Course Description

Human language technology for text and spoken language. Machine learning, syntactic parsing, semantic interpretation, and context-based approaches to machine translation, text mining, and web search. Prerequisite(s): ELEC 2600 OR IEDA 2510 OR IEDA 2520 OR IEDA 2540 OR MATH 2411 OR MATH 2421 OR MATH 2431; Exclusion(s): COMP 5221

List of Topics

We will choose from the topics is in the list below.

<table>
<thead>
<tr>
<th>topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Welcome; introduction; survey; administrivia (honor statement, HKUST classroom conduct)</td>
</tr>
<tr>
<td>Does God play dice? Assumptions: scientific method, hypotheses, models, learning, probability; linguistic relativism and the Sapir-Whorf hypothesis; inductive bias, language bias, search bias; the great cycle of intelligence</td>
</tr>
<tr>
<td>Languages of the world</td>
</tr>
<tr>
<td>Learning to translate: engineering, social, and scientific motivations</td>
</tr>
<tr>
<td>Language structures thought (Sapir-Whorf hypothesis); &quot;It’s all Chinese to me&quot;: linguistic complexity; challenges in modeling translation [at tutorial]</td>
</tr>
<tr>
<td>Is machine translation intelligent? Interactive simulation; Turing test; translation based test of AI; error analysis; what's still missing in AI</td>
</tr>
</tbody>
</table>
Introduction to search

State spaces; the anagram problem

Conditional decomposition; Markovian approximations; memoryless property; Markov models

Character n-gram models; goal states and objective functions

Uninformed search; BFS, DFS, depth-bounded search, iterative deepening, bidirectional search

Informed search; greedy best-first search; agenda-driven search; search fringe; Dijkstra's shortest path algorithm

implementing agenda-driven search for finding best anagrams (Assignment due TBA); physical demo of agenda-driven search for Dijkstra's algorithm; anagrams with replacement

Basic probability theory; conditional probabilities; Bayes' theorem

Chinese anagrams; word vs character n-grams; first- and second-order n-grams; Shannon

hidden Markov models, finite-state models; weighted and stochastic FSAs; parts of speech; generation vs recognition/parsing

Converting state-based to transition based FSAs for both logical and stochastic FSAs; segmental HMM/SFA/WFSAs; WFST: finite-state translation models

HMM/SFSA/WFSA decoding, evaluation, learning: unrolling in time; ways to search the lattice; formalization for Viterbi decoding and evaluation
Introduction to neural networks and their language biases

Implementing feedforward neural network n-gram models; model design following scientific method for machine learning/adaptation in practice (Assignment due TBA)

Search bias; forward algorithm for HMM/SFSA/WFSAs

Training HMM/SFSA/WFSAs: backward algorithm; completing missing data; expectations; EM (expectation maximization); Baum-Welch parameter estimation

RNN: recurrent neural networks

Implementing RNNs; RNN language models; recurrent model design (Assignment due TBA); memoization/caching; HMM alignment models; memory-bounded search; heuristics; admissibility; A* search; memory-bounded heuristic search

Revisiting knowledge representation and language bias; propositional logic; conjunctive normal form; AND/OR graphs; AND/OR hypergraphs; forward chaining

Definite clauses; definite clause grammars; Knuth's algorithm; context-free grammars (CFGs); weighted and stochastic CFGs

Probabilistic dynamic programming based chart parsing; probabilistic Cocke-Kasami-Younger (CKY) parsing; inside-outside algorithm; parsing by theorem proving; productions as inference rules; backward chaining; abductive (diagnostic or explanatory) inference vs deductive inference

From CFGs to ITGs (monolingual vs bilingual modeling); how bilingual conditions make grammar induction easier; the mystery of the magic number 4 in semantic frames; simple and full syntax-directed transduction grammars (SDTGs); tree vs matrix constituent alignment visualizations
Exploring inversion transduction grammars (ITG); ITG characteristics; stochastic ITGs; polynomial-time transduction and learning; resolving the mystery of the magic number 4

Evaluating translation quality: alignment; aligning semantic frames: Interactive exercise

Evaluating translation quality: MEANT

Evaluating translation quality: semantic role labeling (SRL); case frames, semantic frames, predicate-argument structure

Automatic semantic role labeling (ASRL)

Implementing a feedforward neural network based part-of-speech tagger
Assignment due TBA; context-independent POS tagging

I/O representations for feedforward networks; context-dependent POS tagging

Basic probability theory; conditional probabilities; Bayes' theorem

Example-based, instance-based, memory-based, case-based, analogy-based, lazy learning for classification; translation via nearest neighbors (NN); k-NN; weighted k-NN

Learning vs performance components in machine learning; supervised learning; Word sense disambiguation; lexical choice; example-based prediction models; nearest neighbor classifiers; similarity metrics; kNN classifiers

Exploring different feedforward neural network architectures for POS tagging; model design following scientific method for machine learning in practice
Assignment due TBA
Naive Bayes classifiers for WSD and lexical choice

Modern approaches to SRL

Implementing chunkers and shallow parsers via IOBES tagging plus a POS tagger
Assignment due TBA

Chunking via IOBES representations; shallow bracketing

Shallow syntactic parsing; shallow semantic parsing; language bias of IOBES representations, bags of words, and one-hot representations

Introduction to word embeddings

Vector space models; classic word vector approaches

Learning word embeddings via prediction tasks; skip-grams; word2vec;

Recursive autoencoders (RAEs) and recursive auto-associative memories (RAAM); learning word embeddings by making RAEs predict

Context-free grammars (CFGs); generative vs parsing models; top-down vs bottom-up parsing; dynamic programming based chart parsing; Cocke-Kasami-Younger (CKY) parsing

Recursive autoencoders (RAE) and recursive auto-associative memory (RAAM); TRAAM (transduction RAAM); recursive neural network realizations of ITGs; a self-learning rap battle bot
Textbook


Reference books

N/A

Grading Scheme

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pop quizzes</td>
<td>~10%</td>
</tr>
<tr>
<td>Class participation</td>
<td>~15%</td>
</tr>
<tr>
<td>Forum participation</td>
<td>~10%</td>
</tr>
<tr>
<td>Assignments</td>
<td>~65%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

Course Intended Learning Outcomes

At the end of the Natural Language Processing course, you will have achieved the following outcomes.

1. **General**
   1. Possess solid understanding of the fundamental concepts of natural language processing
   2. Possess solid understanding of the fundamental concepts of machine translation, and grasp how it stress tests all aspects of human intelligence and language processing

2. **Transduction**
   1. Know foundational input-output formulations of transduction, such as alignment, chunking, classification, dependency relations, and parsing
   2. Understand the relationship between noisy channel and loglinear models of string transduction, and their Bayesian interpretations

3. **Syntax**
1. Understand the relationship between word segmentation and phrasal lexicons, the relationship to transduction and alignment, and associated algorithms
2. Understand the relationship between traditional grammatical formalisms versus stochastic and weighted grammars
3. Understand the strengths and weaknesses of part-of-speech models, and associated tagging algorithms
4. Understand the various fundamental approaches to parsing, and how they deal with syntactic ambiguity

4. **Alignment**
   1. Understand how bilingual models of syntax generalize upon monolingual models to improve learnability
   2. Understand the combinatorial and empirical trade-offs between various learning models of alignment and compositionality, and their associated algorithms
   3. Understand the core methods for inducing lexicons, translation lexicons, phrasal translation lexicons, as well as permutation and reordering models

5. **Decoding**
   1. Understand the combinatorial and empirical trade-offs between various runtime models for translation, and their associated algorithms
   2. Understand how bilingual transduction models generalize upon monolingual parsing models

6. **Semantics**
   1. Understand lexical semantics models for word sense disambiguation, their relationship to phrasal lexicons and transduction, and associated ambiguity resolution algorithms
   2. Understand lexical semantics models for semantic frames (predicate-argument structures), and associated semantic role labeling algorithms

**Assessment Rubric**

N/A