Microsoft Research at RTE-2: Syntactic Contributions in the Entailment Task: an implementation

RTE - Recognizing Textual Entailment, focuses specifically on Hypothesis and Entailment between a pair of sentences.

This particular system, M(icrosoft) ENT(ailment) uses Sytactic features and a thesaurus. They do this:

- Dependency Parse the pair of sentences
- Align the tree's nodes according to grammar heuristics
- Check for predefined entailment models for a syntax level comparison
- If no syntactic model matches, back off to a lexical similarity model

Some points from their grammar heuristics and lexical similarity model, that we tried to implement too -

Exact and Synonym match
Align a node in the first tree to any node in the second tree that has both the same part of speech and either words are identical, or belong to the same synset in WordNet.

Antonym Match
If two aligned noun nodes (h1,t1) are both subjects or both objects of verb nodes (h0,t0) in their respective sentences, then they check for an antonym match. They construct the set of verb antonyms using WordNet.
(for us, we completely avoid finding relations by using Karakas)

Some points from their grammar heuristics, that we couldn't implement -

Acronym Match, Numerical Value Match, Superlative Mismatch etc.

Using Maximal Embedded Syntactic Subtrees for Textual Entailment Recognition

They use “tree mining and matching” technique. As an abstract view, they look at dependency trees similar to a directed acyclic graph which is labelled. Then they look at the structure of the tree and how much of it can be overlapped. Linguistically, if one sentence had a subordinate clause then they would try to find a similar structure in the other sentence as well, because they both look like “nested trees”.

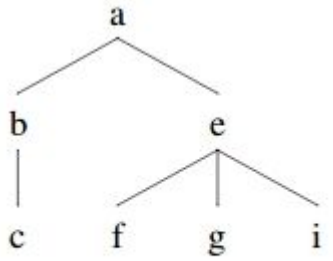

Figure 1: Tree A

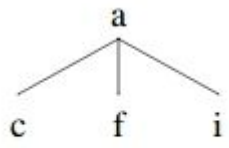

Figure 2: Tree B

Here, there is a high chance that Tree B’s sentence’s meaning is contained in Tree A’s meaning i.e. Entailment. Since, efgi and acfi look similar and efgi is a “nested” structure in Tree A.

They did two runs of this (way more detailed) method, one with realising semantic features and one without. The one which doesn’t use lexical relations has higher precision, like per case, if a result is positive then it as to be high in percentage. But the run with semantic features actually provides better results overall.

Things we derived -

Since the entire paper heavily bases on a good dependency parser, and we could not, we could just use the fact that the “tree branches” could be in a reordered state in Indian languages, and thus we added the PSP feature with our Karaka relations (explain the PSP part)

XNLI: Evaluating Cross-lingual Sentence Representations

Here are some examples from the XNLI dataset :

Language	Premise / Hypothesis	Genre	Label
English	You don't have to stay there. You can leave.	Face-To-Face	Entailment
French	La figure 4 montre la courbe d'offre des services de partage de travaux. Les services de partage de travaux ont une offre variable.	Government	Entailment
Spanish	Y se estremeció con el recuerdo. El pensamiento sobre el acontecimiento hizo su estremecimiento.	Fiction	Entailment
German	Während der Depression war es die ärmste Gegend, kurz vor dem Hungertod. Die Weltwirtschaftskrise dauerte mehr als zehn Jahre an.	Travel	Neutral
Swahili	Ni silaha ya plastiki ya moja kwa moja inayopiga risasi. Inadumu zaidi kuliko silaha ya chuma.	Telephone	Neutral
Russian	И мы занимаемся этим уже на протяжении 85 лет. Мы только начали этим заниматься.	Letters	Contradiction
Chinese	让我告诉你，美国人最终如何看待你作为独立顾问的表现。 美国人完全不知道您是独立律师。	Slate	Contradiction
Arabic	تحتاج الوكالات لأن تكون قادرة على قياس مستويات النجاح. لا يمكنالوكالات أ اتعرف ما إذا كانت ناجحة أم لا	Nine-Eleven	Contradiction

Table 1: Examples (premise and hypothesis) from various languages and genres from the XNLI corpus.

Then the full dataset summarised looks something like this :

sentence1	sentence2	gold_label
और उसने कहा, माँ, मैं घर आया हूं।	जब ही उसकी स्कूल बस ने उसे उतरा उसने अपनी माँ को बुलाया	neutral
और उसने कहा, माँ, मैं घर आया हूं।	उसने एक शब्द नहीं कहा।	contradiction
और उसने कहा, माँ, मैं घर आया हूं।	Uski maa ne bataya ki wo ghar pahuch gaya	entailment
मुझे नहीं पता था कि मैं क्या कर रहा था या कुछ भी, इसलिए वाशिंगटन में किसी निर्दिष्ट स्थान पर रिपोर्ट करना था ।	मैं बिल्कुल जानता था कि मुझे वाशिंगटन जाने के लिए क्या करना है।	contradiction
उसे जाने नहीं मिला	उसे संग्रहालय के उद्घाटन में जाने की इजाजत नहीं थी।	neutral
और मै तो था की भई ठीक है , और बस वही था ।	मैंने कहा हाँ के बाद, यह समाप्त हो गया।	entailment
और, उह, वे वास्तव में परिवार का दौरा बंद कर देते हैं क्योंकि वे केवल, केवल निर्धारित कर चुके हैं कि वे वाईट होने जा रहे हैं।	वे हर दिन यात्रा करते रहे।	contradiction

We used a similar system to prepare our dataset (show dataset) and added separate columns for binary marking whether it's an Entailment, Paraphrase, Contradiction or Neutral pair. Then we compare Accuracy on the Gold Label column. Our Dataset was initialized with 32 pairs, crafted to check a lot of cases, then we make a n*n sized sized one.

Basic Dependency Parser in NLI

The paper starts off by describing parsing in detail, how a parser breaks down a sentence to its constituents, and how grammar, parts of speech and syntactic relations affect it. It defines what tokens are and how "sequences" of rules are important for a parser. Then it gives a brief introduction to Part of Speech and Ambiguities involved.

It describes the problem as follows:

> When sentences are fed into the parser, in most cases it gives incorrect output. It is quite difficult to use those dependency trees in the inference system, and there is little or nothing to be done regarding the output from the parser.

After some definitions for beginners it moves to describe Dependency Parsing, what goes in as an input, POS tags, relations between tokens etc. and how the Stanford parser generates it.

This completes the starting off point for someone trying to understand the basics of what we did in our project, then it talks about Normalization approach, which we haven't used and finally closes off with Related works, like Shallow Inference (which doesn't take semantics into account) and then about textual entailment and syntactic graph distances, which are alternative models to our approach here.