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

Date	14 November 2019, Thursday
Time	9:00am to 4:30pm
Venue	I³ Building, 21 Heng Mui Keng Terrace, Singapore 119613

9:00 – 9:30	Registration
9:30 – 9:40	Opening
9:40-10:20	**Informativeness of corporate filing sentiments: new evidence from a contextual deep learning approach** Allen HUANG (The Hong Kong University of Science and Technology)
10:20 – 11:00	**Measuring the Information Content of Financial News** Bin KE (National University of Singapore)
11:00 - 11:30	**Tea Break**
11:30 – 12:10	**TBA** Min-Yen KAN (National University of Singapore)
12:10 – 14:00	Lunch break and Discussion
14:00 – 14:40	**News Co-Occurrence, Attention Spillover, and Return Predictability** Jun TU (Singapore Management University)
14:40 – 15:20	**Media Sentiment and Stock Market Prediction** Stefan LESSMANN (Humboldt-Universität zu Berlin)

15:20 – 15:50	Tea Break
15:50 – 16:30	**Topic Sentiment Asset Pricing with DNN Supervised Learning** Hitoshi IWASAKI (National University of Singapore)
16:30 – 17:10	**RatingBot: a credit risk rating methodology based on text mining** Diana HRISTOVA (Deutsch Bank AG)
18:00	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.

INFORMATIVENESS OF CORPORATE FILING SENTIMENTS: NEW EVIDENCE FROM A CONTEXTUAL DEEP LEARNING APPROACH

Allen HUANG

Associate Professor at Department of Accounting and Faculty Associate at Institute for Emerging Market Studies, The Hong Kong University of Science and Technology

Abstract: We customize the state-of-the-art deep learning language representation model from Google, Bidirectional Encoder Representations from Transformers (BERT), with large amount of financial texts (5 billion word tokens) to more accurately measure sentiments in corporate filings. Unlike prior algorithms used in accounting and finance to measure sentiments, BERT can learn language context (e.g., polysemy) and semantics from sentence compositions and their sequences of textual documents. We show that financial-customized BERT can more accurately measure sentiments in financial texts than prior methods used in business research such as Loughran and McDonald (2011) word list and the naïve Bayes algorithm, and other deep learning natural language processing models that do not consider contextual meaning of words, such as word2vec. Last, sentiments measured finance-customized BERT have stronger associations with market reactions and future firm fundamentals compared to those measured with traditional methods.

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.

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 in natural language Processing in the recent past, advanced deep learning models promise a substantially more accurate extraction of sentiment indicators from in 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.