

Natural Language Processing and Big Data for Psychology and Computational Linguistics

Overview

Over the past decade, there has been a remarkable cross-fertilization between the fields of psycholinguistics and natural language processing with insights from each field enhancing the development of theory, methodology, and resources in the other.

Developments in statistical NLP have led to a much larger availability of textual resources for psycholinguistic research and have made it possible for psycholinguists to quickly develop specific resources such as corpora and word-frequency lists. As a result, there has been an explosive growth in behavioral and linguistic data that can be leveraged for psycholinguistic research.

Our increasingly networked and technological society has also spurred development of techniques such as natural language understanding and machine translation. This has led to a re-birth of artificial intelligence known as deep learning, which can be traced back to learning theory developed in psychology in (Rescorla & Wagner, 1977) and can again be applied to research on human language processing (e.g., Mandera, Keuleers, & Brysbaert, 2017).

These developments are quickly changing psycholinguistic research. From a field that for decades has been dominated by small-scale one-time controlled laboratory experiments it is becoming a dynamic research enterprise relying on reusable and distributed data generation processes leveraging crowd participation (Keuleers & Balota, 2015).

Knowledge of these developments and techniques is becoming indispensable for researchers working in any area of cognitive science that deals directly or indirectly with language or linguistic resources.

Objectives

The primary objectives for this course are:

- 1) To explore the mutual connections between the fields of natural language processing and language psychology, from a historical perspective.
- 2) To teach participants natural language processing techniques that can be applied to behavioral research.
- 3) To instill in participants the mindset to do creative psycholinguistic research that takes full advantage of NLP techniques and big data.

Modules	Dates: December 10th, 2018 – December 15th, 2018.
	Day 1 (10th December)
	Module A: Human and Machine Learning of Language
	Lecture 1 (EK, 1.5 hrs) Language, behaviorism, and cognitive-science.
	Lecture 2 (EK, 1.5 hrs) Classical conditioning, operant conditioning, the Rescorla-Wagner model, the perceptron, recurrent neural networks, deep learning.
	Day 2 (11th December)
	Lecture 3 (EK, 1.5 hrs) Discriminative and associative learning of human language, distributional semantics
	Module B: Natural Language Processing
	Tutorial 1 (EK, 2 hrs) Basic python, regular expressions, tokenizing.
	Day 3 (12th December)
	Tutorial 2 (EK, 2 hrs) Lemmatizing, named-entity recognition.
	Tutorial 3 (EK, 2 hrs) Counting, word frequency distributions.
	Day 4 (13th December)
	Tutorial 4 (EK, 2 hrs) TFxIDF, Feature Extraction, Classification, Machine Learning algorithms

	<p>Tutorial 5 (EK, 2 hrs) Vector-Space models</p> <p>Day 5 (14th December)</p> <p>Module C: Crowdsourcing and Big Data</p> <p>Lecture 4 (EK, 1.5 hrs) Social Intelligence, Social Media Analytics 2</p> <p>Lecture 5 (EK, 1.5 hrs) Crowdsourcing of behavioral data</p> <p>Day 6 (15th December)</p> <p>Module D: Theoretical applications</p> <p>Lecture 6 (EK, 1.5 hrs) Frequency, visual word recognition, grammaticality judgments.</p> <p>Lecture 7 (EK, 1.5 hrs) Aging, multilingualism, transaction records.</p>
Who can attend?	<p>The course can be attended by students from diverse fields as computer science, linguistics, cognitive psychology, cognitive sciences, psycholinguistics, Natural Language Processing, Data Mining, Big Data and so on. Students at all levels BTech/MTech, BSc./MSc. and from early level PhDs across these disciplines and also interested young faculty from reputed universities and technical institutions are welcome to attend the course.</p>
Fees	<p>The participation fees for taking the course is as follows:</p> <p>Participants from abroad : US \$200 Industry/ Research Organizations: 15000 INR Academic Institutions: 5000 INR Students: 2000 INR</p> <p>The above fee include all instructional materials, computer use for tutorials and assignments, laboratory equipment usage charges, 24 hr free internet facility. The participants will be provided with accommodation on payment basis.</p>

The Faculty



Dr. Emmanuel Keuleers is an Assistant Professor in Department of Cognitive Science and Artificial Intelligence at Tilburg University (The Netherlands). His research is situated at the intersection of natural language processing, machine learning, theoretical linguistics, and psychology. He is the author of numerous well-cited publications on new techniques, datasets, and theoretical findings in these fields. His research has been instrumental in establishing the use of techniques from natural language processing and computational linguistics in psycholinguistic investigation. He was the editor of a recent special issue of the Quarterly Journal of Experimental Psychology on "Megastudies, crowdsourcing, and big datasets in psycholinguistics". He is a consulting editor for Behavior Research Methods, a member of the editorial board of the Mental Lexicon and a regular contributor to top journals in his fields of research. His current work focuses on the computational modeling of human language processing following design principles from information theory, discrimination learning, and network science.

Course Coordinator

Dr. Ark Verma,
Assistant Professor of Psychology,
IIT Kanpur.
Phone: +91-512-6847 (O), +91 - 9506304191
E-mail: arkverma@iitk.ac.in
