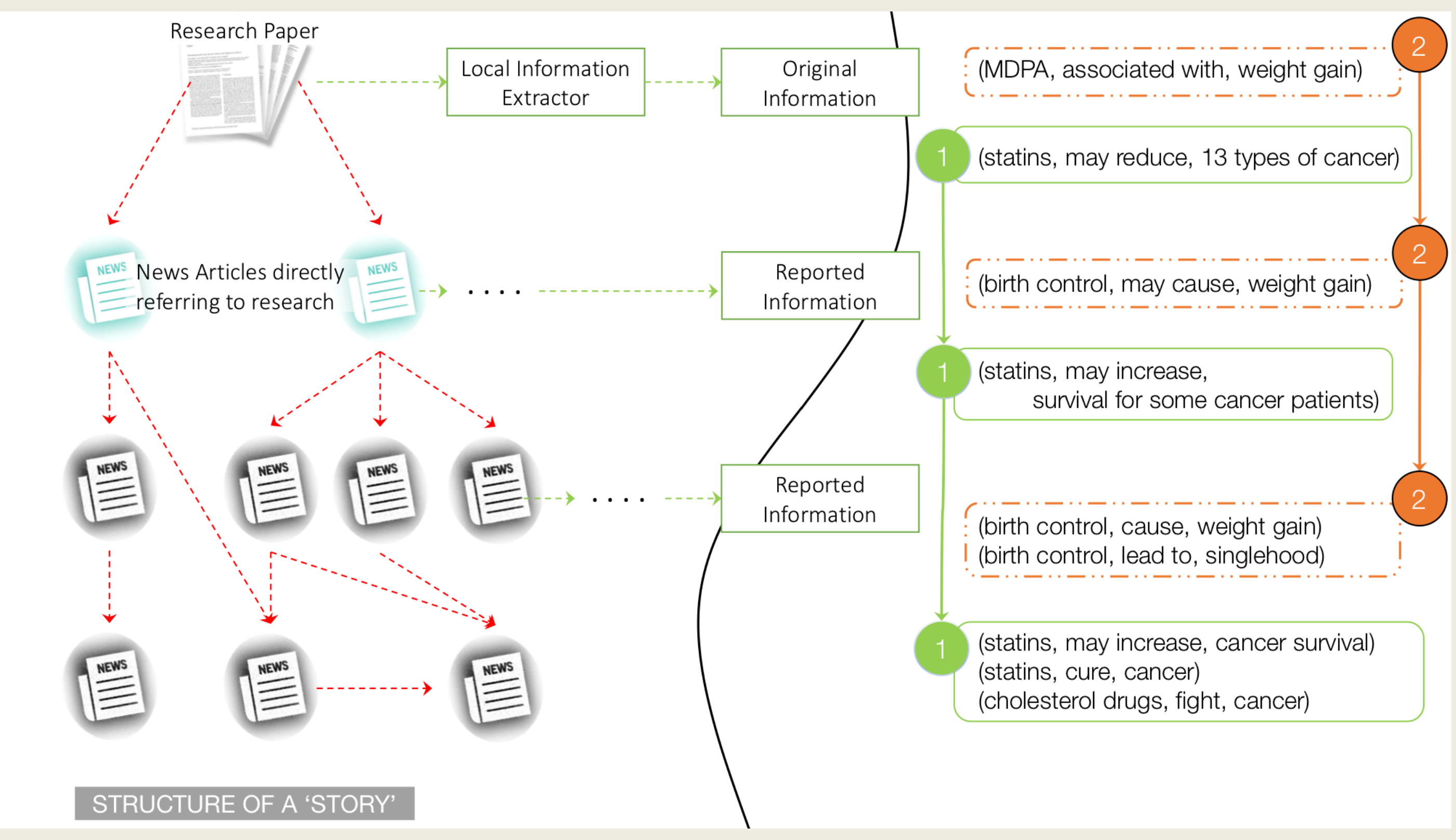
- Collect medical news as “stories” starting with a research publication, and developing over time, spanning multiple news articles (starting with those directly citing the research).

Dynamic event representations:

- Develop event representations to capture the temporal dynamics of a narrative, *i.e.*, where event parameters may change over time.

Analysis of information distortion:

- Machine learning tasks to exploit stylometry and event representations to identify/categorize information distortion.
- Quantitative measures of distortion in natural language.


Research Paper
Local Information Extractor
Original Information
News Articles directly referring to research
NEWS
. . . .
Reported Information
Reported Information
2
(MDPA, associated with, weight gain)
1
(statins, may reduce, 13 types of cancer)
2
(birth control, may cause, weight gain)
1
(statins, may increase, survival for some cancer patients)
2
(birth control, cause, weight gain)
(birth control, lead to, singlehood)
1
(statins, may increase, cancer survival)
(statins, cure, cancer)
(cholesterol drugs, fight, cancer)

STRUCTURE OF A ‘STORY’

Challenge: Event extraction technology is in its infancy, especially from medical language.

Scientific Impact

+ Open information extraction. In particular, advance state-of-the-art in event extraction.
+ Knowledge base design.
+ Information evolution in cyberspace.

Education & Outreach

› PI mentoring: (i) three undergraduate research projects (linguistics and computer science), (ii) twelve graduate research projects (computer science), and (iii) two doctoral students (computer science).

› Student presentation at the 2018 Conference and Labs of the Evaluation Forum (CLEF) at Avignon, France.

Social Potential

+ Assistive technology to improve quality of information (*e.g.*, by notifying readers of misrepresented claims).
+ Benefits journalism with tools for quantitative longitudinal analysis of information.

Publications

› C. Zuo, A. Karakas, and R. Banerjee. **A Hybrid Recognition System for Check-worthy Claims Using Heuristics and Supervised Learning**. In *Working Notes of CLEF 2018 – Conference and Labs of the Evaluation Forum, CLEF – Vol. 2125*. CEUR-WS, **2018**.

› C. Zuo, A. Karakas, and R. Banerjee. **To Check or not to Check: Syntax, Semantics, and Context in the Language of Check-worthy Claims**. In Crestani et al. (Eds.) *Experimental IR Meets Multilinguality, Multimodality, and Interaction: Proc. 10th Int. Conference of the CLEF Association, CLEF – LNCS Vol. 11696*. Springer, **2019**.

Research in Progress

› Using citations to identify reified claims in text

› Identifying event contexts and parameters in language.

› Information propagation with and without distortion.

› Identifying event parameter changes in language.

Award ID#: SES-1834597