Big Data and Networks in Social Sciences - Abstracts of talks

Mining socio-political and socio-economic signals from social media content

Date: 21 September
Speaker: Vasileios Lampos
Abstract:

Over the past few years user-generated content, primarily originating from social media platforms, has been the focus of considerable research effort. One of the main motivations is the potential of using this data as an additional source of information for enhancing our understanding of societal or more personalised signals. This talk will showcase a series of ideas which explored that basic hypothesis, proposing statistical natural language processing frameworks for the estimation of collective patterns (e.g. emotions or voting intentions) or of more focused demographic user attributes (e.g. the occupation, income and socioeconomic status). Apart from the essential statistical evaluation of the proposed methods, interesting insights, stemming from a qualitative analysis, will be presented as well.

Compositional and Multimodal Distributional Semantics

Date: 21 September
Speaker: Stephen Clark
Abstract:

There have been two main approaches to modelling the meaning of language in Computational Linguistics and Natural Language Processing. The first, the compositional approach, is based on classical ideas from Philosophy and Mathematical Logic, and implements the ideas from Formal Semantics which followed Montague's work in the 1970s. The second, more recent approach focuses on the meanings of the words themselves. This is the distributional approach to lexical semantics and is based on the ideas of structural linguists such as Firth, and is sometimes related to the Wittgensteinian philosophy of "meaning as use". The idea is that the meanings of words can be determined by considering the contexts in which words appear in text.

In this talk I will describe two shortcomings of distributional semantics and our attempts to overcome them. The first shortcoming is that distributional semantics is designed to induce word meanings, whereas meaning more generally is mediated through phrases and sentences. Hence there is a new research problem in natural language semantics, which is to combine compositional and distributional semantics. The second shortcoming is that distributional models are typically induced from text alone, whereas the meaning of a concept is often related to its interaction with other modalities, such as vision. Hence we are attempting to enrich the distributional representation of a concept by considering feature norms describing the concept, images and auditory representations of the concept, and even an olfactory (smell) representation. The final part of the talk will describe attempts to induce mappings between these various modalities.

Title: Practical machine learning for social media analysis

Date: 21 September
Speaker: Isabelle Augenstein
Workshop outline:

Social media is a rich resource which can help us understand how people interact, what topics are discussed, what the opinion towards public figures is and what rhetoric they use. This tutorial will introduce machine learning methods commonly used for data analysis tasks, illustrated with social media data and Python code examples.

Participants will learn:

1. How to use Python, scikitlearn, gensim and TensorFlow for data analysis tasks
2. How to represent features
3. How to detect patterns in data, such as topics discussed in social media
4. How to classify data, such as the sentiment towards public figures
5. How to determine the similarity between words and phrases
6. How to model sequential data with neural networks

Participant information: Participants are expected to have basic familiarity with machine learning concepts and Python.

Understanding Unemployment in the Era of Big Data

Date: 22 September
Speaker: Omar Guerrero
Abstract:

On one hand, unemployment is a central issue in all countries. On the other the economic policies designed to mitigate it are usually built on theoretical grounds that are validated at an aggregate level, but have little or no validity from a micro point of view. This situation is a cause for concern because policies are designed and implemented at the level of individuals and organisations, so ignoring realistic micro-mechanisms may lead to undesirable outcomes in the real world. Ironically, the data to inform theoretical frameworks at the micro-level has existed in labour studies since the 1980's. However, it is only now that we count with analytical methods and computational tools to take full advantage of it. In this paper we argue that big data from administrative records, in conjunction with network science and agent-computing models offer new opportunities to inform theories of unemployment and improve policies. We introduce a data-driven model of unemployment dynamics that is validated at both the micro and macro-levels. At a first glance, validation at the micro-level seems unnecessary since we focus on aggregate unemployment. However, by establishing a connection between our model and the ones commonly used to advice policy we show that overlooking micro-level validity leads to erroneous predictions with significant real-world consequences.

Intermediation networks

Date: 22 September
Speaker: Daniele Condorelli
Abstract:

Too many questions, not enough data.

Date: 23 September
Speaker: Thomas House
Abstract:
A common problem when dealing with complex phenomena, whether social, physical or biological, is that the data we receive have typically not been generated specifically to answer the hypotheses of interest. Even if these data are very big' indeed. I will outline one broad methodological approach to this problem based on mechanistic modelling coupled to computationally intensive statistical methods.

Hiring through Networks: Favors or Information?

Date: 23 September
Speaker: Yann Bramoullé
Abstract: