1、情感因子/舆情因子（NLP）

**\*代码参考**（研究对象为中国，仅做代码参考）：<https://github.com/Coldog2333/Financial-NLP>

文章1

Tobias Daudert, Paul Buitelaar, Sapna Negi, Leveraging News Sentiment to Improve Microblog Sentiment Classification in the Financial Domain

<https://www.aclweb.org/anthology/W18-3107/>

简介：通过机器学习模型对microblog中显示有关金融情况的情绪信息进行处理。

数据：

（1）the microblog dataset (M) from the Semeval 2017 Task 5 - subtask 1 (Cortis et al., 2017). This dataset contains 2,488 microblogs retrieved from Twitter3 collected between March 11th and 18th **2016** as well as StockTwits.

<https://www.acl-bg.org/proceedings/2017/RANLPStud%202017/pdf/RANLPStud002.pdf>

（2）the newly created microblogs-related news dataset (MRN) consists of 106 news, specifically, it contains the news’ titles, urls, time and date, a sentiment score, and, if available, a description for each news. The news data was gathered from multiple sources such as wsj.com or bloomberg.com.

**方法**：

The ML algorithm chosen for this task is a Support Vector Machine (SVM).

文章2

Johan Bollen, Huina Mao, Xiaojun Zeng, Twitter mood predicts the stock market

<https://www.sciencedirect.com/science/article/pii/S187775031100007X?via%3Dihub>

简介：

从大型Twitter提要的文本内容中测量了7个不同方面的公众情绪状态。

公众情绪状态的每日变化与道琼斯工业平均指数收盘价的每日变化在统计上具有显着相关性。

调查从大型Twitter提要中获得的集体情绪状态的度量值是否与道琼斯工业平均指数（DJIA）的值随时间相关。

通过两种情绪跟踪工具来分析每日Twitter提要的文本内容，即用来衡量积极情绪与消极情绪的OpinionFinder和可以从6个维度（平静，警报，肯定，重要）衡量情绪的Google情绪状态简介（GPOMS），善良和快乐）。

通过比较民意测验公众对2008年总统选举和感恩节的反应的能力，对所得的情绪时间序列进行交叉验证。然后，使用格兰杰因果分析和自组织模糊神经网络来调查公众情绪的假设。状态由OpinionFinder和GPOMS情绪时间序列测得，可以预测DJIA收盘价的变化。

数据：

（1）这些推文已于**2008**年2月28日至12月19日进行了记录（大约270万 用户发布了9,853,498条推文）。对于每条推文，这些记录都提供一条推文标识符，提交的日期-时间（GMT + 0），其提交类型以及该Tweet的文本内容，该文本内容设计为限制为140个字符。删除停用词和标点符号后，我们将在同一日期提交的所有推文归为一组。

（2）从Yahoo!中提取每日DJIA收盘价的时间序列。

从雅虎获取数据方法（代码）：<https://github.com/lukaszbanasiak/yahoo-finance>

文章3

Merve Alanyali, Helen Susannah Moat & Tobias Preis, Quantifying the Relationship Between Financial News and the Stock Market

<https://www.nature.com/articles/srep03578>

简介：分析2007-2012年之间《金融时报》上有关金融市场的新闻，发现在新闻发布前一天和新闻发布当天，公司在《金融时报》上的每日提及次数与公司股票的每日交易量之间存在正相关关系。

结论：金融市场的动向和金融新闻的动向是内在联系的。

方法：

（1）Data retrieval and preprocessing.

（2）Company names used for news corpus analysis.

（3）Statistical analyses of basic characteristics of the data.

文章4

Finding Value in Earnings Transcripts Data with AlphaSense

<https://extractalpha.com/wp-content/uploads/2015/06/Finding-Value-in-Earnings-Transcripts-Data-with-AlphaSense1.pdf>

简介：使用 AlphaSense（商业化工具） 进行情感分析。

（不是很重要的一篇）

2、Technical Indicators

文章1

Jeremy Goh, Fuwei Jiang, Jun Tu, Guofu Zhou, Forecasting Government Bond Risk Premia Using Technical Indicators

<https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1914227>

简介：研究各项技术指标对于经济发展的影响。

选用63项技术指标。

指标构建：

（1）第一类技术指标是基于滞后远期价差的移动平均值构建的。以趋势跟踪技术分析的标准方法，根据过去远期息差的移动平均值构建了前48个技术指标。

（2）第二组技术指标将建立在“平衡”量的基础上。根据股票市场交易量构建了15个技术指标。

结论：技术指标在经济和统计上都具有显著的预测能力，无论是在样本内还是样本外，以及短期和长期政府债券。技术指标比经济变量更有用，将技术指标中的信息与经济变量相结合的预测模型大大优于仅使用经济变量的模型的预测。

数据

（1）We obtain the **short-term zero coupon U.S. Treasury bond prices** with maturities from onethrough five-years from Fama-Bliss dataset available at the Center for Research in Securities Prices(CRSP) spanning the period 1964:01−2011:12. The **long-term U.S. Treasury bond** data with maturities from seventeen- to twenty-years are from the Federal Reserve’s website, which provides updated data from Gurkaynak, Sack, and Wright (2007) beginning in 1981:07.

<http://www.federalreserve.gov/pubs/feds/2006>

（2）The **macroeconomic** fundamental data are obtained from Sydney C. Ludvigson’s web page and used in Ludvigson and Ng (2009, 2011).

<http://www.econ.nyu.edu/user/ludvigsons/Data&ReplicationFiles.zip>

（3）In addition, we use monthly S&P 500 index and stock market trading volume data from Google Finance.

Methodology

（1）Technical indicator construction

The two types of technical indicators that we consider (bond price and trading volume-based) conveniently capture the trend-following idea that is at the heart of technical analysis. These are representative of the technical indicators that are often analyzed in the academic literature (e.g., Brock, Lakonishok, and LeBaron, 1992; Sullivan, Timmermann, and White, 1999).

<https://onlinelibrary.wiley.com/doi/full/10.1111/j.1540-6261.1992.tb04681.x>

<https://onlinelibrary.wiley.com/doi/full/10.1111/0022-1082.00163>

To investigate the latter question, we include Cochrane and Piazzesi (2005) forward rate factor CPt and Ludvigson and Ng (2009) macroeconomic variable factor LNt as control variables.

<https://www.aeaweb.org/articles?id=10.1257/0002828053828581>

<https://academic.oup.com/rfs/article-abstract/22/12/5027/1577464>