

Relevant Sentence Selection for Harvesting Loss and Damage Information of Disasters from Online News

Assoc/Prof Mao Kezhi
Mr Li Qi

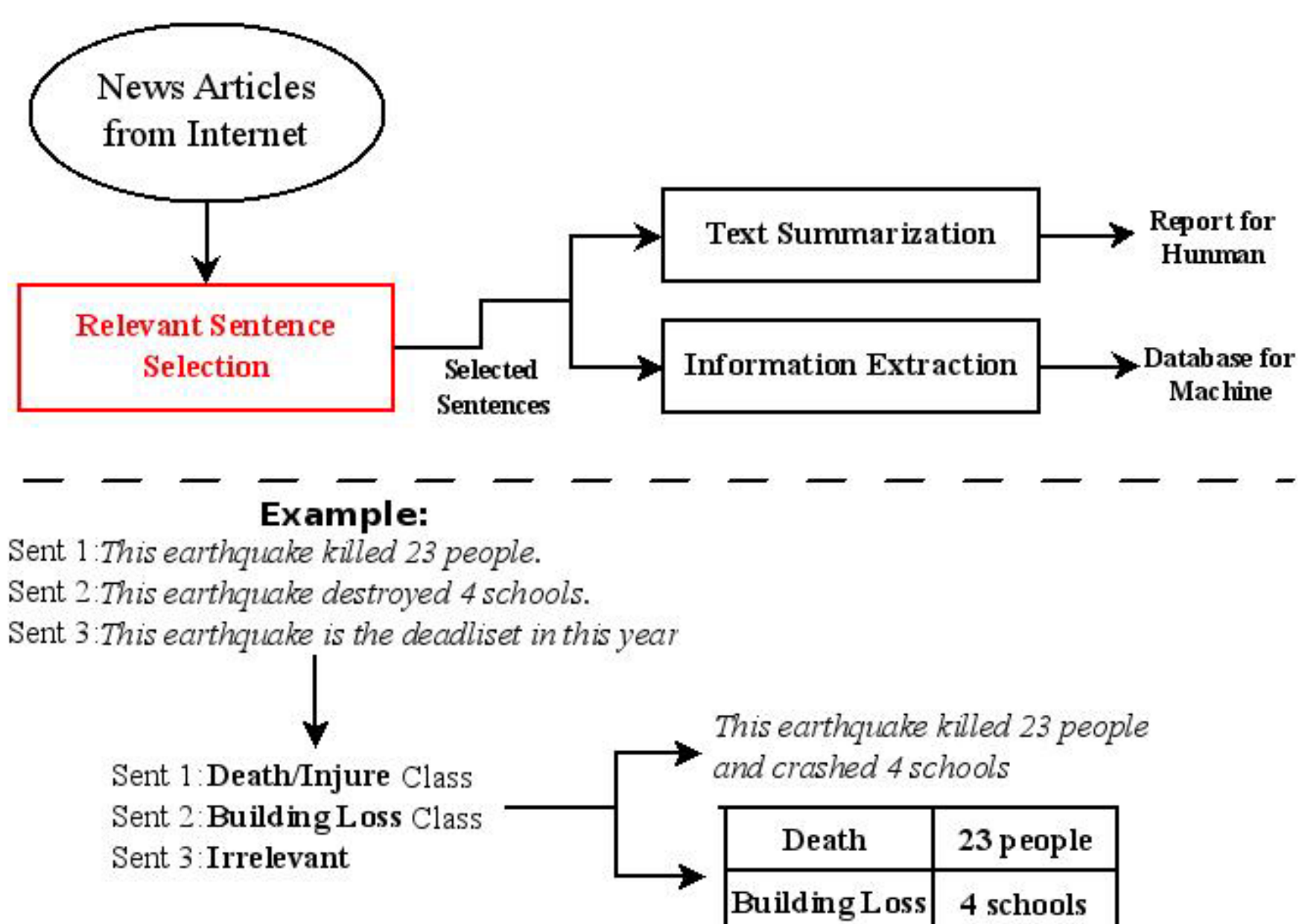
Objectives

Studies show that online news and social media contain rich information on disaster loss and damage. Relevant sentence selection is the preliminary module for the information collection system. In light of features in news articles related to disasters, we propose Recursive Data-Pruning Convolutional Neural Network (ReDP-CNN) to do the sentence selection.

System Architecture

This system consists of a few components, including relevant sentence selection, information extraction, and text summarization.

- **Relevant Sentence Selection**
 - Pick relevant sentences from documents and assign a label
- **Information Extraction**
 - Extract useful information from sentences for machine to analyze
- **Text Summarization**
 - Generate disaster summarization for human to read



Relevant Sentence Selection

We propose Recursive Data-Pruning Convolutional Neural Network (ReDP-CNN) to do the relevant sentence selection. News contains many task-irrelevant words which are useless to the sentence selection. Hence, if these task-irrelevant words are eliminated from the dataset, simple convolutional neural network can be trained better without increasing the network complexity. This model is mainly composed of three parts:

- **Convolutional Filter Evaluation**
 - Evaluate convolutional filters based on the discriminative power
- **Training Data Pruning**
 - Pruning the training data based on the evaluated convolutional filters
- **Convolutional Filter Fine-tuning**
 - Fine-tune the convolutional filters to be more discriminative

