Substance Abuse Knowledge and AI Systems

Rao Wasim, Avaanika Nair, Damir Cavar, Danny Valdez

NLP-Lab, Indiana University, Bloomington, IN 47405, USA

School of Public Health, Indiana University, Bloomington, IN 47405, USA

Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN 47405, USA

Introduction

- Social Media Surveillance of Drug Abuse
- Identification of
  - New street names for drugs
  - New adverse reactions
- Existing data
  - Tweets scraped prior to the closure of the Twitter API
  - YouTube Comments
  - BlueSky
  - Reddit
- NIDA A-Z Drug Dictionary
- DEA Slang Terms and Code Words for Drugs
- Issues:
  - Social Media Posts use coded language and (regional) slang
- Missing Tech
  - Natural Language Processing tools
  - Classifiers

Objective

Build an OWL ontology of NIDA drug terminology and slang to improve LLM-based detection and classification of substance-abuse content on social media, detecting new slang terms.

Methods

- Conversion of NIDA Drug Dictionary to a Knowledge Graph and Description Logic OWL (RDF-based ontology), including
  - Pharmacological names
  - Slang terms
  - Brand names
  - Emojis
- Human annotation of social media posts
- Training and evaluation of supervised classifiers for "post talks about drugs or not"
  - BERT Tweet
  - Llama 3.2
- Scraping other sources
  - Dictionaries about drugs
  - Knowledge sources
- Linking to other existing drug (abuse) related ontologies
  - The Drug Ontology
  - DrOn
- Use LLMs to extract:
  - Concepts and Relations from definitions and descriptions
  - Generating sub-graph relations and hierarchies by pulling hypernyms

Results

- NIDA Drug Ontology V. 1
  - Full hierarchical organizations of concepts by hypernyms
  - Multi-lingual base with English and Spanish
  - Concept relations expressing:
    - Usage
    - Health risks
    - Treatments
    - Effects (Adverse reactions)
- Social Media Post Classifiers
  - Feature Selection and Identification using Logistic Regression
  - Binary classifiers: +/- is about drugs
  - BERT Tweet: F1 score .935
- Corpus of social media posts
  - 890,000 social media posts
  - Selected human annotated posts for binary classification
- Grant submission'
  - Submitted to an RO1 NIH proposal
  - One talk at AI2 at IU Indianapolis
  - Cooperation with Public Health
- Tests
  - LLMs using the NIDA Ontology

Discussion and Future Work

- Experiments with LLMs and Ontology
  - Identification of all new slang words and emoji for drugs
  - Discover newly reported adverse reactions in social media
  - Identify new drugs based on context (e.g., Kraton)
  - Extend the NIDA Ontology
- Add new languages to NIDA Ontology

References

NIDA A-Z Drugs: https://nida.nih.gov/research-topics/drugs-a-to-z

The Drug Ontology: https://bioportal.bioontology.org/ontologies/DRON

DEA Slang Terms and Code Words: https://www.dea.gov/sites/default/files/2018-07/DIR-022-18.pdf

Acknowledgements

- PIs
  - Dr. Damir Cavar (NLP-Lab)
  - Dr. Danny Valdez (Public Health)
  - Shruthi Nagappan (NLP-Lab)
  - Anuj Prakash (NLP-Lab)