Lecture series: Data Science in Action

Zoom, Thursdays, 12 – 1.30 pm

Please note that the syllabus will be updated if necessary

15 September Prof. Dr. Simone Paolo Ponzetto (Chair of Information Systems III: Enterprise Data Analysis, University of Mannheim)

Lecture topic: Towards fair human language technologies through debiasing of semantic spaces

In recent years, much work has started looking at ethical aspects and limitations of Natural Language Processing (NLP) systems. NLP models, for instance, have been found to encode human biases, most notably gender and racial biases. This is indeed timely and essential work, since NLP systems are becoming part of our everyday life at a breathtaking pace and, as such, they provide us with yet another example of the societal challenges posed by the widespread adoption of AI technologies. In this lecture, I will give an introductory overview of digital ethics in NLP systems and focus on a few examples of methods and frameworks to measure and mitigate stereotypical bias in semantic spaces acquired from textual data.

22 September Prof. Dr. Hartmut Wessler (Institute for Media and Communication Studies, University of Mannheim)

Lecture topic: “Glass Box" Machine Learning in Text Classification

The talk first introduces the task of measuring the “integrative complexity” of social media posts as an example for studying the quality of public debate. The traditional manual coding approach is then juxtaposed with computational approaches to measuring integrative complexity. In particular, a combination of a theory-informed dictionary approach with “glass box ML” (as opposed to “black box” machine learning algorithms) is introduced. The talk will demonstrate how this combination cannot only process larger amounts of data in less time, but also aid in theory development.

29 September Prof. Dr. Heiko Paulheim (Chair for Data Science, University of Mannheim)

Lecture topic: Knowledge Matters! The Role of Knowledge Graphs in Modern AI Systems

When we think about modern AI systems, we often think about Machine Learning in the first place. But there is more to AI than simply training models - in particular, AI systems also require some knowledge about the domain they operate in. In this talk, I will shed some light on knowledge graphs, which are a universal mechanism to represent and process such knowledge. I will also look into large-scale knowledge graphs which already exist and can be readily used in AI systems, which provides a leap start over starting to build AI from scratch.

6 October Prof. Dr. Teresa Naab (Institute for Media and Communication Studies, University of Mannheim)

Lecture topic: Gaining insights from social media data

In social media, people share information, publish their opinions, and react to others' postings to show support or devaluation. This user-generated content is the subject of many small scale as well as large scale analyses. The talk addresses challenges in the analysis and interpretation of social media data from a social science perspective. It considers questions of the representativity of social media data in general and biases during sampling such data. It discusses the meaning of standardized indicators (e.g. Likes) and the difficulties of analyzing discursive data (e.g. user comments). It considers challenges due to the dynamic nature of social media data, its hypertextuality, and personalization through algorithms. The talk provides examples from current research projects on online hate to illustrate these challenges.

13 October Prof. Dr. Richard Traunmüller (Chair of Political Science, Empirical Democracy Research, University of Mannheim)

Lecture topic: Improving Studies of Sensitive Topics Using Prior Evidence

Estimates of sensitive questions from list experiments are often much less precise than desired. We address this well-known inefficiency problem by presenting a unified Bayesian framework which combines indirect measures with prior information. Specifying informed priors amounts to a principled combination of information which increases the efficiency of model estimates. This framework generalizes a whole range of different design and modeling approaches for list experiments, such as the inclusion of direct items, auxiliary information, the double list experiment, and the combination of list experiments with other indirect questioning techniques.

20 October Prof. Dr. Dirk Ifenthaler (Chair of Economic and Business Education – Learning, Design & Technology, University of Mannheim)

Lecture topic: Adoption of Learning Analytics – Data-driven vs. Data-demand

Recent research focussing on the adoption of learning analytics reports high interest among educational organisations. At the same time, the maturity level of available organisation-wide learning analytics systems is still low; i.e., organisations are aware of learning analytics and start experimenting with dashboards for students and teachers. However, they are far from organisational transformation. Accordingly, learning analytics remains an interest for educational organisations rather than a major priority. In this presentation, current research findings regarding the effectiveness of learning analytics are critically examined and implications for organisational change as well as pedagogical practice will be provided. Examples of actionable frameworks and adoption models for successful integration of learning analytics systems into educational organisations will be presented, and challenges encountered during the adoption of learning analytics in educational organisations will be highlighted.

27 October Prof. Dr. Andreas Witt (Head of the Department of "Digital Linguistics" at the Leibniz Institute for the German Language, University Professor for Computational Humanities & Text Technology at the University of Mannheim)

Lecture topic: Automatic Measuring of the Non-Scientific Imapct of Research Projects

Within the project TextTranfer II at Leibniz Institute for German Language in Mannheim, a corpus-based method for measuring the impact of publicly funded research is being developed. Unlike standard bibliometric metrics, which do not allow for generalization to societal or economic impact of research outside of academia, it uses supervised machine learning technique to estimate the potential impact of research projects.

For this purpose, we built a corpus of freely available final reports on publicly funded research projects from four different domains, namely electro mobility, artificial intelligence, linguistics, and musicology. As a first step, we developed a model which automatically detects passages in the reports that indicate (any kind of) impact of the projects. We experimented with different models such as Random Forest, Naïve Bayes or embedding-based techniques, which we optimized with hand-crafted rules and heuristics such as keyword lists or the length of passages.

In a deductive, bottom-up process, we then developed an annotation scheme for categorizing different kinds of impact, such as economic, societal or technical impact, including various more fine-grained subcategories. This annotation scheme is then used for manually labelling the passages retrieved in step 1 with different impact categories.

The final annotations will then be leveraged as training- and testing data for machine learning models that are able to automatically detect and categorize impact in research reports. In a pilot study with annotated reports from the domain electro mobility, we experimented and combined various models such as Support Vector Machines (SVM), lexical (tf-idf scores) and syntactical features (part of speech), and Synthetic Minority Oversampling Techniques (SMOTE). The classification results show that we can predict impact categories with 79 F1 score for the domain electro mobility. Thus, this approach, which is mixed in its methodology, can complement bibliometric and scientometric solutions for assessing impact from research and provide insights into the extent of the influence of research results on different fields of application.

In our ongoing work, we develop machine learning models for domain independent classification of impact, using the annotated project reports from all of our four domains - electro mobility, artificial intelligence, linguistics, and musicology for training and testing.

The project is funded by the Federal Ministry of Education and Research and collaborates with the the Technical Information Library Hanover (TIB), the University of Illinois in Urbana-Champaign (UIUC, iSchool) and the Görgen & Köller Gmbh.

3 November Prof. Dr. Irene Finocchi (Professor of Computer Science, LUISS Guido Carli University, Rome)

Lecture topic: An analysis of computer science research: evolution and trends over the last 30 years

Even if computer science is a relatively young discipline, it has witnessed a quick growth and deep changes in a very short time. The talk will shed light on the evolution of the computer science research community over the past 30 years. Analyzing data from the full Scopus database, it will investigate how aspects such as the community size, gender composition, and academic seniority of its members changed over time. It will also discuss the varying popularity of specific research areas, as derived from the ACM's Special Interest Groups and IEEE classifications. The analysis spans 19 nations of the G20 group and involves a total of 728,374 authors and 8,412,543 publications.

10 November Prof. Dr. Georg Alpers (Chair of Clinical and Biological Psychology and Psychotherapy, University of Mannheim)

Lecture topic: Applications of Machine Learning in Emotion Science and Clinical Psychology

17 November Prof. Dr. Florian Stahl (Chair of Quantitative Marketing and Consumer Analytics, University of Mannheim)

Lecture topic: Computer Vision in Advertising

Computer Vision is a new technology that exploits the power of artificial intelligence to analyse images. In advertising, it has been attributed to identifying people, cars, and many more objects in a scene or advertisement. In this lecture, I will provide an overview of machine learning techniques to analyse ad content over time. In particular, I will show on a sample consisting of all issues of "The Economist" in the time period 1900 - 2014 how to analyse both textual and visual cues provided in the ads, as well as the articles of the magazine to understand how the display of emotions has developed over time. Specifically, we investigate how advertisers have expressed emotions in their print advertisements over the past century. Thereby, we consider how the emotion portrayed in ads and articles compares and whether it changes due to historical events. From a managerial perspective, our study predicts to what extent the emotions of ads and articles are synchronized with the general societal sentiment.

24 November Prof. Dr. Laurent Perrussel (Full Professor in Computer Science, University Toulouse 1 Capitole)

Lecture topic: Automated Mechanism Design

Mechanism Design aims at defining mechanisms that satisfy a predefined set of properties, and Auction Mechanisms are of foremost importance. In this talk we show that Strategy Logic provides a formal framework fit to model mechanisms, express classical properties, and verify them. To do so, we consider a quantitative variant of Strategy Logic. We will show how to express the implementation of social choice functions. Second, we will show how fundamental mechanism properties can be expressed as logical formulas, and thus evaluated by model checking.

1 December Prof. Dr. Umberto Grandi (Institut de Recherche Informatique de Toulouse, University of Toulouse 1 Capitole)

Lecture topic: Computational social choice and algorithmic game theory

Societies are taking more and more decisions mediated by algorithms, from e-democracy applications to school-student matching, to agent-mediated e-commerce. In this talk I will sketch some of the main results obtained in the literature on computational social choice and algorithmic game theory, with a strong focus on recent work I conducted on social choice on social networks and iterative voting.

8 December Prof. Dr. Moritz Fleischmann (Chair of Logistics and Supply Chain Management, University of Mannheim)

Lecture topic: Data Driven Supply Chain Planning