Computational Linguistics

CSC 2501 / 485
Fall 2019

1. Introduction to computational linguistics

Gerald Penn
Department of Computer Science, University of Toronto
(many slides taken or adapted from others)

Reading: Jurafsky & Martin: 1.
Bird et al: 1, [*2.3, 4*].

Copyright © 2019 Graeme Hirst, Suzanne Stevenson and Gerald Penn. All rights reserved.

Why would a computer need to use natural language?

Why would anyone want to talk to a computer?

- Computer as autonomous agent.
  Has to talk and understand like a human.

- Computer as servant.
  Has to take orders.

- Computer as personal assistant.
  Has to take orders.

Schedule a meeting tomorrow with George. Book me a flight to Vancouver for the conference. Find out why our sales have dropped in Lithuania. And write a thank-you note to my grandma for the birthday present.

- Computer as researcher.
  Needs to read and listen to everything.

- Computer as researcher.
  Brings us the information we need.

Find me a well-rated hotel in or near Stockholm where the food is good, but not one that has any complaints about noise.

Did people in 1878 really speak like the characters in *True Grit*?

Is it true that if you turn a guinea pig upside down, its eyes will fall out?

- Computer as researcher.
  Organizes the information we need.

- Computer as researcher.
  Wins television game shows.

IBM's Watson on *Jeopardy!*, 16 February 2011

https://www.youtube.com/watch?v=yJptrICVDHI

https://www.youtube.com/watch?v=Y2wQQ-xSE4s

- Computer as language expert. Translates our communications.

- **Input:**
  Spoken
  Written
- **Output:**
  An action
  A document or artifact
  Some chosen text or speech
  Some newly composed text or speech

Intelligent language processing

- Document applications
  Searching for documents by meaning
  Summarizing documents
  Answering questions
  Extracting information
  Content and authorship analysis
  Helping language learners
  Helping people with disabilities
  …

Example: Answering clinical questions at the point of care

Example: Early detection of Alzheimer's

- Look for deterioration in complexity of vocabulary and syntax.
- Study: Compare three British writers

Iris Murdoch

Died of Alzheimer's
Alzheimer's

P.D. James

No Alzheimer's

Agatha Christie

Suspected

Increase in short-distance word repetition

Rise, $p < .01$

Rise, $p < .01$

n.s.

Spoken documents

- “Google for speech”
  Search, indexing, and browsing through audio documents.
- Speech summarization
  Automatically select the 5–20% most important sentences of audio documents.

Speech recognition for

- Use a peech recog disab
- Creat speec

Bloorview KIDS REHAB

Speech-Language Pathology
UNIVERSITY OF TORONTO

Speech transformation for dysarthria

- Transform dysarthric speech to improve comprehensibility

thotra.

Speech-Language Pathology
UNIVERSITY OF TORONTO

Models of human language processing

- Highly multidisciplinary approach
- Exploit the relation between linguistic knowledge and statistical behaviour of words

Models of children's language acquisition

- Models of how children learn their language just from what they hear and observe
- Apply machine-learning techniques to show how children can learn:
  - to map words in a sentence to real world objects
  - the relation between verbs and their arguments

Mathematics of syntax and language

- Fowler's algorithm (2009): first quasi-polynomial time algorithm for parsing with Lambek categorial grammars
- McDonald's algorithm (2005): novel dependency-grammar parsing algorithm based upon minimum spanning trees
- Parsing in freer-word-order languages

Machine Learning
Linguistics
Information Science
CL/NLP
Psycho-linguistics
Knowledge representation and reasoning
Signal processing

Computational linguistics 1

- Anything that brings together computers and human languages …
  - … using knowledge about the structure and meaning of language (*i.e.*, not just string processing).
- *The dream:* "The linguistic computer".
  - Human-like competence in language.

Computational linguistics 2

- The development of computational models with natural language as input and/or output.
- ***Goal:*** A **set of tools** for processing language (semi-) automatically:
  - To access linguistic information easily and to transform it — *e.g.,* summarize, translate, ….
  - To facilitate communication with a machine.
- "NLP": Natural language processing.

Computational linguistics 3

- Use of computational models in the study of natural language.
- ***Goal:*** A **scientific theory** of communication by language:
  - To understand the structure of language and its use as a complex computational system.
  - To develop the data structures and algorithms that can implement/approximate that system.

What does it mean to “understand” language?

The Turing Test

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?

It wouldn't scan.

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

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

Alan Turing, "Computing machinery and intelligence", *Mind*, 59, 1950, 433–460.

The Turing Test

Would you say Mr Pickwick reminded you of Christmas?

In a way.

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

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.

Alan Turing, "Computing machinery and intelligence", *Mind*, 59, 1950, 433–460.

Using language
is not necessarily
understanding language

A computer psychiatrist

Men are all alike.

In what way?

They're always bugging us about something or the other.

Can you think of a specific example?

Well, my boyfriend made me come here.

Your boyfriend made you come here.

He says I'm depressed.

Joseph Weizenbaum, *Computer Power and Human Reason*, W.H. Freeman, 1976.

A computer psychiatrist

I'm sorry to hear you are depressed.

It's true. I am unhappy.

Do you think coming here will help you not to be unhappy?

Perhaps I could learn to get along with my mother.

Tell me more about your family.

Joseph Weizenbaum, *Computer Power and Human Reason*, W.H. Freeman, 1976.

Levels of understanding 1

- ***0. Keyword processing:***
  Limited knowledge of **particular words** or **phrases**, or their collocations.
  - Chatbots.
  - Information retrieval.
  - Web searching.

Levels of understanding 2

- ***1. Limited linguistic ability:*** Appropriate response to simple, highly constrained **sentences**.
  - Database queries in NL.
    "Show all sales staff who exceeded their quota in May."
  - Simple NL interfaces.
    "I want to fly from Toronto to Vancouver next Sunday."

Levels of understanding 2

Levels of understanding 3

- ***2. Full text comprehension:*** Understanding **multi-sentence** text and its relation to the “real world”.
  - Conversational dialogue.
  - Automatic knowledge acquisition.

- ***3. Emotional understanding/generation:***
- Responding to **literature, poetry, humour**
- Story narration.

Current research trends

- Emphasis on large-scale NLP applications.
  - *Combines:* language processing *and* machine learning.
- Availability of large text corpora, development of statistical methods.
  - *Combines:* grammatical theories *and* actual language use.
- Embedding structure into known problem spaces (especially neural networks!).
  - *Combines:* statistical pattern recognition *and* some sophisticated linguistic knowledge.

Building blocks of CL systems 1

- Language interpretation, language generation, and machine translation.
  - Part-of-speech (PoS) tagging.
  - Parsing and grammars.
  - Reference resolution.
  - Dialogue management.

Natural language interpretation

Does Flight 207 serve lunch?

↓

YNQ ($\exists$e SERVING(e) $\wedge$ SERVER(e, flight-207)
$\wedge$ SERVED(e, lunch))

Natural language generation

```
(spray-1  (OBJECT  paint-1)
          (PATH      (path-1
                     (DESTINATION wall-1))))
          (CAUSER sally-1)
```

Sally sprayed paint onto the wall.

Machine translation

- Current systems based purely on statistical associations.
- Getting incrementally better as they learn from more data.
- Probably more emergent knowledge of linguistics in there than we give them credit for, but it’s nearly impossible for us to extract.

http://www.duchcov.cz/gymnazium/

Gymnázium Duchcov

Home

History

Studied

Current school year

Admission

Maturity

People

Subjects

Google Apps

upc

NADACE ČEZ

INFOPHARM

History Gymnasium Duchcov

General desire Czech population after the establishment of secondary schools and English language teaching has been filled up after the First World War. On 6 October 1919 became the Czech grammar school provisionally operate in the German real grammar school. Huge interest in the study (Czech grammar school in Duchcov was only for Teplice, duchcovský and bílinský District) led the Directorate of the Institute to open a few classes and outside the German grammar school. Such a universal desire to acquire its own building stood for the birth of intent to build a school for duchcovské Czech efficient and dignified building. From the desire of the year 1919 to be implemented yet expired long eight years. New was transferred to use on Sunday 22nd May 1927. Over the decades, the school became not only an important educational institution, but also the centre of minority life, a number of outstanding graduates shows the level of instruction. It was built facilities, established tradition, the school into a broad awareness.

Like our country prožívala its famous and less famous days. The worst, perhaps during the war, when it was completely interrupted instruction. From the perspective of building a building suffered in the years 1948 - 1953, when it was cancelled real school building to pieces and various types of schools, which are for their operational needs to carry out structural adjustments.

http://www.duchcov.cz/gymnazium/ Translated by Google Translate, 14 July 2008

History Grammar Duchcov

General desire of the Czech population after the establishment of secondary schools in the Czech language of instruction was filled to the First World War. 6 October 1919 Czech Grammar School began to operate provisionally in the German grammar school. The enormous interest in the study (in Czech high school was only for Duchcov Teplice, Duchcovská and Bilina district) headquarters forced the institute to open a few classes outside of the German school. This and a general desire to get their own building was the genesis of the project to build a gymnasium for Duchcovská Czech efficient and dignified building. From the desire of the year 1919 to implement more long passed for eight years. New building was handed over for use on Sunday 22 May 1927. Over the decades, the school became not only an important educational institution, but also the center of the minority lives, many outstanding graduates shows the level of teaching. It was built facilities, created a tradition, the school turned into a broad awareness.

Like our country has endured its famous and less famous days. Perhaps the worst of the war, when teaching was completely interrupted. In terms of construction the building was seriously damaged in the years 1948 - 1953, when school was canceled and the real object is rotated different kinds of schools here for their operational needs to carry out construction work.

http://gymdux.sokolici.eu/index.php/informace/historie-koly
Translated by Google Translate, 3 August 2010.

History Gymnasium Duchcov

General desire Czech population after the establishment of secondary school teaching in Czech language was filled to the First World War. On 6 October 1919 began Czech grammar provisionally act in the German grammar school. The enormous interest in the study (Czech secondary school in Duchcově was only for Teplitz, and Duchcov Bílina district) headquarters induced Institute opened a few classes outside of German grammar. This and the general desire to obtain its object was the genesis of the project to build a high school for Duchcov Czech efficient and dignified building. From the desire of 1919 for implementation has passed even eight long years. The new building was handed over for use on Sunday 22 May 1927. Over the decades, the school has become not only a major educational institution, but also the center of the Czech minority, many outstanding graduates testifies to the level of instruction. It was built facilities created tradition, the school came into widespread awareness.

Like our entire country had its famous and less famous days. The worst perhaps during the war when teaching was completely interrupted. In terms of construction building was seriously damaged in the years 1948 - 1953, when it was abolished grammar school building and alternated different types of schools, which are for its operational needs in construction work.

http://gymdux.sokolici.eu/index.php/informace/historie-koly
Translated by Google Translate, 17 June 2013.

History Grammar School Duchcov

The general desire Czech population after the establishment of secondary school teaching in Czech language was filled to the First World War. On October 6, 1919 began Czech gymnasium provisionally act in the German grammar school. The huge interest in the study (Czech secondary school in Duchcově was only for Teplice, Bílina Duchcov and district) caused the Directorate of the Institute opened a few classes outside the German Gymnasium. This and a general desire to get your own object was the genesis of the project to build a high school for Duchcov Czech efficient and dignified building. From the desire of 1919 for implementation has passed even eight long years. The new building was handed over for use on Sunday 22 May 1927. During the decade, the school has not only become an important educational institution, but also the center of the minority, many outstanding graduates indicative of the level of education. It was built facilities, established tradition, the school came into widespread awareness.

Like our whole country had its famous and less famous days. The worst perhaps during the war when teaching was completely interrupted. In terms of construction and the building was seriously damaged in the years 1948 - 1953, when the grammar school was canceled in the building and alternated different types of schools, which are for its operational needs in construction work.

http://www.gspsd.cz/historie/historie-skoly
Translated by Google Translate, 26 May 2014.

History of the school

Added: 14.11.2014 | Author: Holikova | Viewed: 45142x |

The school was formed by the merger of two institutions - a grammar school and a secondary technical school. The tradition of both schools dates back to the 1920s. At present, both schools are located in the building of the grammar school, which was built in 1927. After extensive reconstruction, which took place in 1990 1996, it dominates the town of Duchcov. As one of the most beautiful buildings in the north of Bohemia is listed in the regional list of architectural monuments

HISTORY OF INDEPENDENT INDUSTRIAL SCHOOL

1919 - 1938

Soon after the proclamation of the Czechoslovak Republic on October 28, 1918, it seemed paramount to establish a Czech vocational school. Until then, only a German vocational school (Bergschule) existed in Duchcov. On **1 December 1919, the** Ministry of Education established the **State Vocational Secondary School** based in Duchcov. 46 Czech miners signed up for the first year.

The first year of teaching took place under difficult conditions. The school had only one room at the Duchcov German grammar school. The teaching staff consisted practically of two professors, the director of Dr. Ing. Alois Parma and his assistant Ing. Václav Přibáně. Later the school moved to the more spacious rooms of the Waldstein chateau in Duchcov.

1938 - 1941

The time of occupation was very difficult for the school as well as for the whole republic. In 1938 the borderland was seized by the Germans. On September 23, the class was stopped and on October 10 the Ministry of Education and National Enlightenment ordered the school to be transferred to Kladno, saying that pupils enrolled in Duchcov would leave the school and new ones would not be admitted. At the end of the school year on 30 June 1941 the school was liquidated.

1945 - 1959

After the end of World War II in May 1945 the building of the former German upper school (in Bezručova Street) was secured by representatives of the MNV in Duchcov. The former headmaster of the school, Ing. Hugo Malý to take over the school building and inventory. 37 pupils were admitted to the first year. The period of study was set at 2 years. It was taught only in the morning. In the afternoon the pupils went down the mines.

The school year 1952 - 53 began in the new building of the former national school Hynek Holub in Kubicových Street. In the school year 1954 - 55, the study of the previous field of Coal Mining was extended to include the field of Mining Measurement. In the school year 1955 - 56, a friendship was established with the mining school of E. Thälmann in Senftenberg in the GDR, which lasted until the 1980s.

1959 - 1969

The beginning of the 1960s was a milestone in the development of the school. The boom in mining in the 1960s led to an increase in the requirements for the number of graduates of the **Secondary Technical Mining School** in Duchcov, as the school was called since 1961. In order to speed up the training of graduates, a three-year field of study **Coal Mining** was introduced for trained miners. The study field **Mining Mechanization and Electrification** and the specialization **Coal Treatment** was also established.

Between 1960 and 1970, the study forms were extended by distance study, **Geology and Drilling** . At the end of the 1960s, the fields of **Power Electrical** and **Geodesy Equipment** were introduced **.**

1969 - 1979

On September 1, 1979, the school changed its name to the **Secondary Technical School.** In the second half of the 1980s, the mining branch changed its name to **Mining and Mining Geology.**

https://www.gspsd.cz/index.php?type=Post&id=256&ids=249
Translated by Google Translate, 5th September 2019.

Building blocks of CL systems 2

- Information extraction
  - Chunking (instead of parsing).
  - Template filling.
  - Named-entity recognition.

Information extraction

"Bridgestone Sports Co. said Friday it has set up a joint venture in Taiwan with a local concern and a Japanese trading house to produce golf clubs to be shipped to Japan. The joint venture, Bridgestone Sports Taiwan Co., capitalized at 20 million new Taiwan dollars, will start production in January 1990."

Tie-up-1: *Relation:* Tie-up
Entities: Bridgestone Sports Co.
a local concern
a Japanese trading house
Joint venture: Bridgestone Sports Taiwan Co.
Activity: Activity-1
Amount: NT $ 20,000,000

Activity-1: *Company:* Bridgestone Sports Taiwan Co.
Product: golf clubs
Start date: January 1990

Building blocks of CL systems 3

- Lexical semantics
  - Word sense disambiguation (WSD).
  - Taxonomies of word senses.
  - Analysis of verbs and other predicates
  - Embeddings of words into continuous vector space (word2vec, BERT, XLNet, etc.) .

- Computational morphology

Why is understanding hard? 1

- The structures that we are interested in are richer than strings – often *hierarchical* or *scope-bearing*.

Nadia knows Ross left.

KNOWS(Nadia, LEFT(Ross))

Why is understanding hard? 2

- Mapping from ***surface-form*** to meaning is many-to-one: Expressiveness.

Nadia kisses Ross. Ross is kissed by Nadia.

KISS (Nadia, Ross)

Nadia gave Ross a kiss. Nadia gave a kiss to Ross.

Why is understanding hard? 3

- Mapping is one-to-many:
  Ambiguity at all levels.
  - Lexical
  - Syntactic
  - Semantic
  - Pragmatic

Lexical ambiguity

The lawyer walked to the bar and addressed the jury.

The lawyer walked to the bar and ordered a beer.

You held your breath and the door for me. (Alanis Morissette)

- Computational issues
  - Representing the possible **meanings** of words, and their **frequencies** and their **indications**.
  - Representing **semantic relations** between words.
  - Maintaining adequate **context**.

used to strain microscopic plant life from the
zonal distribution of plant life .
close-up studies of plant life and natural
too rapid growth of aquatic plant life in water
the proliferation of plant and animal life
establishment phase of the plant virus life cycle
that divide life into plant and animal kingdom
many dangers to plant and animal life
mammals . Animal and plant life are delicately

automated manufacturing plant in Fremont
vast manufacturing plant and distribution
chemical manufacturing plant , producing viscose
keep a manufacturing plant profitable without
computer manufacturing plant and adjacent
discovered at a St. Louis plant manufacturing
copper manufacturing plant found that they
copper wire manufacturing plant , for example
‘s cement manufacturing plant in Alpena

vinyl chloride monomer plant , which is
molecules found in plant and animal tissue
Nissan car and truck plant in Japan is
and Golgi apparatus of plant and animal cells
union responses to plant closures .
cell types found in the plant kingdom are
company said the plant is still operating
Although thousands of plant and animal species
animal rather than plant tissues can be

Decision list for *plant*

LogL	Collocation	Sense
8.10	*plant life*	→ A
7.58	*manufacturing plant*	→ B
7.39	*life* (within ±2-10 words)	→ A
7.20	*manufacturing* (in ±2-10 words)	→ B
6.27	*animal* (within ±2-10 words)	→ A
4.70	*equipment* (within ±2-10 words)	→ B
4.39	*employee* (within ±2-10 words)	→ B
4.30	*assembly plant*	→ B
4.10	*plant closure*	→ B
3.52	*plant species*	→ A
3.48	*automate* (within ±2-10 words)	→ B
3.45	*microscopic plant*	→ A
	...	

Syntactic ambiguity 1

Nadia saw the cop with the binoculars.

Syntactic ambiguity 2

Put the book in the box on the table.

Put the book in the red book box.

Visiting relatives can be trying.

Noun phrase

Adj *Noun*

Verb *Noun*

Verb phrase

Syntactic ambiguity 3

- These are absolutely everywhere. Some real headlines:
  Juvenile Court to Try Shooting Defendant
  Teacher Strikes Idle Kids
  Stolen Painting Found by Tree
  Clinton Wins on Budget, but More Lies Ahead
  Hospitals are Sued by 7 Foot Doctors
  Ban on Nude Dancing on Governor's Desk
- Usually we don't even notice – we're that good at this kind of resolution.

Syntactic ambiguity 4

- Most syntactic ambiguity is **local** — resolved by syntactic or semantic context.
  Visiting relatives is trying.
  Visiting relatives are trying.
  Nadia saw the cop with the gun.
- Sometimes, resolution comes too fast!
  [ ][ ][[????
  The cotton clothing is made from comes from Mississippi.
  [[ ] [ ][[ ]]
  “Garden-path” sentences.

Syntactic ambiguity 5

- Computational issues
  - Representing the possible **combinatorial structure** of words.
  - Capturing syntactic **preferences** and frequencies.
  - Devising **incremental parsing** algorithms.

Semantic ambiguity

- Sentence can have more than one meaning, even when the words and structure are agreed on.

Nadia wants a dog like Ross's.

Everyone here speaks two languages.

Iraqi Head Seeks Arms.

DCS Undergrads Make Nutritious Snacks.

Pragmatic ambiguity

- A sample dialogue
  Nadia: *Do you know who's going to the party?*
  Emily: *Who?*
  Nadia: *I don't know.*
  Emily: *Oh … I think Carol and Amy will be there.*
- Computational issues
  - Representing **intentions** and **beliefs**.
  - **Planning** and **plan recognition**.
  - **Inferencing** and diagnosis.

Need for domain knowledge 1

Derivatization of the carboxyl function of retinoic acid by fluorescent or electroactive reagents prior to liquid chromatography was studied. Ferrocenylethylamine was synthesized and could be coupled to retinoic acid. The coupling reaction involved activation by diphenylphosphinyl chloride. The reaction was carried out at ambient temperature in 50 min with a yield of ca. 95%. The derivative can be detected by coulometric reduction (+100 mV) after on-line coulometric oxidation (+400 mV). The limit of detection was 1 pmol of derivative on-column, injected in a volume of 10µl, but the limit of quantification was 10 pmol of retinoic acid.

S. El Mansouri, M. Tod, M. Leclercq, M. Porthault, J. Chalom, "Precolumn derivatization of retinoic acid for liquid chromatography with fluorescence and coulometric detection." *Analytica Chimica Acta*, 293(3), 29 July 1994, 245–250.

Need for domain knowledge 2

In doing sociology, lay and professional, every reference to the "real world", even where the reference is to physical or biological events, is a reference to the organized activities of everyday life. Thereby, in contrast to certain versions of Durkheim that teach that the objective reality of social facts is sociology's fundamental principle, the lesson is taken instead, and used as a study policy, that the objective reality of social facts *as* an ongoing accomplishment of the concerted activities of daily life, with the ordinary, artful ways of that accomplishment being by members known, used, and taken for granted is, for members doing sociology, a fundamental phenomenon.

Harold Garfinkel, Preface, *Studies in Ethnomethodology*, Prentice-Hall, 1967, page vii.

Levels of linguistic structure and analysis 1

- Phonology
  - The **sound system** of a language.
- Morphology
  - The **minimal meaningful units** of language (root of a word; suffixes and prefixes), and how they combine.
- Lexicon
  - The semantic and syntactic **properties of words**.

Levels of linguistic structure and analysis 2

- Syntax
  - The means of expressing meaning: **how words can combine**, and in what order.
- Semantics
  - The **meaning** of a sentence (a logical statement?).
- Pragmatics
  - The **use** of a sentence: pronominal referents; intentions; multi-sentence structure.

Focus of this course 1

- Grammars and parsing.
- Resolving "syntactic" ambiguities.
- Determining "argument structure."
- Lexical semantics, resolving word-sense ambiguities.
- “Compositional” semantics.
- Understanding pronouns.

Focus of this course 2

- Current methods
  - Integrating statistical knowledge into grammars and parsing algorithms.
  - Using text corpora as sources of linguistic knowledge.

Not included

- Machine translation, language models, text classification, part-of-speech tagging…*
- Graph-theoretic and spectral methods%
- Speech recognition and synthesis*¶
- Cognitively based methods§
- Semantic inference,% semantic change/drift^
- Understanding dialogues and conversations¶
- Bias, fake news detection, ethics in NLP$

* csc 401 / 2511. % csc 2517. ¶ csc 2518. § csc 2540. ^ csc 2519. $csc 2528.

What about “Deep Learning?”

- Yes, we’ll definitely cover neural methods.
- Deep learning is more of a euphemism in NLP.
- The depth of the networks hasn’t really paid off to the same extent that it has in other areas.
- It would be more accurate to call what we do “fat learning.”
- But deep/fat learning isn’t the answer to all of our problems…

The Case of Text-to-Speech Synthesis

Google Research Blog

The latest news from Research at Google

Tacotron 2: Generating Human-like Speech from Text

Tuesday, December 19, 2017

Posted by Jonathan Shen and Ruoming Pang, Software Engineers, on behalf of the Google Brain and Machine Perception Teams

- "We find that the proposed system significantly outperforms all other TTS systems, and results in an MOS comparable to that of the ground truth audio."

The Case of Text-to-Speech Synthesis

Google Research Blog

The latest news from Research at Google

“mean opinion score”

human-generated speech

Tacotron 2: Generating Human-like Speech from Text

Tuesday, December 19, 2017

Posted by Jonathan Shen and Ruoming Pang, Software Engineers, on behalf of the Google Brain and Machine Perception Teams

- “We find that the proposed system significantly outperforms all other TTS systems, and results in an MOS comparable to that of the ground truth audio.”

Mean Opinion Score

- Ask this guy and his friends to label the “naturalness” of speech samples, some of them human-generated, some synthesized.

- …but assume all humans are professional readers
- …and don’t look too closely at what “naturalness” means.

The Case of Text-to-Speech Synthesis

Google Research Blog

The latest news from Research at Google

“almost failed to prove incomparability”

Tacotron 2: Generating Human-like Speech from Text

Tuesday, December 19, 2017

Posted by Jonathan Shen and Ruoming Pang, Software Engineers, on behalf of the Google Brain and Machine Perception Teams

- “We find that the proposed system significantly outperforms all other TTS systems, and results in an MOS comparable to that of the ground truth audio.”

Disclaimer 1

NATURAL TTS SYNTHESIS BY CONDITIONING WAVENET ON MEL SPECTROGRAM PREDICTIONS

Jonathan Shen[1], *Ruoming Pang*[1], *Ron J. Weiss*[1], *Mike Schuster*[1], *Navdeep Jaitly*[1], *Zongheng Yang*[*2], *Zhifeng Chen*[1], *Yu Zhang*[1], *Yuxuan Wang*[1], *RJ Skerry-Ryan*[1], *Rif A. Saurous*[1], *Yannis Agiomyrgiannakis*[1], *and Yonghui Wu*[1]

[1]Google, Inc., [2]University of California, Berkeley,
{jonathanasdf,rpang,yonghui}@google.com

L] 16 Feb 2018

ABSTRACT

This paper describes Tacotron 2, a neural network architecture for speech synthesis directly from text. The system is composed of a recurrent sequence-to-sequence feature prediction network that maps character embeddings to mel-scale spectrograms, followed by a modified WaveNet model acting as a vocoder to synthesize time-domain waveforms from those spectrograms. Our model achieves a mean opinion score (MOS) of 4.53 comparable to a MOS of 4.58 for profes-

the authors note, this was simply a placeholder for future neural vocoder approaches, as Griffin-Lim produces characteristic artifacts and lower audio quality than approaches like WaveNet.

In this paper, we describe a unified, entirely neural approach to speech synthesis that combines the best of the previous approaches: a sequence-to-sequence Tacotron-style model [12] that generates mel spectrograms, followed by a modified WaveNet vocoder [10, 15]. Trained directly on normalized character sequences and correspond-ing speech waveforms, our model learns to synthesize natural sound-

- "The overall mean score of -0.270 ± 0.155 shows that raters have a small but statistically significant preference towards ground truth over our results." (ICASSP 2018)

Disclaimer 2

- We asked Google to provide us with Tacotron prompts for the text that we used in our study, and they refused.
- So we used Polly, Amazon's deep-learning-based TTS system.
- We asked Google which statistical significance test they used in the ICASSP 2018 paper, and they did not respond to us.
- So what we are going to describe now is what most recent publications have used for significance testing.

“Significantly More Natural”

- What most people do is run Friedman’s test, together with a post-hoc pairwise comparison called the Wilcoxon signed rank test.
- This was probably selected because it is a “paired” test. Pairs = naturalness judgements by a single human subject on 2 systems/speakers.
- But paired implies a within-subjects protocol. MOS computes means across human judges.
- And the null hypothesis is very strong: the naturalness distribution of the two systems is *the same*.
- Better null hypothesis: $P(A < B) = P(B < A)$.
- Weaker null hypothesis → harder to refute.

Other Corrections

- Broaden the “human” population:
  - Professional speakers
  - Native non-professional speakers of North American English
  - Native non-professional speakers of Indian English
  - 4 dialects of Polly (AU, GB, IN, US)
- Include competitive non-deep-learning baselines:
  - 4 (non-DL) TTS systems from 2013 Blizzard Challenge
- Compare paired AB tests in addition to MOS calculations
  - Lower variance than absolute 1-5 judgements
  - Known to correlate with 6/8 ITU P.85 scales for “quality”
  - Nominal data: need Stuart-Maxwell test with post-hoc binomial sign test.

Post-hoc Binomial Sign Test

	Expert 1	Expert 2	Expert 3	Expert 4	Indian 1	Indian 2	Indian 3	Indian 4	American 1	American 2	American 3	American 4	Polly GB	Polly US	Polly AU	Polly IN	TTS C	TTS I	TTS K	TTS M
Expert 1																				
Expert 2	M																			
Expert 3	M	M																		
Expert 4	S	S	M																	
Indian 1	S	S	S	S																
Indian 2	N	S	N	S	S															
Indian 3	S	M	S	N	N	N														
Indian 4	S	S	S	S	N	S	N													
American 1	N	N	N	N	S	S	S	S												
American 2	N	N	S	M	S	S	S	S	N											
American 3	S	M	N	N	S	S	S	S	N	N										
American 4	S	S	M	S	S	S	S	M	M	S	M									
Polly GB	S	S	S	S	N	M	S	N	N	S	N	S								
Polly US	S	S	N	S	N	N	S	S	S	S	N	N	N							
Polly AU	N	M	N	S	S	N	N	N	S	S	S	N	N	N						
Polly IN	S	S	N	N	N	N	N	N	N	S	S	S	S	S	S					
TTS C	S	N	S	S	S	S	N	S	S	S	S	S	S	N	S	S				
TTS I	S	S	S	S	N	S	N	N	S	S	S	S	S	N	N	S	M			
TTS K	S	S	S	S	N	M	N	S	S	S	S	N	N	N	N	N	N	M		
TTS M	N	S	S	S	N	N	N	N	S	S	S	S	S	S	N	S	N	M	N	

- 2013 TTS was already indistinguishable from Indian speakers.
- Polly still not distinguishable from 2013 Blizzard systems.
- First studied by USAF for cockpit speech interfaces (Sampson & Navarro, 1984).

Thurstone's Law of Comparative Judgement

- The dialectal provenance of the sample is an important factor in determining judgements of naturalness.
- AB tests can be mapped onto an absolute scale.[1]
- As observed earlier, foreign accent helps TTS naturalness, but harms human naturalness.

-0.7054	Expert 2
-0.6786	American 2
-0.6031	American 4
-0.561	American 1
-0.4395	American 3
-0.4075	Expert 4
-0.3675	Expert 1
-0.2299	Expert 3
-0.0785	Polly IN
0.1511	Polly GB
0.1695	Polly AU
0.2282	TTS K
0.3044	Indian 1
0.305	TTS M
0.3328	Polly US
0.395	Indian 3
0.4167	Indian 2
0.4353	Indian 4
0.6229	TTS C
0.7098	TTS I

[1]Here we used the Bradley-Terry-Luce (BTL) model with Morrisey-Gulliksen scaling.

The Morals of the Speech Synthesis Story

- Did deep learning usher us into a new age of human-like speech production? Well, it didn't drag us out of it.
- Deep learning is more competitive against professional readers and native speakers of North American English.
- The purpose of a scientific evaluation is to create reasonable circumstances under which the experimental hypothesis might *fail*.
- In technically competent circles, the success of deep learning is more often ascribed to the achievements of "liberated practitioners," e.g., speech and language researchers, who, thanks to a string of achievements by machine learning researchers in modularizing their own discipline beginning in the late 1990s, were able to focus on higher-level design issues and *stop* researching ML.