Text Analysis

Period: a.y. 2022/23 – II sem.

Class times:

Instructor:

Prof. **Qiaoni Shi**
Dept. of Marketing - Room 4-D1-14
qiaoni.shi@unibocconi.it

Course description

A growing number of papers applied textural analysis appeared in top Marketing, Management, Accounting journals in recent years.

To help enrich your research agenda, this course will provide you with an understanding of common and emerging methods of analyzing large collections of textual data. In addition to the survey of text analytics methods, this course will teach you how to apply these text analytics methods in Python.

In summary, the focus of this course is on concepts and methods of text analytics, their implementations in Python, and applications to empirical research. Ph.D. students of all disciplines may benefit from this course.

Learning Objectives

By the end of the seminar, you are expected to have developed the following skills:

- An understanding of how text analytics are used in Marketing, Management
- An understanding of traditional and emerging methods for analyzing textual information
- Familiarity with basic syntax in Python
- An understanding of how to apply text analytics methods in Python

Course Requirement

1. Final Project – Research Proposal

This final project asks you to come up with a promising research idea in which text analytics is applied to your area of interest. The purpose of the project is to help jumpstart your publication pipeline by prodding you to explore

opportunities in applying text analysis to your field. The goal is to come up with an idea that you can pursue to write a publishable manuscript, therefore, the idea should be novel in some ways (e.g., new research question, new hypothesis, or revisit old research question but apply text analysis on new data) and feasible.

By the end of the course, you have to submit a proposal that includes these sections:

1) Research question: 1-2 paragraphs formally describing your research question.
2) Motivation: describe why your research question is important
3) Literature review: discuss extant theory and prior related published works
4) Proposed research design: describe in detail how you will use text analysis to answer your research questions
5) Data: describe in detail what data you plan to use, and how you plan to obtain or collect these data. You should show that the required data for this research idea are not prohibit difficult to get.
6) Proposed analyses: describe the analyses that you plan to do
7) Expected findings and managerial implications
8) Original contribution: describe the original and innovative contribution of this paper to the existing body of knowledge

2. Final Project - Presentation

You will be asked to present the proposal of your research idea during the last session. You will have 12 minutes to present your research proposal, followed by a short Q&A session.

Topics

Lecture

1. Applications of text analytics in Marketing, Management
2. Natural Language Processing

   Regular Expression, Stop-words, Tokenization, Normalization, Stemming and Lemming, Part of Speech Tagging, Word vector representation: word2vec
3. Text Analysis

   Frequency analysis, Term weighting, Clustering, Classification, Probabilistic topic model (Latent Dirichlet Allocation)
4. Information extraction

   Name entity and Relation, Affective meaning: Sentiment, Emotion

Computer Lab

1. Python Syntax
2. Data Acquisition from the Web: website scraping
3. Text Analytics in Python

Python Software

Google Colaboratory will be used in this course for the Python programming.

Tentative Reading List

Text Analytics

Methods

- **Traditional/Statistical Approaches**

Tim Loughran, Bill Mcdonald (2016) Textual Analysis in Accounting and Finance: A Survey. Journal of Accounting Research 54(4): 1187-1230.

David Blei (2012) Probabilistic Topic Models. Communications of the ACM 55(4) 77-84

Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schutze. Introduction to Information Retrieval. Cambridge University Press. https://nlp.stanford.edu/IR-book/#anchor01

- **Deep Learning Approaches**

Ronan Collobert et al. (2011) Natural Language Processing (Almost) from Scratch. Journal of Machine Learning Research 12: 2493-2537

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. (2013) Efficient Estimation of Word Representations in Vector Space. Proceedings of Workshop at ICLR 2013. Available at: https://arxiv.org/abs/1301.3781

Socher Richard et al. (2013) Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. Conference on Empirical Methods in Natural Language Processing (EMNLP 2013). Available at: https://nlp.stanford.edu/~socherr/EMNLP2013_RNTN.pdf

Ozan. Irsoy, Claire Cardie (2014) Opinion Mining with Deep Recurrent Neural Networks. Conference on Empirical Methods in Natural Language Processing (EMNLP 2014). Available at: https://www.cs.cornell.edu/~oirsoy/files/emnlp14drnt.pdf

Dan Jurafsky and James H. Martin. Speech and Language Processing (3rd ed. draft). Available at: https://web.stanford.edu/~jurafsky/slp3/

- **Text Analytics in Python**

Natural Language Processing with Python. Available at: https://www.nltk.org/book/

Applications in Marketing

Nikolay Archak, Anindya Ghose, Panagiotis Ipeirotis (2011) Deriving the Pricing Power of Product Features by Mining Consumer Reviews. Management Science 57(8): 1485-1509.

Lee Thomas, Eric Bradlow (2011) Automated Marketing Research Using Online Customer Reviews. Journal of Marketing Research 48(5): 881-894.

Oded Netzer, Ronen Feldman, Jacob Goldenberg, Moshe Fresko (2012) Mine Your Own Business: Market-Structure Surveillance Through Text Mining.

Marketing Science 31(3):521-543.

Stephan Ludwig et al. (2013) More Than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates. Journal of Marketing 77(1): 87-103.

Hyoryung Nam, Yogesh Joshi, P.K. Kannan (2017). Harvesting Brand Information from Social Tags. Journal of Marketing 81(4): 88-108

Lee, Dokyun, Kartik Hosanagar, and Harikesh S. Nair. (2018) Advertising content and consumer engagement on social media: Evidence from Facebook. Management Science 64.11: 5105-5131.

Gabel, S., Guhl, D., & Klapper, D. (2019). P2V-MAP: Mapping market structures for large retail assortments. Journal of Marketing Research, 56(4), 557-580.

Fanglin Chen, Xiao Liu, Davide Proserpio, and Isamar Troncoso Cortez (2021). "Product2Vec: Understanding Product-level Competition Using Representation Learning," Working Paper. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3519358

Applications in Management

Wei Guo, Tieying Yu, Javier Gimeno (2017) Language and Competition: Communication Vagueness, Interpretation Difficulties, and Market Entry. Academy of Management Journal 60(6): 2073-2098

Manjit Yadav, Jaideep Prabhu, Rajesh Chandy (2007) Managing the Future: CEO Attention and Innovation Outcomes. Journal of Marketing 71(4) 84-101.

Gerard Hoberg and Gordon Phillips (2010) Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis. Review of Financial Studies 23(10): 3773-3811.

Angelo Fanelli, Vilmos F. Misangyi, Henry L. Tosi (2009) In Charisma We Trust: The Effects of CEO Charismatic Visions on Securities Analysts. Organization Science 20(6): 941-1076.

Donal Crilly, Morten Hansen, Maurizio Zollo (2016) The Grammar of Decoupling: A Cognitive-Linguistic Perspective on Firms' Sustainability Claims and Stakeholders' Interpretation. Academy of Management Journal 59(2): 705-729.

Tentative schedule

Session	Lecture	Lab/Discussion
1	Introduction	Intro to Python and Google Colaboratory
2	Text analysis basics	Python basics
3	Dictionary-based Text analysis	Data acquisition, Regular expression
4	Text representation, e.g., Word2Vec	Literature Discussion
5	Latent Dirichlet Allocation (LDA)	LDA, Sentiment analysis
6	Presentation	