Research Methods in Computational Linguistics - An Overview

Sowmya Vajjala

Iowa State University, USA

07 March 2018
Guest lecture in ENGL 501:
Introduction to Research in Rhetoric, Composition, and Professional Communication

Outline

- Introduction
  1. My three questions
  2. What is Computational Linguistics (CL)?
- Typical research problems in CL
- Research Methods: Theoretical and Empirical CL
- Methods in practice: Example of empirical CL research

Introduction

Three questions + the recommended reading

1. What comes to your mind when you think about doing content analysis with a computer?

Three questions + the recommended reading

1. What comes to your mind when you think about doing content analysis with a computer?
2. How do you think we should evaluate the analyses shown by a computer?

Three questions + the recommended reading

1. What comes to your mind when you think about doing content analysis with a computer?
2. How do you think we should evaluate the analyses shown by a computer?
3. If we distinguish research methods as: exploratory, constructive (i.e., theory building) and empirical - from what you read in the recommended book chapter, what is a more prominent research method for computational linguists?

What is Computational Linguistics?

- Study of developing computational models to process and understand human language.

What is Computational Linguistics?

- Study of developing computational models to process and understand human language.
- What is involved?
  - Understanding of how language works (Linguistics)
  - Understanding how computers work (Computer Science)
  - Understanding how to combine these to make computers understand human language (Computational Linguistics)

Some applications of Computational Linguistics

- Apple's Siri, Microsoft's Cortana, OK Google etc.
- Google Translate
- Spam classification in your email applications
- Information extraction in gmail or Facebook (e.g., identifying meeting times)
- Search engines
- Chat bots (you may not even know you are talking to a machine!)
- Booking hotel accommodations, flights etc online with voice input systems
- Spelling and grammar checkers

.... and many more!

What are some big questions computational linguists think about?

- What does a computer see when we write/type?
- What does it see when we speak?
- How can a computer "understand" human language?

What are some big questions computational linguists think about?

- What does a computer see when we write/type?
- What does it see when we speak?
- How can a computer "understand" human language?
- What can a computer easily do with language that we cannot do easily and quickly?
- What is difficult for a computer that is simple for humans?

What are some big questions computational linguists think about?

- What does a computer see when we write/type?
- What does it see when we speak?
- How can a computer "understand" human language?
- What can a computer easily do with language that we cannot do easily and quickly?
- What is difficult for a computer that is simple for humans?
- How do we make software that understands human language?
- How do we make such software work for many languages?

CL is inter-disciplinary by definition

CL is inspired by methods and ideas from:

- Linguistics
- Computer Science and Engineering
- Human-Computer Interaction
- Cognitive Science
- Psychology
- Logic

- thus, there are science, engineering, humanities and social sciences - a little bit of everything within CL!

my take on different aspects of CL:
engineering $>$ science $>$ social science $>$ humanities

Research in CL

Research in CL

- Theoretical research
  - Research driven by scientific questions
- Empirical research
  - Research driven by real-world applications
  - Research driven by technological innovations

Some theoretical research questions in CL

1. How do we represent morphology on computers? (word structure)?
2. How do we represent a language's grammar?
3. How do we represent semantics (i.e., meanings) ?
4. How do we represent general discourse (beyond sentences)?
5. How do we understand the rhetorical structure of a piece of text?

Some research questions in CL, grounded in real-world applications

1. How can a computer perform machine translation?
2. How can it answer questions? (beyond searching for keywords)
3. Can it summarize content? (news summaries)
4. Can it create new content? (generate weather reports)
5. How can we extract important information from text? (e.g., meeting scheduling from email)
6. Can we get leverage from recent advances in deep learning?

... and so on

Research Methods in CL

Research Methods in CL

- Since CL is inter-disciplinary, research methods also vary.
- Depending on the research question, you incorporate methods from CS, Linguistics, Cog. Sci., HCI etc.,

Theoretical CL research methods

1. Methods to formally specify different aspects of a language
2. Methods to study the complexity of such programs, and optimization techniques
3. Methods to develop and validate theories about language from large amounts of data
4. Computer programs that convert such descriptions into machine language

Some examples

1. finite state machines, automata theory
2. graph theory
3. linguistic theories
4. different grammar formalisms and their computational models
5. search algorithms

.. .. ..

Empirical research methods

1. Computer programs that process and analyze human language
2. Methods to collect and store large corpora
3. Methods to do predictive analysis
4. Methods to evaluate user reactions

(methods derive from multiple sources, depending on purpose)

Some examples

1. probabilistic models of language
2. machine learning methods (learning from data)
3. human computer interaction
   (e.g., in speech input, in typing interfaces etc)
4. statistical significance testing (which model gives a better result? Is it by chance? or real?)
5. A/B testing (Which modeling change impacts users?)

.. .. ..

CL Today

- Early CL research focused on formal theories of language, building grammar rules for languages etc.
- Current day CL research has a strong empirical focus.
- Primary goal is to automatically derive knowledge about language structure, instead of specifying rules and logic.

Empirical CL

- What does empirical mean?

Empirical CL

- What does empirical mean?
  $\Rightarrow$ Relying on data/observations/experiments
- Not all empirical CL research starts with some theory.
  Sometimes, empirical research is used to actually understand complex, large scale real world data
  (e.g., is the new movie getting good reviews from the public?)

Empirical CL

- What does empirical mean?
  $\Rightarrow$ Relying on data/observations/experiments
- Not all empirical CL research starts with some theory.
  Sometimes, empirical research is used to actually understand complex, large scale real world data
  (e.g., is the new movie getting good reviews from the public?)
- Sometimes, it is motivated by technological relevance
  (e.g., how do we build and evaluate machine translation software?)

Empirical CL

- What does empirical mean?
  $\Rightarrow$ Relying on data/observations/experiments
- Not all empirical CL research starts with some theory.
  Sometimes, empirical research is used to actually understand complex, large scale real world data
  (e.g., is the new movie getting good reviews from the public?)
- Sometimes, it is motivated by technological relevance
  (e.g., how do we build and evaluate machine translation software?)
- Sometimes, empirical research complements theories (testing theories with observations, identifying errors in theories etc)

Empirical CL

- What does empirical mean?
  $\Rightarrow$ Relying on data/observations/experiments
- Not all empirical CL research starts with some theory.
  Sometimes, empirical research is used to actually understand complex, large scale real world data
  (e.g., is the new movie getting good reviews from the public?)
- Sometimes, it is motivated by technological relevance
  (e.g., how do we build and evaluate machine translation software?)
- Sometimes, empirical research complements theories (testing theories with observations, identifying errors in theories etc)
- Sometimes, empirical research may build theories too!

Typical Process of Empirical CL Research

- What is the problem to be solved?
- Where is the data relevant to this problem?
- Collect and explore data (sometimes, data is already there from somewhere else)
- Form some hypotheses
- Test them by developing probabilistic and statistical models based on a part of the given data
- Evaluate the "model" on the other part
- Keep refining the model repeating the last two steps, until some stopping point.

Empirical CL research methods- an example

Towards Evaluating Narrative Quality In Student Writing

Swapna Somasundaran[1], **Michael Flor**[1], **Martin Chodorow**[2]
Hillary Molloy[3] **Binod Gyawali**[1] **Laura McCulla**[1]
[1]Educational Testing Service, 660 Rosedale Road, Princeton, NJ 08541, USA
[2]Hunter College and the Graduate Center, CUNY, New York, NY 10065, USA
[3]Educational Testing Service, 90 New Montgomery Street, San Francisco, CA 94105, USA
{ssomasundaran,mflor,hmolloy, bgyawali,LMcCulla}@ets.org
martin.chodorow@hunter.cuny.edu

https://www.transacl.org/ojs/index.php/tacl/article/viewFile/1237/284

Evaluating Narrative Quality In Student Writing

- Goal: develop an automatic scoring software that evaluates narrative essays written by students

Evaluating Narrative Quality In Student Writing

- Goal: develop an automatic scoring software that evaluates narrative essays written by students
- Need: Existing scoring engines (e.g., by ETS) focus on informational, argumentative, persuasive essays, and on evaluating language proficiency.

Evaluating Narrative Quality In Student Writing

- Goal: develop an automatic scoring software that evaluates narrative essays written by students
- Need: Existing scoring engines (e.g., by ETS) focus on informational, argumentative, persuasive essays, and on evaluating language proficiency.
- Contributions of this research:
  1. Manually annotated 942 essays for narrative quality, following the rubric adapted from Common Core State Standards.
  2. They then used these essays as the basis to develop an computer based scoring model
  3. They then performed an analysis of which essays are easy to score for a machine, which are not, and why.

Research Questions

1. Is it possible to score narrative quality traits in essays using automated methods?
2. Which of our feature sets are effective for scoring narrative quality traits?
3. How do these features perform compared to a generic essay scoring system predictions?
4. How does the overall score differ from individual traits?
5. What are the best features to do automated narrative scoring?

Methods: Corpus

- Narrative essays written by school students (grades: 7, 10, 12) in the Criterion program by ETS.
- Each essay is around 300 words long, was written in response to one of the 18 prompts. Example prompt looks like these:

[Personal Experience] There are moments in everyone's lives when they feel pride and accomplishment after completing a challenging task. Write a story about your proudest moment.
[Hypothetical Situation] Pretend that one morning you wake up and find out that you've become your teacher for a day! What happened? What do you do? Do you learn anything? Write a story about what happens. Use your imagination!
[Fictional Story] Throughout the years, many have placed messages in sealed bottles and dropped the bottles into the ocean where they eventually washed up on foreign shores. Occasionally the finder has even contacted the sender. Write a story about finding your own message in a bottle.

Methods: Corpus - 2

- Scoring rubric consisted of 3 traits (organization, development, convention), and several sub-traits.
- Each essay was scored by two experienced annotators, and adjudicated by a third researcher.
- Scoring scale: 0-4 for Organization, 0-4 for development, 0-3 for convention (0-11 total score range)
- Sub-traits were also scored in the 0-4 range.
- Annotators did not participate in software development.

Methods: Corpus - 3

Trait:Sub-trait	**QWK**
Organization	0.71
:Plot	0.62
:Characters/Setting/POV	0.65
:Transitioning	0.57
:Sequencing	0.63
:Opening/Closing	0.66
Development	0.73
:Characters/Setting/Events	0.68
:Narrative Techniques	0.64
:Language	0.59
:Source Materials	0.52
:Style	0.58
Convention	0.46
Narrative (Org. + Dev.)	0.76
Total (Org. + Dev. + Conv.)	0.76

Table 2: Inter-annotator agreement

Next Step

- Use the scoring rubric as the guideline to explore construct relevant linguistic features that can do automated scoring.
- Use the corpus as a gold standard to develop a software system to automatically score new narrative essays.

Methods: Linguistic Feature Engineering

Computer programs extract the following information from the essays are written:

- Discourse cues used in writing stories (e.g., location/temporal connections)

Methods: Linguistic Feature Engineering

Computer programs extract the following information from the essays are written:

- Discourse cues used in writing stories (e.g., location/temporal connections)
- Event chain (sequences of event that shares something e.g., an actor, a object)

Methods: Linguistic Feature Engineering

Computer programs extract the following information from the essays are written:

- Discourse cues used in writing stories (e.g., location/temporal connections)
- Event chain (sequences of event that shares something e.g., an actor, a object)
- Subjective language (e.g., adjectives used to describe character traits)

Methods: Linguistic Feature Engineering

Computer programs extract the following information from the essays are written:

- Discourse cues used in writing stories (e.g., location/temporal connections)
- Event chain (sequences of event that shares something e.g., an actor, a object)
- Subjective language (e.g., adjectives used to describe character traits)
- Providing specific details in the text, diversity of vocabulary
- Usage of pronouns, stative verbs

... and so on.

Methods: Building automated models

- Baseline: E-Rater (ETS) score (this is used in GRE too)
- Modeling methods: Different regression methods, random forests, elastic net (machine learning)
- Choice of linguistic features: Models with different subsets of linguistic features.
- Bootstrapping experiments were performed to test for statistical significance. (i.e, do multiple experiments by sampling a part of the data each time).
- Evaluation measure to compare models: Quadratic Weighted Kappa (QWK) between human and machine scores.

Methods: qualitative analysis

- Manual analysis of cases where a machine prediction diverges from human prediction a lot (say: human gives a score of 10 and machine gave it a 1).
- Finding: human-machine disagreement is related to the length of an essay (machine tends to give high scores are longer, even if they are poorly written).

Then what?

- Draw conclusions based on all the results obtained by following these methods
- Do the results address our questions, and to what extent?
- What is left? or what needs to be explored further?

... and it goes on.

Some Ethical questions in CL research

- Transparency: Issues related to reproducibility of research (releasing code and data)
- Privacy: Things to take care of when collecting user data (e.g., working with tweets)
- Fairness: Is there some bias in the machine learnt models? (e.g., an essay scoring algorithm being biased towards German learners of English)

Summary

- CL is very inter-disciplinary, and has a mix of research methods.
- Research Methods depend on the kind of research problem.
- Research in CL has a strong empirical bent.
- Some methods for empirical research include: computer programming, machine learning, linguistic feature engineering, statistical analyses
- There are some ethical questions in CL research, and one needs to be aware of those.

Thank you!
Questions?

Thank you!

Questions?

For further questions, contact: sowmya@iastate.edu

References

- Hossein Hassani. 2017. Research in Computer Science: The Challenges and Issues. https://arxiv.org/pdf/1703.04080.pdf
- Gordana Dodig-Crnkovic. 2002. Scientific Methods in Computer Science. http://www.mrtc.mdh.se/publications/0446.pdf
- Somasundaran, S., Flor, M., Chodorow, M., Molloy, H., Gyawali, B., & McCulla, L. (2018). Towards Evaluating Narrative Quality In Student Writing. Transactions of the Association for Computational Linguistics, 6, 91-106.