EMNLP-IJCNLP 2019

ECONLP — Economics and Natural Language Processing

Proceedings of the Second Workshop on Economics and Natural Language Processing (ECONLP 2019)

November 4, 2019
Hong Kong, China

Introduction

Welcome to ECONLP 2019! After launching the first Workshop on Economics and Natural Language Processing (ECONLP) at ACL 2018 in Melbourne, Australia, the second ECONLP workshop is held at EMNLP-IJCNLP 2019 in Hong Kong, China, on November 4, 2019.

This workshop reflects the increasing relevance of natural language processing (NLP) for regional, national and international economy, both in terms of already operational language technology products and systems, as well as newly emerging methodologies and techniques as a response to new requirements at the disciplinary intersection of economics and NLP. The focus of the workshop will be on the many ways how NLP influences business relations and procedures, economic transactions, and the roles of human and computational actors involved in commercial activities using natural language as primary communication mode.

The main topics ECONLP addresses include (definitely not excluding other areas of relevance for the broad workshop theme)

- NLP-based (stock) *market analytics*, e.g., prediction of economic performance indicators (trend prediction, performance forecasting, etc.), by analyzing verbal statements of enterprises, businesses, companies, and associated legal or administrative actors

- NLP-based *product analytics*, e.g., based on social and mass media monitoring, summarizing reviews, classifying and mining complaint messages and other (non)verbal types of customer reactions to products or services

- NLP-based *customer analytics*, e.g., client profiling, tracking product/company preferences, screening customer reviews or complaints, identifying high-influentials in economy-related communication networks

- NLP-based *organization/enterprise analytics* (e.g., tracing and altering social images of organizational actors, risk prediction, fraud analysis, predictive analysis of annual business, sustainability and auditing reports)

- Analysis of *market sentiments and emotions* as evident from consumers' and enterprises' verbal behavior and their communication strategies about products, services or market performance

- *Competitive intelligence* services based on NLP tooling

- Relationships and interactions between *quantitative (structured) economic data* (e.g., contained sales databases and associated time series data) and *qualitative (unstructured verbal) economic data* (press releases, newswire streams, social media contents, etc.)

- *Organizational information management* based on the content-based assembling, packaging and archiving of verbal communication streams in organizations/enterprises (emails, meeting minutes, business letters, internal reporting, etc.)

- *Credibility and trust models* for business agents involved in the economic process (e.g., as traders, sellers, advertisers) extracted from text/opinion mining their current communication as well as legacy data

- *Deceptive or fake information* recognition related to economic objects (such as products, advertisements, etc.) or economic actors (such as industries, companies, etc.), including opinion spam targeting or emanating from economic actors and processes

- Verbally fluent *software agents* (chat bots for counseling, sales and marketing) as reliable actors in economic processes serving business interests, e.g., embodying models of persuasion, information biases, fair trading

- *Enterprise search engines* (e-commerce, e-marketing) involving NLP analytics

- *Consumer search engines*, market monitors, product/service recommender systems involving NLP analytics

- *Client-supplier interaction platforms* (e.g., portals, helps desks, newsgroups) and transaction support systems based on written or spoken natural language communication

- *Multi-media and multi-modality interaction platforms*, including written/spoken language channels, triggering or supporting economic processes

- Specialized modes of *information extraction and text mining* in economic domains, e.g., temporal event or transaction mining

- *Information aggregation* from economy-related discourse, from and across single sources (e.g., review summaries, automatic threading)

- *Text generation* in economic domains, e.g., review generation, complaint response generation

- *Ontologies for economics*, lexicons for the economy domain or adaptation of general-domain lexicons for economic NLP – acquisition, maintenance or update of such terminological resources

- *Corpora and annotations policies* (guidelines, metadata schemata, etc.) for economics-related NLP

- *Economy-specific text genres* (business reports, sustainability reports, auditing documents, product reviews, economic newswire, business letters, law documents, etc.) and their usage for NLP

- Dedicated *software resources* for economic NLP (e.g., NER taggers, sublanguage parsers, pipelines or end-to-end systems for processing economic discourse)

Two types of papers were solicited for the ECONLP 2019 workshop:

- Long papers (8+1 pages) should describe solid results with strong experimental, empirical or theoretical/formal backing,

- Short papers (4+1 pages) should describe work in progress where interesting novel, yet still preliminary results have been worked out.

We received this year 17 submissions (from which 1 was withdrawn during the review process), and based on a rigorous review process, we accepted 5 of them as long papers, 3 as short papers and rejected 8 from the remaining 16 papers. Accordingly, the acceptance and rejection rate was equally 50%. The acceptance/rejection ratio amounts to 1.0.

Compared with ECONLP 2018, the first workshop on Economics and Natural Language Processing, we received one more submission, decreased the acceptance rate from 64% to 50% (conversely, we increased the rejection rate from 36% to 50%); the acceptance/rejection ratio fell from 1.9 in 2018 to 1.0 this year. Also the proportion of long papers relative to short ones increased substantially this year, up from 2 (2018) to 5 for long papers and down from 7 (2018) to 3 for short papers in 2019. Cautiously, one might interpret this change as an indication that work in this field gets deeper and riper.

Overall, the theme of this workshop seems to become more and more attractive for a wide range of neighbouring scientific communities, including NeurIPS, IJCAI, KDD, and SIGIR. Without aiming for completeness, the list below enumerates some major events with a similar thematic scope in 2019:

- *Robust AI in FS 2019 — Workshop on Robust AI in Financial Services: Data, Fairness, Explainability, Trustworthiness, and Privacy @ NeurIPS 2019*, Vancouver, B.C., Canada, December 2019

- *FinNLP & FinSDB 2019 — First Workshop on Financial Technology and Natural Language Processing (FinNLP) with a Shared Task for Sentence Boundary Detection in PDF Noisy Text in the Financial Domain (FinSBD) @ IJCAI-2019*, Macao, China, August 2019

- *ADF 2019 — Second Workshop on Anomaly Detection in Finance @ KDD 2019*, Anchorage, Alaska, USA, August 2019

- *eCom 2019 — 3rd SIGIR ACM SIGIR Workshop on eCommerce @ SIGIR 2019*, Paris, France, July 2019

- *FNP 2019 — Second Financial Narrative Processing Workshop @ NoDaLiDa 2019*, Turku, Finland, September 2019

Hence, ECONLP might constitute the institutional hub for gathering the contributions of the ACL-affine NLP community dealing with a wide range of economic topics.

Last but not least, we want to thank those colleagues who submitted their work to our workshop and hope that their efforts will strengthen a process of productive and sustainable activities in this exciting interdisciplinary domain. In particular, we also want to thank our PC members whose thorough and timely reviews were the basis for properly selecting the best-quality papers presented at this workshop. Finally, we hope the attendants of the workshop will enjoy the presentations and discussions in Hong Kong.

The organizers of ECONLP 2019

Udo Hahn
Véronique Hoste
Zhu (Drew) Zhang

Organizers:

Udo Hahn	Friedrich-Schiller-Universität Jena, Germany (Chairman)
Véronique Hoste	Ghent University, Belgium
Zhu (Drew) Zhang	Iowa State University, Ames, IA, USA

Program Committee:

Sven Buechel	Friedrich-Schiller-Universität Jena, Germany
Hsin-Hsi Chen	National Taiwan University, Taipei City, Taiwan
Paulo Cortez	University of Minho, Guimarães, Portugal
Sanjiv Ranjan Das	Santa Clara University, CA, USA
Brian Davis	Maynooth University, Ireland
Flavius Frasincar	Erasmus Universiteit, Rotterdam, The Netherlands
Petr Hájek	Univerzita Pardubice, Czech Republic
Allan Hanbury	Technische Universität Wien, Austria
Qing Li	Southwestern University of Finance and Economics, Sichuan Sheng, China
Pekka Malo	Aalto University, Finland
Julian J. McAuley	University of California, San Diego, CA, USA
Viktor Pekar	Aston University, Birmingham, England, U.K.
Paul Rayson	Lancaster University, England, U.K.
Samuel Rönnqvist	University of Turku, Finland
Sameena Shah	S&P Global Ratings, New York City, NY, USA
Kiyoaki Shirai	Japan Advanced Institute of Science and Technology (JAIST), Nomi, Ishikawa, Japan
Padmini Srinivasan	University of Iowa, Iowa City, IA, USA
Amanda Stent	Bloomberg LP, New York City, NY, USA
Heiner Stuckenschmidt	Universität Mannheim, Germany
Chuan-Ju Wang	Academia Sinica, Taipei City, Taiwan
Wlodek W. Zadrozny	University of North Carolina, Charlotte, NC, USA
Ming Zhou	Microsoft Asia Research, Beijing, China

Invited Speaker:

Marco Enriquez	Division of Economic and Risk Analysis, U.S. Securities and Exchange Commission (SEC), Washington, D.C., USA

Keynote talk: NLP for Financial Regulation

Marco Enriquez

Division of Economic and Risk Analysis, U.S. Securities and Exchange Commission (SEC),
Washington, D.C., USA

Abstract[1]

Narrative disclosures have been long-standing components of regulatory requirements (e.g., the U.S. SEC's requirement for corporate issuers to file 10-Ks). As the length of these disclosures grow in size and as the number of registrants increase, however, manual review of these corpora has become impractical or infeasible. Hence, there has been increasing interest in Natural Language Processing (NLP) among financial regulators across the world. In this talk I cover use-cases of NLP to aid in regulatory workflows: from processing to risk assessment and surveillance. I also discuss main challenges that regulators have faced to systematically deploy NLP capabilities in production environments. Finally, I detail shortcomings of current NLP technologies for financial regulation, and areas for future research and improvement.

Table of Contents

Conference Program

Monday, November 4, 2019

9:00–9:30   *Introduction to the ECONLP workshop*
Udo Hahn

9:30–10:40   SESSION 1: FINANCIAL EVENT EXTRACTION

9:30–10:00   *Extracting Complex Relations from Banking Documents*
Berke Oral, Erdem Emekligil, Seçil Arslan and Gülşen Eryiğit

10:00–10:20   *Financial Event Extraction Using Wikipedia-Based Weak Supervision*
Liat Ein-Dor, Ariel Gera, Orith Toledo-Ronen, Alon Halfon, Benjamin Sznajder, Lena Dankin, Yonatan Bilu, Yoav Katz and Noam Slonim

10:20–10:40   *A Time Series Analysis of Emotional Loading in Central Bank Statements*
Sven Buechel, Simon Junker, Thore Schlaak, Claus Michelsen and Udo Hahn

10:40–11:00   *Coffee Break 1*

11:00–12:30   SESSION 2: FINANCIAL FORECASTING AND PREDICTION

11:00–11:30   *Forecasting Firm Material Events from 8-K Reports*
Shuang (Sophie) Zhai and Zhu (Drew) Zhang

11:30–12:00   *Incorporating Fine-grained Events in Stock Movement Prediction*
Deli Chen, Yanyan Zou, Keiko Harimoto, Ruihan Bao, Xuancheng Ren and Xu Sun

12:00–12:30   *Group, Extract and Aggregate: Summarizing a Large Amount of Finance News for Forex Movement Prediction*
Deli Chen, Shuming Ma, Keiko Harimoto, Ruihan Bao, Qi Su and Xu Sun

12:30–14:00   *Lunch Break*

14:00–14:45   *Invited Talk: NLP for Financial Regulation*
Marco Enriquez (Division of Economic and Risk Analysis, U.S. Securities and Exchange Commission)

Monday, November 4, 2019 (continued)

14:45–15:35   **SESSION 3: MISCELLANEOUS TOPICS**

14:45–15:15   *Complaint Analysis and Classification for Economic and Food Safety*
João Filgueiras, Luís Barbosa, Gil Rocha, Henrique Lopes Cardoso, Luís Paulo Reis, João Pedro Machado and Ana Maria Oliveira

15:15–15:35   *Annotation Process for the Dialog Act Classification of a Taglish E-commerce Q&A Corpus*
Jared Rivera, Jan Caleb Oliver Pensica, Jolene Valenzuela, Alfonso Secuya and Charibeth Cheng

15:35–16:00   *Coffee break 2*

16:00–16:30   **Feedback round from the audience - future directions of ECONLP**

16:30–18:00   **POSTER PRESENTATIONS**