CSC2604: Topics in Human Centered and Interdisciplinary Computing

Computational Models of Semantic Change

Date/Time: Wednesday, 1-3pm

Location: BA026 (basement of Bahen)

Instructor: Yang Xu
Contact: yangxu@cs.toronto.edu
Office: DL Pratt 390
Office Hours: By appointment

This syllabus may be subject to adjustments as the course progresses.

Course Description: Words are core components of language, but their meanings tend to vary over time, e.g., *face* ('body part'→'expression'), *grasp* ('action'→'understanding'), *gay* ('happy'→'homosexual'). Meaning changes like these allow the lexicon to express emerging ideas with existing words, but they present potential challenges for computers to learn word meanings and their extensions. This seminar course explores theoretical and computational approaches to semantic change. Main topics include word representation, detection and modeling of meaning change, semantic chaining, and metaphor.

Note: Proficiency in Python is assumed. Some knowledge of natural language processing, probability, and information theory will be helpful.

Objectives: This course is aimed at the following three objectives.

1. Develop a broad foundation for the interdisciplinary study on semantic change.
2. Develop extensive technical skills in computational modelling of text data.
3. Develop essential communicative skills in scientific presentation and writing.

Textbook: No single textbook will be used. Instead, the course will rely on a combination of published papers and book chapters.

Recommended Readings:

- Traugott, E.C., & Dasher, R.B. *Regularity in semantic change*. CUP. 2001.
- Sweetser, E. *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. CUP. 1991.
- Hopper, P.J., & Traugott, E.C. *Grammaticalization*. CUP. 2003.
- Lakoff, G. *Women, fire, and dangerous things: What categories reveal about the mind*. UCP. 1987.

Deliverables and Assessments:

Presentation	30%
In-class discussion and labs	20%
Project proposal	5%
Project report	15%
Project presentation	15%
Code repository	15%

Letter Grade Scale:

90 - 100%	A+	77 - 79%	B+
85 - 89%	A	73 - 76%	B
80 - 84%	A-	70 - 72%	B-
		0 - 69%	Fail

Course Policies:

- **General**
  - Students are expected to present and lead discussion on at least 1 technical paper.
  - Students with scheduled presentations are required to send the slides to the instructor on the day(s) before the presentations.
  - Late assignments will receive a 1 point deduction per delayed hour until no point can be further deducted.
- **Labs**
  - Students should attempt labs in class and may work in pairs.
- **Projects**
  - Students are expected to work independently on projects.
  - Students may obtain instructor's permission to work on their own research projects, provided they are relevant to the course.
  - Students may proceed with their projects only if the initial proposals have been approved by the instructor. Otherwise they may do so until the revised proposals have been approved.
  - Students who plan to work with exeedingly large amounts of data may consult with the instructor on feasibility and server access.
- **Attendance**
  - Attendance is expected generally and required on days of presentation.
  - Students are responsible for all missed assignments due to absence, unless they notify the instructor at least two days prior to the due date.

Tentative schedule (items with '*'s are optional):

Week	Content	Reading
Week 1	• Semantic change and computational intelligence • Lab 0: Probabilistic n-gram model	TD01(ch2); Sha48*
Week 2	• Latent models of word meaning (LSA) • Lab 1: Meaning construction from text	RG65; LD97
Week 3	• Latent models of word meaning (word2vec) • Extension of Lab 1 (Lab 1 due)	Mik13a; Mik13b; GL14*
Week 4	• Limitations of word embeddings • Lab 2: Diachronic analysis of meaning (Extension of Lab 1 due)	Nem17; BM17*
Week 5	• Automatic detection of semantic change	Ham16; Dub17; CS10*; Sag09*
Week 6	• Probabilistic models of semantic change • Project requirements and resources (Lab 2 due)	FL16; BL06*
Week 7	• Cognitive models of semantic change • Project proposal due	Ram18; Lak87(ch6&7)*
Week 8	• Word sense disambiguation and induction	Yar95; Bar15*; Nun79*
Week 9	• Research Topic (TBD)	TBD
Week 10	• Project clinic	
Week 11	• Final project presentation • Final report due in 1 week	

Resources:

- Python:

  Jupyter: https://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/

  Natural Language Processing with Python: http://www.nltk.org/book/

  Natural Language Toolkit: http://www.nltk.org/

  Bare essentials: http://www.cs.toronto.edu/~yangxu/PythonBookletV4.pdf

- GitHub:

  Creating a repo: https://help.github.com/articles/create-a-repo/

  Common commands: https://gist.github.com/jedmao/5053440

- Word embeddings:

  Word2vec: https://code.google.com/archive/p/word2vec/

  GLOVE: https://nlp.stanford.edu/projects/glove/

  Lda2vec: https://github.com/cemoody/lda2vec

  tSNE: https://github.com/paulorauber/thesne

  HistWords: https://nlp.stanford.edu/projects/histwords/

- Longitudinal text corpora:

  Project Gutenberg: https://www.gutenberg.org/

  Google N-grams: http://storage.googleapis.com/books/ngrams/books/datasetsv2.html

  Syntactic N-grams: http://commondatastorage.googleapis.com/books/syntactic-ngrams/index.html

  Helsinki Corpus of English: http://www.helsinki.fi/varieng/CoRD/corpora/HelsinkiCorpus/

  Early English Books Online: https://corpus.byu.edu/eebo/

  CHILDES: https://childes.talkbank.org/

- Lexical resources:

  WordNet: https://wordnet.princeton.edu/

  MetaNet: https://metanet.icsi.berkeley.edu/metanet/

  Metaphor Map of English: http://mappingmetaphor.arts.gla.ac.uk/

Historical Thesaurus of English: http://historicalthesaurus.arts.gla.ac.uk/

Dictionary of Old English: https://www.doe.utoronto.ca/pages/index.html

- Benchmark data:

  WordSimilarity-353: http://www.cs.technion.ac.il/~gabr/resources/data/wordsim353/

  SimLex-999: https://www.cl.cam.ac.uk/~fh295/simlex.html

  SemEval-2017: http://alt.qcri.org/semeval2017/index.php?id=tasks

  Stanford Question Answering: https://rajpurkar.github.io/SQuAD-explorer/

- Human behavioural data:

  University of South Florida Free Association Norms: http://w3.usf.edu/FreeAssociation/

  Human Brain Cloud: http://www.humanbraincloud.com/

  Word concreteness ratings: http://crr.ugent.be/archives/1330

  Word affectiveness ratings: http://crr.ugent.be/archives/1003

  Word age-of-acquistion norms: http://crr.ugent.be/archives/806