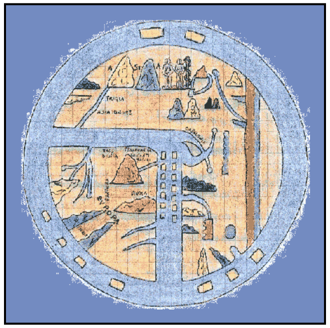
Lecture 15
Processing Human Language by Computer: Advances and Challenges (Part I)

Rob Gaizauskas



Outline

- The Significance of Language
- Turing and the Imitation Game
- Why Language Processing is so hard for a Computer
- Two Frameworks for the Computational Analysis of Language
  - Symbolic
  - Statistical
- Applying Partial Knowledge: Text Processing Applications
- Summing up and the Road Ahead

The Significance of Language

- Language is a distinguishing feature of human beings
  - Other animals can **communicate** (transmit signals that convey information)
    - bee dances, vervet danger calls
  - No other animal can
    - build arbitrarily complex sentence structures
    - refer to non-present times/places or to non-existent situations

  despite, e.g., extensive attempts to teach apes language

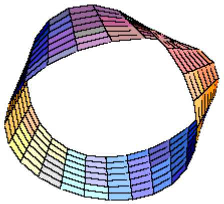
The Significance of Language

- Language is the key medium of social interaction
  - Hard to think of a social structure that is NOT predicated on language: business, education, courts, government, news, …
  - Most scholars agree writing systems arose as a means of recording business transactions
  - Dunbar (1996) argues gossip replaces grooming for humans
    - up to 10-20% of apes' time spent grooming – necessary not for hygiene but to maintain social cohesion, defuse conflict
    - argues speech took over this role in hairless apes with abnormal vocal fluency

The Significance of Language

- Language is the basis of collective memory/culture
  - Written language literally makes "history" possible
  - Permanent, objective record of thought/discourse renders possible sharing of
    - knowledge/ideas
    - experience/descriptions of events
    - feelings/emotions

    across space and time
  - Facilitates building on ideas/discoveries of previous generations – without which we would not have witnessed the spectacular expansion of the human species over the past 5,000 years



The Significance of Language

- Language is intimately bound up with thought
  - Philosophers have long argued whether non-linguistic thought is possible – i.e. whether all thought takes place in language
    - Dennett, who rejects this strong claim, still argues that language constrains the brain/thinking:

      "Language infects and inflects our thought at every level. The words in our vocabularies are catalysts that can precipitate fixations of content as one part of our brain tries to communicate with another. The structures of grammar enforce a discipline on our habits of thought, shaping the ways in which we probe our own "data bases" … The structures of the stories we learn provide guidance at a different level, prompting us to ask ourselves the questions that are most likely to be relevant to our current circumstances." (D. Dennett, Consciousness Explained, 1991)

  - Psychologists, such as Vygotsky, interested in the cognitive development of the child, argue that learning word meaning and concept acquisition in the child go hand in hand

Outline

- The Significance of Language
- Turing and the Imitation Game
- Why Language Processing is so hard for a Computer
- Two Frameworks for the Computational Analysis of Language
  – Symbolic
  – Statistical
- Applying Partial Knowledge: Text Processing Applications
- Summing Up and The Road Ahead

Turing and the Imitation Game

- Reflecting the centrality of language to our conception of human intelligence, Turing proposed conversation as the decisive test for machine intelligence ("Computing Machinery and Intelligence", 1950)

“Turing Test”

From: www.alanturing.net
©Copyright B.J. Copeland, July 2000

Turing and the Imitation Game

- Turing's example of an exchange that might occur during the test.

  Interrogator 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

  Computer It wouldn't scan.

  Interrogator How about 'a winter's day'? That would scan all right.

  Computer Yes, but nobody wants to be compared to a winter's day.

  Interrogator Would you say Mr. Pickwick reminded you of Christmas?

  Computer In a way.

  Interrogator Yet Christmas is a winter's day, and I do not think Mr Pickwick would mind the comparison

  Computer I don't think you're serious. By a winter's day one means a typical winter's day, rather than a special one like Christmas.

Turing and the Imitation Game

- In his 1950 paper Turing speculated that his test would be passed by 2000
- Now 2013 and no computer has passed the test
  - Loebner competition* runs annually
    - $100,000 prize for first computer to fool the judges
    - Annual prize ($4000 in 2013) for most convincing (="least bad") computer

- Why not?
- How have computer scientists, computational linguists and artificial intelligence researchers been pursuing the project getting computers to "understand" human language?

* http://www.loebner.net/Prizef/loebner-prize.html

Outline

- The Significance of Language
- Turing and the Imitation Game
- Why Language Processing is so hard for a Computer
- Two Frameworks for the Computational Analysis of Language
  - Symbolic
  - Statistical
- Applying Partial Knowledge: Text Processing Applications
- Summing Up and The Road Ahead

Why Language Processing is so hard for a Computer

- A computer's primitive capabilities can be captured in the following minimal set of operations:
  - add one to a number in a named location
  - subtract one from a number in a named location
  - repeat a number of operations until a number in a stored location tests equal to 0
- Given these capabilities it is *relatively* easy (60 years of computer science!) to create programs to "interpret" artificial languages whose commands bundle multiple such primitive commands into common operations.

  E.g.
  - 'print all lines in a text file that contain the string "giraffe"'
  - 'change all lower case "a"s to "B"s'
- Examples of such programming languages:
  - Java, Fortran, Perl, Python …

Why Language Processing is so hard for a Computer

- Such languages are **artificial**:
  - Constrained, unambiguous vocabulary
  - Fixed, simple, unambiguous grammar
  - Interpretation requires no knowledge beyond the text
  - …
- None of these characteristics holds of **natural** languages

Why Language Processing is so hard for a Computer: Vocabulary/lexicon

- New “words” constantly appearing
  - Names (people, places, companies, products, buildings, films…)
  - Terminology in technical domains
  - Neologisms (“ecotourist”, “cyberterrorist”, “blog” …)
- Words can have multiple senses
  - “bridge”, “bank” …
    - on average 1.34 senses per open class word (nouns, adjectives, adverbs,verbs) in the WordNet 3.0 lexical database of 155,000 words
  - Which sense catalogue is correct?
  - Which sense is correct in context: “I saw the crane near the bank”
    - ~20 sense combinations for sentences with 10 open class words; ~80 for 15
- Sense of existing words shifts with time or new senses are acquired
  - “literally” – “*The new iMac's screen **literally** floats in the air*”
  - “gay”, “programmatic”, “twitter”

Why Language Processing is so hard for a Computer: Syntax/Grammar

- Boundaries of what is/is not grammatical are fuzzy

  It were right exciting

- Grammar is complex – no complete, generally accepted grammatical description of any NL
- Grammar is ambiguous – more than one acceptable interpretation of the same string

?

He saw [the man] [with the binoculars]

?

  – Problem non-trivial: corpus analysis shows on average each sentence in newswire text contains two ambiguous prepositional phrases (Mitchell, 2003)

Why Language Processing is so hard for a Computer: Syntax/Grammar

- Boundaries of what is/is not grammatical are fuzzy

  It were right exciting
- Grammar is complex – no complete, generally accepted grammatical description of any NL
- Grammar is ambiguous – more than one acceptable interpretation of the same string

> One morning I shot an elephant in my pajamas.
> How he got into my pajamas I don't know.
> Groucho Marx, *Animal Crackers*, 1930.

  - We generally chose one meaning unreflectively, relying on knowledge of typical event patterns

Why Language Processing is so hard for a Computer: World knowledge/conceptual knowledge

- Consider this narrative fragment

  John plucked up the courage to ask the president for a raise. He …

- What is the antecedent of "he"? -- could be either "John" or "the president"

  John plucked up the courage to ask the president for a raise. He was sitting stonily behind his mahogany desk.

  John plucked up the courage to ask the president for a raise. He was shaking as he entered the office.

- Given the completion how do we know which antecedent to chose? How do we make a computer know this …

Why Language Processing is so hard for a Computer: World knowledge/conceptual knowledge

- Consider these partial sentences from a Medline abstract:

  The three-dimensional structure of Endo H has been determined …

  A shallow curved cleft runs across the surface of the molecule from …

  This cleft contains the putative catalytic residues Asp130 and Glu132 …
- From them a biologist easily understands that *Asp130* and *Glu132* are in the protein *Endo H*
- To do so requires knowing, amongst other things
  - Endo H is a protein
  - All proteins are molecules
  - A cleft in the surface of an object is in the object
  - When X is in Y and Y is in Z then X is in Z
- None of this is knowledge is <u>in</u> the text – it's <u>in</u> the reader

Why Language Processing is so hard for a Computer: Summary

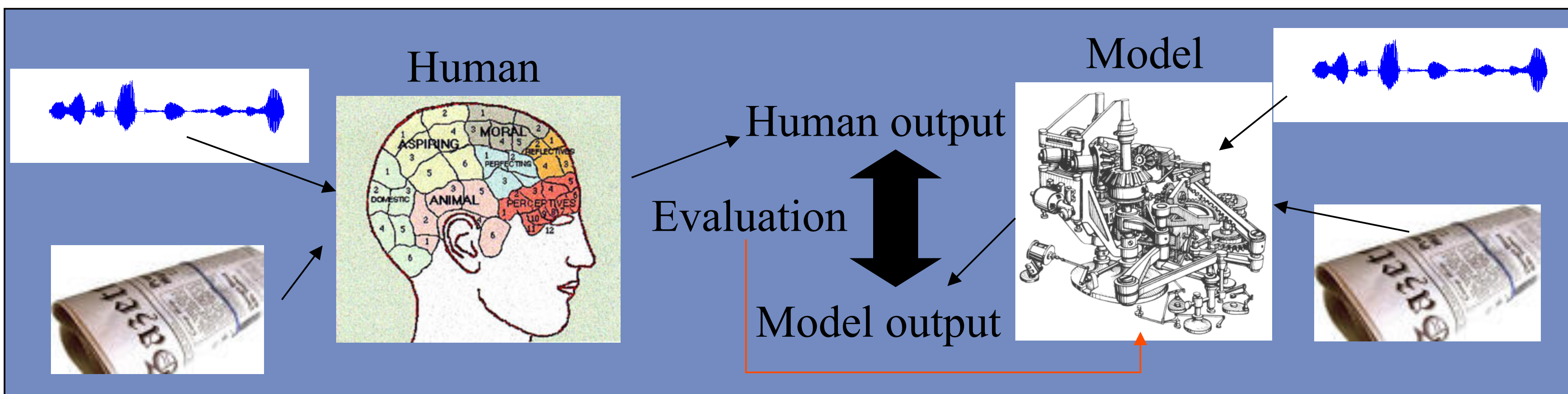
- Language is
  - dynamic – new words/word senses
  - highly ambiguous
    - multiple possible meanings for words
    - multiple possible syntactic structures for word sequences
  - complex – no as yet discovered comprehensive set of grammar rules
  - partial – requires the reader/listener to supplement the text/utterance with world/conceptual knowledge in order to recover the message

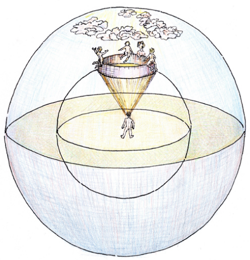
Outline

- The significance of language
- Turing and the Imitation Game
- Why Language Processing is so hard for a Computer
- Two frameworks for the computational analysis of language
  - Symbolic
  - Statistical
- Applying partial knowledge: text processing applications
- The Road Ahead

Computational Models of Language

- To progress with either deeper understanding or application, some conceptual framework for studying language is necessary
- Since observed language is the input/output of processes (speaking/hearing; writing/reading) it is natural to embody any such framework in a computational model
- Such a model forms a conjecture about how human-like linguistic outputs could be produced given certain inputs – and can be tested/evaluated against data



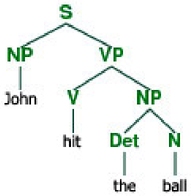


Computational Models of Language

- Models which have been operationalised in computer programs include
  - Descriptive linguistic models based on theories of syntax, semantics and discourse
  - Statistical models derived from analysis of large volumes of data
  - Psycholinguistic models based on observation of human subjects
  - Neurocognitive models based on functional neural imaging techniques
- Most computational linguistic work to date has concentrated on descriptive linguistic models and statistical models and hybrids of these

  Pluses:
  - Descriptive linguistics rich in data and analysis
  - Statistical approaches offer robustness, learning from data
  - Both permit implementations that make no claim to process language in the way that humans do

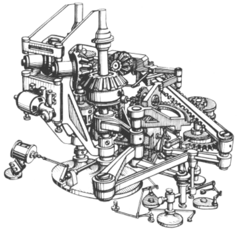
  Minuses:
  - Both permit implementations that make no claim to process language in the way that humans do



Descriptive Linguistic Models

- Linguists traditionally describe written language via a hierarchy of levels, with sub-word components at the bottom and texts or discourses situated in some real world setting at the top

Pragmatics	Study of how language is used in context to achieve communicative ends
Discourse Interpretation	Study of how multi-sentence texts are integrated into a coherent model of a depicted world
Semantics	Study of the literal meanings of words and of how these combine to yield literal meanings of sentences
Syntax	Study of how words/word components combine to form grammatical phrases/sentences
Morphology	Study of the minimal meaningful subunits of words (dis-en-tangle-ed)



Operationalising Descriptive Linguistic Models (1)

- Descriptive linguistics provides no process model, but various processing models can be/have been built in this paradigm

Pragmatic interpretation: implicature/ conversational agents

↑

Discourse Interpretation: reference resolution/ domain/scenario knowledge application

↑

Semantic analysis: word sense selection/mapping to logical form

↑

Syntactic analysis: POS tagging + parsing

↑

Morphological analysis

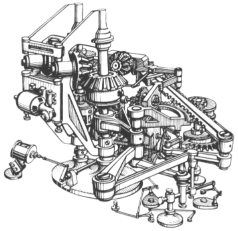
JOHN turned on the OVEN.
He grilled the FISH.

John turned on the oven.
He grilled the fish.
IN

person(e1),gender(e1,male),
grill$_1$(e2),tense(e2,past),
fish$_1$(e3)
lsubj(e2,e1),lobj(e2,e3)

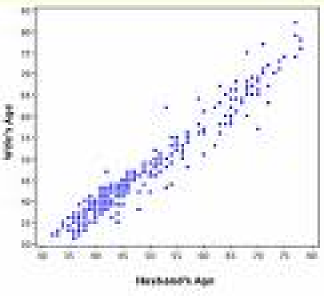
S
NP VP
N V NP
DT N
He grilled the fish

He grill+ed the fish

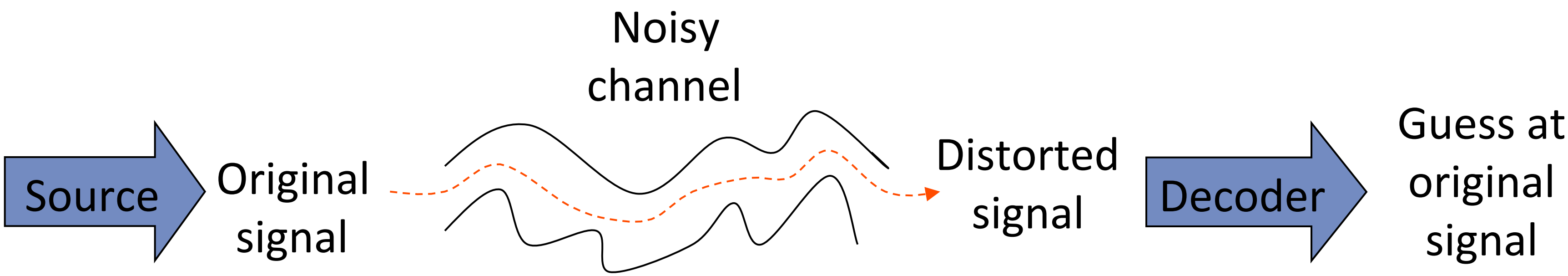


Operationalising Descriptive Linguistic Models (2)

- Processing components in descriptive linguistic models (e.g. POS taggers, parsers, word sense disambiguation modules) may be
  - rule-based or statistical
  - hand-authored or learned automatically from data
- Key assumptions:
  - Presuppose (at least some) theoretical, i.e. non-observed, constructs of linguistics -- e.g. parts of speech, phrasal/grammatical categories
  - Presume language expresses “meaning” which can be represented in some extra-linguistic formalism (e.g. logic) amenable to inference (for this reason this approach sometimes called symbolic)

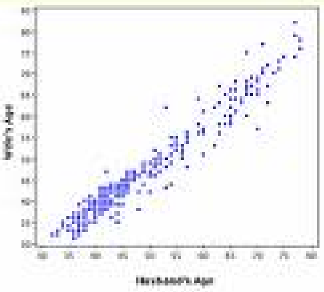


Statistical Models and Information Theory



- Shannon's "noisy channel" model provides a fruitful way to look at a number of language problems

Problem	Original Signal	Distorted Signal
Spelling Correction	Correctly spelled text	Misspelled text
Speech Recognition	Written Sentence	Spoken Sentence
Machine Translation	English text	French text
Part of speech tagging	Tagged sentence	Sentence minus tags
…	…	…



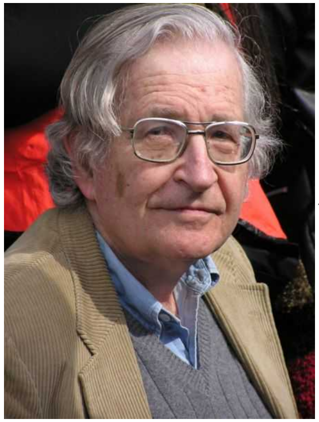
Statistical Models and Information Theory

- Decoder makes "guess" using Bayesian inference, employing
  - prior probability of signal
  - probability of observing distorted signal given the signal

  to compute most probable signal given the observed distorted signal
- Probabilities are estimated from training data, e.g. samples of
  - spelling mistakes + corrected spellings (spelling correction)
  - speech + transcription (speech recognition)
  - original + translation (machine translation)
  - text + tags (part-of-speech tagging)
- Language has characteristic that likelihood of seeing a specific element, e.g. word, at any point in sequence conditioned by preceding sequence
- Cannot get sensible probabilities of entire preceding sequence – so approximate using product of probabilities of short, preceding subsequences, such as word pairs or triples (Markov assumption)

Linguistic vs Statistical Models: Strengths and Weakneses

- History of NLP has seen fortunes of both wax and wane



But it must be recognized that the notion “probability of a sentence’” is an entirely useless one, under any known interpretation of the term.

Noam Chomsky, 1969

Fred Jelinek, 1988

Every time I fire a linguist our system performance improves.

Linguistic vs Statistical Models: Strengths and Weakneses

- Linguistic models:
  - Weaknesses:
    - Methods of rule acquisition: introspection or corpus analysis by linguist?
    - Fragility of rule-based systems
  - Strengths:
    - Can model higher level NL tasks for which data sparseness appears insuperable
    - Models can be inspected by humans; (appear to?) have explanatory power
- Statistical models:
  - Weaknesses:
    - Not suited for certain tasks: question answering? co-reference?
    - Markov assumption clearly overly simplistic
  - Strengths:
    - Robustness in face of noisy data
    - Wide coverage/automatic trainability
- Pragmatic language engineers take best of both (hybrid systems)
- Better models of language understanding may require moving beyond both …

Questions?