

Natural Language Processing

Greg Kondrak

Outline

- AI and Language
- Natural Language Processing
- Syntactic Analysis
 - Grammars
 - Parsing
 - Chart Parsing Algorithm

The Turing test

Q: In the first line of your sonnet which reads "Shall I compare thee to a summer's day," would not "a spring day" do as well or better?
A: It wouldn't scan.
Q: How about "a winter's day"? That would scan all right.
A: Yes, but nobody wants to be compared to a winter's day.
Q: Would you say Mr. Pickwick reminded you of Christmas?
A: In a way.
Q: Yet Christmas is a winter's day, and I do not think Mr. Pickwick would mind the comparison.
A: I don't think you are serious. By a winter's day one means a typical winter's day, rather than a special one like Christmas.

Eliza – sample dialog

P: Men are all alike.
E: In what way?
P: They're always bugging us about something or other.
E: Can you think of a specific example?
P: Well, my boyfriend made me come here.
E: Your boyfriend made you come here.
P: He says I'm depressed much of the time.
E: I'm sorry to hear you are depressed.
P: It's true. I am unhappy.
E: Do you think coming here will help you not to be unhappy?
P: I need some help, that much seems certain.
E: What would it mean to you if you got some help?
P: Perhaps I could learn to get along with my mother.

Eliza – sample patterns

Word	Ran	Pattern	Outputs
alike	10	?X	In what way? What resemblance do you see?
are	3 3	?X are you ? Y ?X are ?Y	Would you prefer it if I weren't ?Y? What if they were not ?Y?
always	5	?X	Can you think of a specific example? When? Really, always?
what	2	?X	Why do you ask? Does that interest you?

What understanding isn't

- Merely reacting in a fixed or random manner to certain words in the text, and ignoring the rest of the text and the wider context.
- Programs like Eliza can give illusion of understanding if circumstances are right. Rarely can illusion be sustained.

Schank's scripts

"A man went into a restaurant and ordered a hamburger. When the hamburger arrived it was burned to a crisp, and the man stormed out of the restaurant angrily, without paying for the hamburger or leaving a tip".

Goals of Language Processing

- A scientific theory of communication by language:
 - to understand the structure of language and its use as a complex computational system,
 - to develop the algorithms and data structure that can implement/approximate the system.

Some applications

- Speech recognition
- Text-to-speech synthesizers
- Web-search engines
- Machine translation
- Information extraction
- Automatic summarization
- Question answering

Levels of Analysis

- Phonetics
- Phonology
- Morphology
- Lexicon
- Syntax
- Semantics
- Pragmatics and Discourse

Syntax

- Syntax is the study of the regularities and constraints of word order and phrase structure.
 - Which words are grouped together into a "phrase"
 - How words within a phrase are related to a common "theme"
 - How different phrases are related to each other

Phrase Structures

- Noun phrases
 - A noun phrase consists of a head noun and a set of modifiers.
 - The meaning of the noun phrase is largely determined by the noun.
- Verb phrases
 - A verb phrase consists of a head verb and a set of modifiers
 - the head verb denotes the action/activity/state

Grammar

- Must specify:
 - the primitive elements
 - how they can combine
- Terminology
 - constituents
 - start symbol, productions, terminals, non-terminals, derivation
 - grammatical vs. ungrammatical

Example Grammar

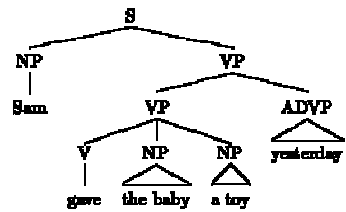
- $S \rightarrow NP VP$
- $NP \rightarrow Art N$
- $NP \rightarrow Art Adj N$
- $NP \rightarrow NP PP$
- $VP \rightarrow V$
- $VP \rightarrow V NP$
- $PP \rightarrow P NP$
- $Art \rightarrow the, a$
- $Adj \rightarrow old, happy$
- $N \rightarrow cat, car, sofa$
- $V \rightarrow laugh, scratch$
- $P \rightarrow in, to, under$

Parsing

- The task of assigning a correct tree to a string given some Context-Free Grammar.
 - "correct" means consistent with the input and the grammar
 - the leaves of the tree cover all and only the input
 - the tree corresponds to a valid derivation according to the grammar

A parse tree

(15)



Parsing as Search

- The search space is the space of trees generated by the grammar.
- The search is guided by the structure of the space and by the input.

Multiplication of parses

- "List the sales of the products produced in 1972 with the products produced in 1972".
- 455 parses
- Extending the coverage of the grammar to obscure constructions increases the number of undesired parses for common sequences.

Parsing: Top-Down vs. Bottom-up

- | | |
|--------------------------------------|---|
| ▪ "goal-driven" | ▪ "data-driven" |
| ▪ start from the root node S | ▪ start from the input words |
| ▪ expand constituents recursively | ▪ group constituents into larger ones |
| ▪ stop when a tree matches the input | ▪ stop when a tree rooted in S is built |
| ▪ always compatible with the grammar | ▪ always compatible with the input |

Main Problems

- ambiguity
 - global ambiguity
 - local ambiguity
- repeated parsing of sub-trees
- left-recursion

A smarter parsing algorithm

- solves the left-recursion problem
- does not do repeated work
- solves an exponential problem in polynomial time
- answer: dynamic programming

Combining the Top-Down and the Bottom-Up Approaches

- Use top-down search control strategy with bottom-up filtering
- Depth-First Search strategy
- Search control issues
 - which node to expand next
 - which of the grammar rules to try
- Top-Down, Depth-First, Left-to-Right

Ambiguity

- Categorical: "I saw that gasoline can explode."
- Lexical: "Joe played baseball/ the piano/ a sonata/ Mozart."
- Structural: "I saw the man with binoculars."
- Reference: "Ross put the wine on the table. Because it wasn't level, it slid off."

Overlapping fields

- Computer science
- Linguistics
- Cognitive psychology
- Philosophy
- Mathematics
- Engineering