Syntax-Directed Translation

CSE 399-005 Final Exam

Liang Huang
A Test on Systran

- **source:** (Matthew 26:41)
  - "the spirit is willing, but the flesh is weak."
- **translate to another language and then back**
  - **Japanese:** Mind is rejoicing but the meat is weak.
  - **Spanish:** The alcohol is arranged but the meat is weak.
  - **Russian:** The vodka is tempting, but the meat's a bit suspect.
Machine Translation

- **Background**
  - one of the earliest attempts of AI
  - yet one of the hardest problems to solve

- **Word-by-word translation**
  - source: *the gunman was shot to death by the police.*
  - google: 这个 枪手 被 射击 对 死亡 由 警察。
  - human: 枪手 被 警察 击 毙。

- **Problems with this approach**
  - reordering
  - context-sensitive lexical choice
Syntax-Directed Translation

• How do we human-beings do translations?
  1. **understand** the source sentence
  2. **generate** the target sentence

• Machine Translation?
  1. **parse** the source sentence into a tree
  2. **recursively transfer** the tree into the target language

  • *this final exam*
I. Parsing

parse the input sentence into a tree structure according to a grammar - Treebank grammar

\[ S \rightarrow NP-C \rightarrow DT \rightarrow the \rightarrow NP-C \rightarrow NN \rightarrow \text{gunman} \rightarrow VP \rightarrow VBD \rightarrow was \rightarrow VP \rightarrow VBN \rightarrow \text{shot} \rightarrow PP \rightarrow TO \rightarrow \text{to} \rightarrow \text{by} \rightarrow \text{by} \rightarrow DT \rightarrow the \rightarrow NN \rightarrow \text{police} \rightarrow PP \rightarrow IN \rightarrow \text{death} \]

- S - sentence
- NP - noun phrase
- VP - verb phrase
- PP - prepositional phrase
- DT - determiner
- NN - noun
- VBD - past tense (-ed)
- VBN - past participle
- TO - “to”
- IN - preposition
2. Recursive Transfer

- converting tree-fragments into the target language
2. Recursive Transfer

- converting tree-fragments into the target language

```
S
  NP-C  VP
    |    |
    DT  NN  VBD  VP-C
      |    |    |
      the gunman was VP PP
        |    |
        VBN  TO  IN  NP-C
         |    |    |    |
         shot TO NP-C by DT NN
             |    |    |
             to NN the police
                 |   death
                    |
```
2. Recursive Transfer

- converting tree-fragments into the target language
2. Recursive Transfer

- converting tree-fragments into the target language

```
 VP
  |  
 VBD | VP-C
   |   |
   was |   |
   VP | PP
     | IN |
     | NP-C |
     |   |
     | VBN |
     |     |
     | shot |
     |     |
     | TO   |
     |   |
     | NP-C |
     |   |
     | IN  |
     | by |
     | DT |
     | NN |
     | the |
     | police |
     |     |
     | death |
```
2. Recursive Transfer

- converting tree-fragments into the target language

枪手

qiangshou
2. Recursive Transfer

- converting tree-fragments into the target language

槍手

qiangshou
2. Recursive Transfer

- converting tree-fragments into the target language
2. Recursive Transfer

- converting tree-fragments into the target language

枪手 被 qiangshou bei

jingcha

警察

被

VP

VBN

the police

shot

to

death

to

NP-C

NP-C

the police

Shot

TO

death
2. Recursive Transfer

- converting tree-fragments into the target language

枪手 被 警察

$qiangshou$ $bei$ $jingcha$

```
(shot VP)
  TO
  NP-C
dead
```

$VBN$ $PP$
2. Recursive Transfer

- converting tree-fragments into the target language

枪手 被 警察

qiangshou  bei  jingcha
2. Recursive Transfer

- converting tree-fragments into the target language
Transfer Rules

converting English into Chinese, Spanish, Arabic, etc.
Transfer Rules

• from (LHS)
  • English subtree with variables

• to (RHS)
  • Foreign string with variables

• variables in RHS is a permutation of those in LHS
  • 1-1 mapping (no deletion, no insertion)
Example Rules

Phrasal Translation

$$VP \rightarrow \text{está cantando}$$

$$\underbrace{\text{VBZ} \quad \text{VBG}}_{\text{is} \quad \text{singing}}$$

Non-contiguous Phrases

$$VP \rightarrow \text{poner, } x_0$$

$$\underbrace{\text{VB} \quad x_0:NP \quad \text{PRT}}_{\text{put} \quad \text{on}}$$

Lexicalized Reordering

$$NP \rightarrow x_1 \text{ de } x_0$$

$$\underbrace{x_0:NP \quad PP}_{\text{of}}$$

Non-constituent Phrases

$$S \rightarrow \text{il y a, } x_0$$

$$\underbrace{\text{PRO} \quad \text{VP}}_{\text{there} \quad \text{are}}$$

$$\underbrace{x_0:NP}_{\text{are}}$$
Another Example

• my friend’s black cat
• le chat noir de mon ami
• el gato negro de mi amigo
• the cat black of my friend
• please refer to the exam paper for this example
Search Algorithms

find the best derivation
Search

- depth-first-search (DFS)
- for each tree node
- try all rules applicable
- recursion on variables
- plugin the results

VP
  VBD was
  VP-C

VP
  VBN shot
  PP
   TO
   NP-C

IN by

NP-C
  DT the
  NN police

VT
  jibi

jingcha

jingcha

bei

beijing

police
Search

- depth-first-search (DFS)
- for each tree node
- try all rules applicable
- recursion on variables
- plugin the results
- many rules applicable
- at each tree node
- exponential run-time!
Memoization

- dynamic programming
- remember each tree node visited before
- if visited again, simply return the best translation
- linear-time algorithm
- linear in input length

The diagram illustrates the structure of the sentences and the process of memoization in natural language processing.
**Algorithm 1** Top-down Memoized Recursion

1: function \textsc{Translate}(\eta) \\
2: \hspace{1em} \textbf{if} cache[\eta] defined \textbf{then} \quad \triangleright \text{this sub-tree visited before?} \\
3: \hspace{1em} \textbf{return} cache[\eta] \\
4: \hspace{1em} \textbf{best} \leftarrow 0 \\
5: \hspace{1em} \textbf{for} r \in \mathcal{R} \textbf{do} \quad \triangleright \text{try each rule } r \\
6: \hspace{2em} \text{matched, sublist} \leftarrow \textsc{PatternMatch}(t(r), \eta) \quad \triangleright \text{tree pattern matching} \\
7: \hspace{2em} \textbf{if} matched \textbf{then} \quad \triangleright \text{if matched, } \text{sublist} \text{ contains a list of matched subtrees} \\
8: \hspace{2em} \text{prob} \leftarrow \Pr(r) \quad \triangleright \text{the probability of rule } r \\
9: \hspace{2em} \textbf{for} \eta_i \in \text{sublist} \textbf{do} \quad \triangleright \text{recursively solve each sub-problem} \\
10: \hspace{3em} p_i, s_i \leftarrow \textsc{Translate}(\eta_i) \\
11: \hspace{3em} \text{prob} \leftarrow \text{prob} \cdot p_i \\
12: \hspace{2em} \textbf{if} \text{prob} > \text{best} \textbf{then} \quad \triangleright \text{plug in the results} \\
13: \hspace{2em} \text{best} \leftarrow \text{prob} \\
14: \hspace{2em} \text{str} \leftarrow \{x_i \mapsto s_i\} s(r) \quad \triangleright \text{caching the best solution for future use} \\
15: \hspace{1em} \text{cache}[\eta] \leftarrow \text{best, str} \quad \triangleright \text{returns the best string with its prob.} \\
16: \hspace{1em} \textbf{return} cache[\eta]
Improvements?

- extra credit problems
  - [easy] print out the derivation
  - [medium] rule indexing
  - [medium] DFS with Branch-and-Bound
  - [challenging] output not only the top translation, but also the 2nd-best, 3rd-best, ... k-th best translations
    - so that you can get “le chat noir de mon amie”
Sample Translations

src: this year, this leading role is still very prominent.

ref: 今年, 这种主导作用依然非常突出。

trs: 今年, 这领导作用还很突出。

src: executive committee of fifa also announced some reform measures.

ref: 国际足联执委会还宣布了一些改革措施。

trs: 国际足联执行委员会还宣布了一些改革措施。

src: at present, some western nations have already announced their termination of economic aid to zimbabwe.

ref: 目前,某些西方国家已经宣布终止对津巴布韦的经济援助。

trs: 目前,一些西方国家已公布终止津巴布韦经济援助。
Thank you!
谢谢！

Merci - Gracias - Grazie -
Obrigado - Danke -
ευχαριστώ - спасибо -
谢谢 - शुक्रिया -
...