How Alignment Impact Language Learning Tasks? A Visual Analytics Tool for Interpreting, Debugging, and Tunning Text Entailment Model

Abstract— With the advances of deep learning, neural network based models have reset the state-of-the-art for nearly all linguistic tasks in natural language processing. The disruptive advance also brings enormous challenges. The opaque nature of neural network model leads to hard to debug system and difficult to interpret mechanism cite. In this work, we use textual entailment task as an example to illustrate how a flexible visualization system can help NLP researchers quickly identify the potential limitation of a models and probe the model inner states for interpreting key mechanism such as attention.

Index Terms—Natural Language Processing, Interpretable Machine Learning, Visual Analytics

- Depth learning vis
- Vis

How to interpret classification task

- What feature / dimension / part-of-the-input should change to alter the outcome
- What Examples can you provide that is similar to the current instance

Papers:

- Interacting with Predictions: Visual Inspection of Black-box Machine Learning Models, Use partial dependency (varying along one feature and observer how prediction changes)
- Why Should I Trust You? Explaining the Predictions of Any Classifier (LIME)

- what is the textaul entailment problem
- the importance of textual entailment problem
- how easily can the visualization method extends other NLP problem

Textual entailment is one of major natural language tasks, namely translation, summarization, question

- how the visualization approaches generalizes to other NLP task
- limitation

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx