A Discriminative Model for Tree-to-Tree Translation (*Original*)

Brooke Cowan, Ivona Kucerova, Michael Collins EMNLP 2006

Discriminatory model-to-Tree Translation Tree *(via Estonian)*
A model of discrimination for Tree-to-Tree Translation (*via French*)
The distinctive pattern of the tree to tree translation (*via Lithuanian*)
An insightful model tree to tree Lyrics *(via Macedonian)*
For a distinctive tree-for-Tree Translation Model (*via Turkish*)

Presented by
Martin McRoy and Ruben Sipos

Motivation / Related Work

Approaches
- Statistical machine translation (Google)
- Rule based (Worldlingo ➡ Microsoft Windows)
- Example based

(Och and Ney, 2004) Phrase-based approach
(Alshawi, 1996) (Wu, 1997) Transductive grammars

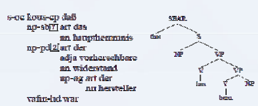
Goal: Learn a model that maps parse trees in the source language to parse trees in the target language

- Clause by clause
- German vs. English syntax
- General approach can be applied to different set of languages

Process Overview

Sentence

Parse into clauses



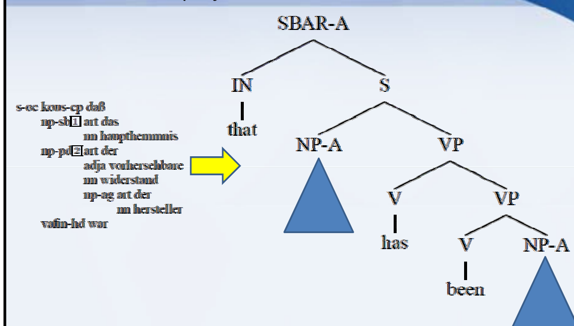
Perceptron

Predict AEP

Translate Modifiers

Combine

Extended projection



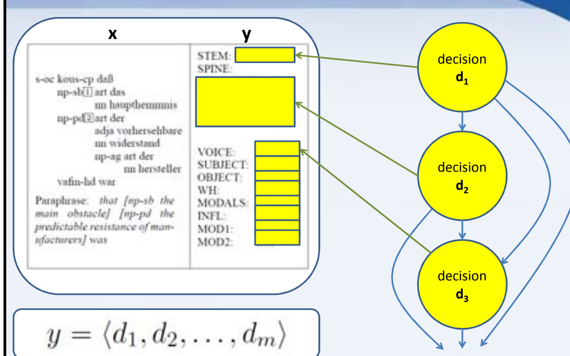
that the main obstacle **has been** the predictable resistance of manufacturers

Aligned extended projection

s-oc kous-cp daß np-sb 1 art das nn haupthemmnis np-pd 2 art der adja vorhersehbare nn widerstand np-ag art der nn hersteller vafin-hd war Paraphrase: *that [np-sb the main obstacle] [np-pd the predictable resistance of manufacturers] was*	STEM: be SPINE: SBAR-A IN that S NP-A VP V NP-A VOICE: active SUBJECT: 1 OBJECT: 2 WH: NULL MODALS: has INFL: been MOD1: null MOD2: null

VOICE: One of two alternatives, `active` or `passive`, specifying the voice of the main verb.

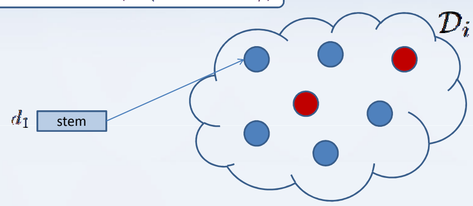
Model



$$y = \langle d_1, d_2, \ldots, d_m \rangle$$

Decisions

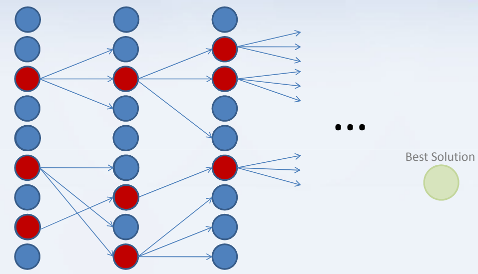
$d_i \in \mathrm{ADVANCE}(x, \langle d_1, \ldots, d_{i-1} \rangle)$



$\bar{\phi}(x, \langle d_1, \ldots, d_{i-1} \rangle, d_i) \in \mathbb{R}^d$

1	main verb
2	any verb in the clause
⋮	
20	are all of the verbs at the end?
21	nonterminal label of the root of the tree
22	terminal labels of all words constituting the subject

Prediction: Beam search



$$\Phi(x, y) = \sum_{i=1}^{m} \bar{\phi}(x, \langle d_1, \ldots, d_{i-1} \rangle, d_i)$$
$$\mathrm{SCORE}(x, y) = \Phi(x, y) \cdot \bar{\alpha}$$

Learning

training pairs: $\langle (e_1, g_1), (e_2, g_2), \ldots, (e_n, g_n) \rangle$

scoring:
$$\Phi(x, y) = \sum_{i=1}^{m} \bar{\phi}(x, \langle d_1, \ldots, d_{i-1} \rangle, d_i)$$
$$\mathrm{SCORE}(x, y) = \Phi(x, y) \cdot \bar{\alpha}$$

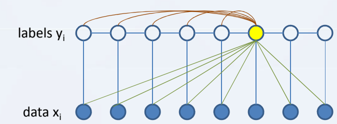
perceptron:

Inputs: Training examples (x_i, y_i)
Initialization: Set $\bar{\alpha} = 0$
Algorithm:
For $t = 1 \ldots T$, $i = 1 \ldots n$
Calculate $z_i = \arg\max_{z \in \mathbf{GEN}(x_i)} \Phi(x_i, z) \cdot \bar{\alpha}$
If$(z_i \neq y_i)$ then $\bar{\alpha} = \bar{\alpha} + \Phi(x_i, y_i) - \Phi(x_i, z_i)$
Output: Parameters $\bar{\alpha}$

Recap

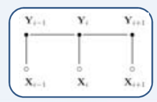
beam search

$$\Phi(x, y) = \sum_{i=1}^{m} \bar{\phi}(x, \langle d_1, \ldots, d_{i-1} \rangle, d_i)$$
$$\mathrm{SCORE}(x, y) = \Phi(x, y) \cdot \bar{\alpha}$$
$$F(x) = \arg \max_{y \in \mathrm{GEN}(x)} \mathrm{SCORE}(x, y)$$



CRF

$$p(\mathbf{Y}_v \mid \mathbf{X}, \mathbf{Y}_w, w \sim v)$$



Results & Conclusions

- per clause based machine translation
- structured prediction of AEPs
- decisions depend on
  - whole input *X* and
  - all previous decisions
- learning based on perceptron
- beam search to solve *argmax*
- comparable results with other approaches