

Models and Muggles: An Evaluation of the PASCAL Textual Entailment Challenge

The PASCAL Recognizing Textual Entailment (RTE) Challenge re-ignited interest in automatically identifying Textual Entailment relations. *Textual entailment* is a directional relationship between pairs of sentences. The first sentence is defined as the *Text* and the second as the *Hypothesis*, denoted by T and H respectively. T is said to entail H if the meaning of H can be inferred from the meaning of T, “as would typically be interpreted by people”.

For example:

- T The incident in Mogadishu, the Somali capital, came as U.S. forces began the final phase of their promised March 31 pull out.
- H The capital of Somalia is Mogadishu.

The task of identifying TE relations is applicable in multiple disciplines within Natural Language Processing such as **Question Answering**, **Multidocument Summarization** and **Machine Translation evaluation**.

The data sets for training and evaluation were criticized as not being representative of real world examples based on the instructions given to annotators. Several participants queried whether the task was achievable by humans let alone computers..

An investigation into the validity of the PASCAL data was carried out by surveying Melbourne University students. The survey was designed to elicit whether the sentence pairs could be classified as TRUE or FALSE by people who don't specialize in the field.

Human evaluation overall obtained 54% true positive agreement in the absence of world knowledge.

While the Pascal challenge presents a necessary forum for systems from a broad range of disciplines to exchange ideas, we believe that the data is too diverse for a human to classify and thus it is unrealistic to expect a system to solve the problem.