



# Workshop on AI Powered Sentiment Analysis – NLP, Data Science and Others

**Date** 14 November 2019, Thursday  
**Time** 9:30am to 4:30pm  
**Venue** Auditorium on the ground floor, I<sup>3</sup> Building, 21 Heng Mui Keng Terrace, Singapore 119613

Time	Topic
9:30 – 10:00	Registration
10:00 – 10:10	Opening by Professor ZHU, Chengbo (Head of Math Department)
10:10 – 10:50	<b>Measuring the information content of financial news</b> Bin KE (National University of Singapore)
10:50 - 11:30	<b>Linking graphical methods and sentiment analysis for fake news detection</b> Min-Yen KAN (National University of Singapore)
11:30 – 13:00	Lunch break and Discussion
13:00 – 13:40	<b>News co-occurrence, attention spillover, and return predictability</b> Jun TU (Singapore Management University)
13:40 – 14:20	<b>Media sentiment and stock market prediction</b> Stefan LESSMANN (Humboldt-Universität zu Berlin)
14:20 – 14:50	Tea Break
14:50 – 15:30	<b>Topic sentiment asset pricing with DNN supervised learning</b> Hitoshi IWASAKI (National University of Singapore)
15:30 – 16:10	<b>RatingBot: a credit risk rating methodology based on text mining</b> Diana HRISTOVA (Deutsch Bank AG)

<b>19:00</b>	Dinner (By Invitation)
--------------	------------------------

Workshop Organizers:

Ying CHEN, *Department of Mathematics, Risk Management Institute, National University of Singapore*

Min DAI, *Centre for Quantitative Finance, Department of Mathematics, Risk Management Institute, National University of Singapore*

Yeneng SUN, *Risk Management Institute, National University of Singapore*

Workshop Coordinators:

Hitoshi IWASAKI, *Department of Statistics and Applied Probability, National University of Singapore*

Hao LEI, *Department of Statistics and Applied Probability, National University of Singapore*

The workshop is jointly organized by the Department of Mathematics, Risk Management Institute and Centre for Quantitative Finance at the National University of Singapore.

## MEASURING THE INFORMATION CONTENT OF FINANCIAL NEWS

---

Bin KE

*Professor of Accounting, Provost's Chair, and Director of Asia Accounting Research Centre, Business School, National University of Singapore*

**Abstract:** We propose a model to automatically measure the information content of news text, trained using news and corresponding cumulative abnormal returns of listed companies. Existing methods in the accounting and finance literature exploit sentiment signal features, which are limited by not considering factors such as events. We address this issue by leveraging deep neural models to extract rich semantic features from news text. In particular, a novel tree-structured LSTM is used to find target-specific representations of news text given syntax structures. Empirical results show that the neural models can outperform sentiment based models, demonstrating the effectiveness of recent NLP technology advances for computational finance.

## LINKING GRAPHICAL METHODS AND SENTIMENT ANALYSIS FOR FAKE NEWS DETECTION

---

Min-Yen KAN

*Associate Professor and Deputy Director (Research) in the NUS Institute of the Application of Learning Sciences and Education Technology (ALSET), School of Computing, National University of Singapore*

**Abstract:** While much fake news detection methods stress fact verification to see whether a source's statements can be validated against factual sources, another thread of work attempts to leverage the regularity of how news and attitudes spreads in our social networks. We build up towards this goal, first by examining the problem of fake news detection and stance detection using sentiment analysis features and state of the art XLNet embeddings on the main common datasets of CSI, Fake News Challenge and RumourEval 2019, and then examining a broader architecture involving a graphical representation where nodes are users, their comments on news articles, the news articles, and the websites / sources where those news articles originate from. Graph Convolution methods then can apply to such a graph enabling appropriate, multi-hop information figure into the computation of the status of each node. A useful output of these forms of analysis is that the nodes within the network (people, articles, and sources) can be ranked on scales. This is a joint work with NGUYEN Van Hoang, Toshiki TOMIHARA, Kazunari SUGIYAMA, Yohei SEKI, Preslav NAKOV.

## NEWS CO-OCCURRENCE, ATTENTION SPILLOVER AND RETURN PREDICTABILITY

---

Jun TU

*Associate Professor of Finance, Lee Kong Chian School of Business, Singapore Management University*

**Abstract:** We examine the effect of investor attention spillover on stock return predictability. Using a novel measure, the News Network Triggered Attention index (NNTA), we find that NNTA negatively predicts market returns with a monthly in(out)-of-sample R-square of 5.97% (5.80%). In the cross-section, a long-short portfolio based on news co-occurrence generates a significant monthly alpha of 68 basis points. The results are robust to the inclusion of alternative attention proxies, sentiment measures, other news- and information-based predictors, across recession and expansion periods. We further validate the attention spillover effect by showing that news co-mentioning leads to greater increases in Google and Bloomberg search volumes than unconditional news coverage. Our findings suggest that attention spillover in a news-based network can lead to significant stock market overvaluations, and especially when arbitrage is limited.

## MEDIA SENTIMENT AND STOCK MARKET PREDICTION

---

Stefan LESSMANN

*Chair Professor of Information Systems, School of Business and Economics, Humboldt-Universität zu Berlin*

**Abstract:** For the past few years, media sentiment has been used as a proxy for investor sentiment to test the effect of the latter on stock market returns. Corresponding studies often rely on dictionary-based methods to quantify sentiment. Given fundamental advancements in the field of natural language Processing in the recent past, advanced deep learning models promise a substantially more accurate extraction of sentiment indicators from text data. The paper considers BERT (Devlin et al. ,2018), a powerful text classification methodology based on transfer learning, and examines the degree to which BERT-based sentiment indices differ from conventional dictionary- and machine learning-based indices when used as a cue to model financial market developments.

## TOPIC SENTIMENT ASSET PRICING WITH DNN SUPERVISED LEARNING

---

Hitoshi IWASAKI

*Department of Statistics & Applied Probability, National University of Singapore*

**Abstract:** We develop an innovative deep neural network (DNN) supervised learning approach to extracting insightful topic sentiments from analyst reports at the sentence level and incorporating this qualitative knowledge in asset pricing and portfolio construction. The topic sentiment analysis is performed on 113,043 Japanese analyst reports and the topic sentiment asset pricing model delivers superior predictive power on stock returns with adjusted R squared increasing from 1.6% (benchmark model without sentiment) to 14.0% (in-sample) and 13.4% (out-of-sample). We find that topics reflecting the subjective opinions of analysts have greater impact than topics of objective facts and justification of the quantitative measures.

## RatingBot: A CREDIT RISK RATING METHODOLOGY BASED ON TEXT MINING

---

Diana HRISTOVA

*Risk Methodology, Deutsche Bank AG, Germany*

**Abstract:** Credit risk is at the core of banking business and its adequate measurement is crucial for financial institutions. Due to lack of historical default data and heterogeneity of customers, qualitative expert-based information is an important factor in measuring the creditworthiness of large companies. However, such information is often extracted manually, causing inefficiencies and possible subjectivity. The RatingBot is a text mining based credit rating approach, which efficiently and objectively models relevant qualitative information based on textual sources. In particular annual reports are analysed with text mining (including sentiment analysis) and machine learning classification techniques (e.g. SVM, NN) to predict the credit rating class of a company. The approach is evaluated on two datasets and also the application of further approaches (e.g. , Random Forest, embeddings) and sources of qualitative information (e.g. news) is discussed.