

Textual Entailment on Large Datasets

Presented By,
SUBALAKSHMI SHANTHOSI S
(186001008) M.E(CSE)

Supervised By,
Dr. Chandrabose Aravindan

REVIEW-1
AUGUST 08,2019

- Introduction
- Motivation
- Problem statement
- Literature Survey
- Existing System
- Proposed Work
- Dataset Description
- References

Textual Entailment:

- Textual entailment (TE) in natural language processing is a directional relation between text fragments. The relation holds whenever the truth of one text fragment follows from another text.
- In other words, Textual entailment is the task of determining whether a hypothesis is true, given a premise.
- Textual Entailment relations:
 - text: If you help the needy, God will reward you.
 - ① A *positive TE*(text entails hypothesis):
hypothesis: Giving money to a poor man has good consequences.
 - ② A *negative TE*(text contradicts hypothesis):
hypothesis: Giving money to a poor man has no consequences.
 - ③ A *non TE*(text does not entail nor contradict):
hypothesis: Giving money to a poor man will make you a better person.

Catchphrase extraction:

- Catchphrases are short phrases within the text of the document which can be extracted by selecting certain portions from the text of the document.
- Catchphrases act as a tool to find similarity between two documents.
- Thus, in our project, we work to retrieve suitable prior case for the given current case and to extract catchphrases for a given document.

- It is critical for legal practitioners to find and study previous court cases,so as to examine how the ongoing issues were interpreted in the older cases.
- Thus,an automated precedence retrieval approach is desired.
- It is essential for legal practitioners to have a concise representation of the core legal issues described in a legal text.
- Thus,an automated catchphrase extraction approach is needed.

Problem Statement

- Given a set of current case documents and a set of prior case documents, the system will extract the catchphrases present in the documents and retrieve the relevant prior case documents for all the current case documents.

Distributed Representation in Information Retrieval[1]

- The given training and test documents are represented as vectors using Doc2Vec.
- Cosine distance between the current and prior document vectors are measured and ranked.
- To extract the catchphrases, cosine distance between the catchphrase vector and the document vectors are calculated.
- Based on the calculated distance the prior cases and catchphrases are ranked.

Catchphrase Extraction from Legal Documents Using LSTM Networks[2]

- The problem is formulated as a classification task and the objective is to learn a classifier using LSTM network.
- The proposed methodology involves a pipelined approach and is divided into four phases:
 - Pre-processing
 - Candidate phrase generation: n-grams with n in range 1 to 4 were created from the text.
 - Create vector representations for the phrases
 - Training a LSTM network

Catch Phrase Extraction From Legal Documents Using Deep Neural Network[3]

- For each file, a set of potential meaningful phrases were created and then are classified using deep neural network.
- The proposed steps are preprocessing,create potential meaningful phrases based on common grammar of phrases,feature selection,label the vectors,classification and training the model.

A Text Similarity Approach for Precedence Retrieval from Legal Documents[4]

- The given text are all preprocessed.
- Linguistic features were extracted from all the legal documents.
- Feature vectors were constructed for the documents using TF-IDF score or Word2Vec.
- Cosine similarity score between each current case with all the prior cases was calculated.
- Prior cases were ranked based on the similarity score for each current case

Literature Survey (Contd...)

A Deep Network Model for Paraphrase Detection in Short Text Messages[5]




- The proposed paraphrase detection model is composed of two main components,i.e.,pair-wise word similarity matching and sentence modelling.
- The pair-wise similarity matching model is used to extract similarity information between pairs of sentences.
- The features generated by the convolutional layer have the form of n-grams, and are fed into the LSTM.
- Sentence modelling component is able to process sequential input with the aim to learn the long-term dependencies in the sentences.
- Then,the two sentences are matched and a similarity matrix is generated.
- The CNN is applied onto the similarity matrix to learn the patterns in the semantic correspondence between the two sentences.

- The existing approaches have used cosine similarity approach to retrieve the suitable prior case for the given current case and then rank the prior case documents[1][4].
- Deep neural network have been proposed to detect and extract catchphrases[2][3].
- Deep learning approach have been used to detect the paraphrase by a hybrid deep neural architecture which composed of CNN and LSTM model[5].

- In our work, the catchphrases from documents will be extracted and suitable prior case for the given current case will be retrieved using deep learning approach.

Dataset Description

- Forum of Information Retrieval Evaluation in 2017(Fire2017 IRLed)
- Precedence Retrieval:
 - The dataset consists of 200 current case which is formed by removing the links to the 2000 prior case and the prior case which have been cited by the case in current case.
- Catchphrase Extraction:
 - The dataset consists of 100 documents and their corresponding gold standard catchphrases for training and the test set consists of 300 separate documents whose catchphrases were to be found.

-  Reshma, U and Kumar, M Anand and Soman, KP, *Distributed Representation in Information Retrieval-AMRITA_CEN_NLP@IRLeD 2017*. FIRE (Working Notes) 69–71, 2017..
-  Bhargava, Rupal and Nigwekar, Sukrut and Sharma, Yashvardhan, *Catchphrase Extraction from Legal Documents Using LSTM Networks*. FIRE (Working Notes) 72–73, 2017.
-  Das, Sourav and Barua, Ranojoy, *Catch Phrase Extraction from Legal Documents Using Deep Neural Network.*, FIRE (Working Notes) 78–79, 2017.

-  Thenmozhi, D and Kannan, Kawshik and Aravindan, Chandrabose, *A Text Similarity Approach for Precedence Retrieval from Legal Documents.*, FIRE (Working Notes)90–91,2017.
-  Agarwal, Basant and Ramampiaro, Heri and Langseth, Helge and Ruocco, Massimiliano, *A Deep Network Model for Paraphrase Detection in Short Text Messages* ,arXiv preprint arXiv:1712.02820,2017.

THANK YOU