NLP5: Identifying Toxic Individuals Online (IdenTox)
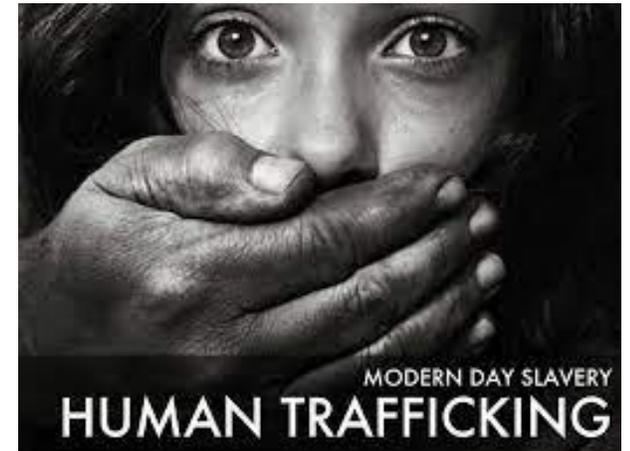
Background

Human Trafficking is a lucrative organized crime that majorly includes sex trafficking and puts many human beings in misery with lifelong traumas. Human trafficking has found on the internet an ideal environment to cultivate more material and potential predators through today's digital communities. Every day prospective abusers find new ways to hide their actions through dubious online advertisements and hidden messages.

Goal

• Design an NLP pipeline for detecting forms of human trafficking through online ads.

Tasks

• Literature review on human trafficking and NLP methods (not many use recent Transformers)

• Provide an empirical evaluation of NLP baselines for the task (Transformers, Word2Vec, tf-idf + SVM)

• Find deficiencies in the baselines

    • Here we can analyze token processing, tokens, embedding relation in vector space, edge cases

    • (One example would include resolving indicators for hidden messages such as "c@$H" ⇔ "cash")

• Devise novel modules targeting these deficiencies

• Run an ablation study on the new modules

Jan Wahle
wahle@uni-wuppertal.de

Terry L. Ruas
ruas@uni-wuppertal.de

Background

Every year, millions of people worldwide are victims of sexual abuse, and physical violence. According to police crime statistics, most (>50%) sexual abuse cases are committed by a known and trusted person, while strangers are rare. Therefore, the most prominent indicator for detecting sexual abuse lies in analyzing conversations between predators and victims. Online text conversations are frequently available while spoken conversations are rare. This project aims to target both domains, online textual conversations as well as spoken conversations to find a relation between the two which can be leveraged for transfer learning from the higher resource to the low resource case.

Goal

- Design and implement an NLP approach to detect potential sexual abuse from online conversations and transfer the acquired knowledge to a small set of real conversations.

Jan Wahle
wahle@uni-wuppertal.de

Tasks

- iterature review on Predatory Conversation Detection.

- Provide an empirical evaluation of NLP baselines for detecting textual (online) conversations.

- Identify deficiencies in the baseline approach

- Iterate over the baseline approach targeting the known deficiencies and provide an ablation study for each component.

- Test the generalization of the trained models for spoken conversations.

Terry L. Ruas
ruas@uni-wuppertal.de

Background

Many machine learning algorithms and verification procedures rely on high-quality data. Supervised machine learning requires a human assessment of the task (the ground-truth), for example about the sentiment of Twitter posts. Language models use self-supervision and receive their ground-truth by removing words from natural text and try to predict the same words again. For both applications, the quality of text is crucial. One of the most recent language models, GPT-3, uses a machine learning classifier to find high-quality data using a set of known high-quality text documents.

Goal

- Develop a machine learning model to select high-quality data based on a gold-standard.

Tasks

- Review literature about NLP datasets and contextual word embeddings (Transformers)

- Establish a gold-standard dataset for the task of human trafficking.

- Train a machine learning classifier to find similar articles to the gold-standard.

- Select text from a larger pool of potential human trafficking text and do a sample analysis on the quality of the automatically selected data.

Jan Wahle
wahle@uni-wuppertal.de

Terry L. Ruas
ruas@uni-wuppertal.de

Background

The more data we have, the more we need to analyze them. Text Summarization is an NLP downstream task in which one creates a shorter version of a given input (or many) containing the most important aspects of it. Aside from being a scientific downstream task, text summarization can be applied to many scenarios such as meetings notes, scientific papers, bug reports, books, news, blog posts, TLDR, and transcripts. At a higher level, summarization can be extractive (unaltered) or abstractive (semantic). We want to investigate the summarization problem through new architectures and systems, possibly for low resource languages.

Goal

• Explore Text Summarization task (Extractive/Abstractive) [low resource languages]

Tasks

• Review literature about NLP approaches working with text summarization (DL, GA?)

• Establish which models and datasets are being used to exercise this task

   • Applied or Benchmark

• Identify (dis)similar between these approaches

• Propose training architecture, data, or paradigm to compete with SOTA systems

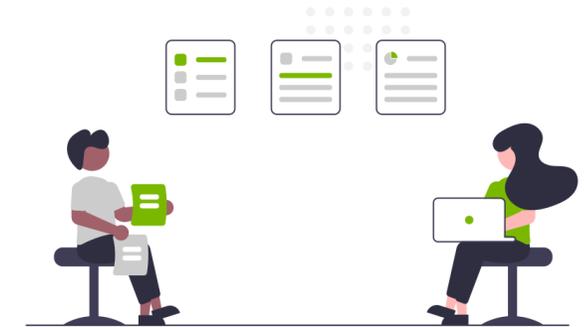• Evaluate approach in known datasets

Jan Wahle
wahle@uni-wuppertal.de

Terry L. Ruas
ruas@uni-wuppertal.de

NLP9: Who needs two-hour meetings? Meeting summarization using transcripts.

Background

An increase in the number of online meetings made clear that typically meetings only have few key topics and a limited amount relevant information for all participants. Therefore, the extraction of their key topics and their summarization became more obvious. Although text summarization has been long explored by the natural language processing (NLP) community, its application to meetings and dialogs is still incipient. Meetings differ from traditional text as their structure is often dynamic. The interaction between multiple participants (e.g., discussions), their deviant formats, irregular sequences, different semantic styles, and topics promote a complex scenario. Short meetings can easily reach thousands of tokens in just a few minutes of conversation. Thus, techniques that produce high quality meeting summaries, including the most important ideas discussed between its participants, are still necessary.

Goal

- Explore Text Summarization task (Extractive/Abstractive) [low resource languages]

Tasks

- Review literature about NLP approaches working with text summarization and meeting summarization.

- Establish which models and datasets are being used to exercise this task

- Propose training architecture, training data, or paradigm.

- Integrate solutions in our repository and contribute to state-of-the-art solutions.

- Evaluate approach in known datasets

Jan Wahle
wahle@uni-wuppertal.de

Terry L. Ruas
ruas@uni-wuppertal.de