SEMANTIC DISTORTION IN MEDICAL INFORMATION

Chaoyuan Zuo, Aakash Bhatia, Nitinkumar Bhyri, Ritwik Banerjee
[chkzuo, asbhatia, nbhyri, rbanerjee]@cs.stonybrook.edu

MOTIVATION
Misinformation is widespread. The meaning of a claim made by a medical news article may change as it passes through the internet.

Our goal is to introduce a novel dataset with semantically distorted medical news claims and study the performance of standard and deep learning-based multi-class classifiers on this dataset.

Such a data-set has not been introduced yet, making our dataset one of a kind.

This data-set has widespread applications in journalism and limiting the spread of misinformation via the internet. It can help us identify the kinds of changes a news claim may undergo. It can be used to detect differences in the news reported by various news outlets and detect distorted news claims spread by people via social media.

DATA-COLLECTION WORKFLOW
Collected 10,507 unique medical news articles from 11 major news websites using a web-scraper. Extracted and cleaned the headlines from these articles to obtain the “claim” made by the news article.

7 distorted versions of each claim will be crowd-sourced via Amazon Mechanical Turk to create a labelled dataset.

The complete data-set will contain approximately 70,000 labelled, semantically distorted news claims.

DISTORTION DEFINITIONS
Paraphrase - Altering the claim without changing its meaning.
Generalization - Replacing a term(s) with more general terms.
Specification - Replacing a term(s) with more specific terms.
Hyperbole: Altering claims from a weaker to stronger meaning.
Meiosis: Altering claims from a stronger to weaker meaning.
Negation - Negating the meaning of a claim.
Unrelated change in entity/relation - Replacing an entity or relation in a claim with a completely unrelated one.

EXAMPLES OF DISTORTED CLAIMS

| Claim: Ramping up exercise tied to lowered heart disease risk in older adults. |
| Paraphrase | Increase in the amount of exercise has been proven to lower risk of heart disease in older adults. |
| Generalization | Increase in the amount of exercise has been proven to lower risk of diseases in people. |
| Specification | Increasing exercise frequency by one-third has been proven to lower risk of heart disease in adults over the age of 45. |
| Hyperbole | Increase in the amount of exercise has been proven to eliminate the risk of heart disease in older adults. |
| Meiosis | A mild increase in the amount of exercise has been proven to reduce the risk of heart disease in older adults. |
| Negation | There is no evidence to suggest that increasing the amount of exercise can lower risk of heart disease in older adults. |
| Unrelated change in entity/relation | Increase in the amount of exercise has been proven to lower risk of muscular stiffness in older adults. |

DISTRIBUTION OF DATA

Counts of Websites

Websites