

基于网络大数据的社会经济预测综述*

苏志 张骥

(首都经济贸易大学 北京 100070)

摘要 整理采用网络大数据进行社会经济预测的分析报告和学术研究,并根据数据信息类型梳理研究主要脉络。其中,网络用户搜索数据研究根据搜索关键词义和搜索频率变化预测经济活动或市场发展趋势;舆情研究从网络信息交流平台提取公众观点倾向和情绪波动信息,研究投资行为以及研判金融市场形势。还总结现有研究存在的问题以及展望未来研究发展方向。

关键词 网络大数据 搜索引擎 微博舆情 经济预测 市场预测 文献综述

中图分类号 G358 F062.5

文献标识码 A

文章编号 1002-1965(2015)07-0018-04

DOI 10.3969/j.issn.1002-1965.2015.07.005

Researches on Socio-economic Forecasting Based on Internet Big Data

Su Zhi Zhang Qi

(Capital University of Economics and Business, Beijing 100070)

Abstract According to the data content, this paper summarizes and organizes some reports and academic researches about using internet big data for socio-economic forecasting. It finds that search engine user data are used to predict trends of economic activities or markets, mainly based on search keyword meaning and search frequency; and sentiment studies get public opinion and emotion information from on-line social media platforms, which is then applied in studying investment behaviors and financial markets. This paper also points out some problems in the current research and presents some perspectives for future studies.

Key words Internet big data search engine microblog sentiment economic forecasting market forecasting literature review

0 引言

互联网的快速普及和广泛应用使人类社会产生和获得的数据与信息出现爆炸式增长。2011年,全球共有21亿互联网用户和5.55亿个网站,新增和复制的数据总量为1.8ZB。预计到2020年,全球的数据总量将达到35ZB,每年产生的数据远远超过人类有史以来所有印刷材料的数据总量(200PB)。网络大数据正在改变人们生产和生活方式,并影响企业经营和政府管理方式,已经有不少研究和项目将网络大数据应用于社会经济活动预测。

基于大数据的社会经济预测研究通过统计性的搜索、比较、聚类和分类等方法分析归纳海量数据,实现“从数据直接到价值知识”。大数据研究的这种特性很适合社会经济预测:人类经济社会是一个开放复杂

的系统,系统之间有着千丝万缕的有机联系,可以通过大数据分析现象关联性或局部系统联系机制,在一定程度上对社会经济现象的发展进行合理推测。因此,如何有效的收集、提取、分析、利用网络大数据中蕴含的丰富社会经济信息成为政府、企业和科研机构的热点研究课题。

1 基于网络大数据的社会经济研究综合性报告

1.1 网络大数据研究具有商业经济价值 最早意识到网络大数据的经济和社会价值并开始研究和利用大数据主要是一些企业、政府和研究机构。麦肯锡2011年发布一份报告估计大数据应用在医疗领域和公共管理部门的潜在年产值以及个人位置数据的市场价值超过千亿美元^[1]。在企业界,众多高科技公司在巨大经济利益的驱动下大力发展大数据处理技术,因

收稿日期:2015-03-05

修回日期:2015-04-10

基金项目:北京市自然科学基金项目“北京房价预期微博舆情研究”(编号:9144025)。

作者简介:苏志(1979-),女,博士,讲师,研究方向:应用经济学、网络舆情等;张骥(1990-),男,硕士研究生,研究方向:经济学、网络舆情等。

为已经实施或部署大数据试验项目的企业组织比依靠传统分析方法的企业组织获得高约15%的优势^[2]。

1.2 网络大数据研究具有社会经济价值 各国政府、学术机构、以及一些国际机构也积极推动网络大数据研究,致力于将大数据的价值转化为社会和经济发展的动力。MIT学者创建的Billion Prices Project (BPP)通过采集网络零售商品价格建立零售价格通货膨胀指数,为宏观价格和国际价格比较研究提供新数据来源。2012年的世界经济论坛的报告重点探讨在发展中国家使用移动通信设备所产生的大数据如何帮助政府和机构了解低收入人群的情况,并相应的提供服务和进行危机预警^[3]。联合国推出Global Pulse项目根据社交网络等数据研究社会舆情,对贫困国家的失业率增加、物价上涨、或是疾病爆发等危机进行预测,通过数字化预警信号调整人道主义援助项目政策,从而更有效的帮助更多地区摆脱贫困^[4]。

2 基于网络搜索数据的社会经济预测研究

谷歌和百度等搜索引擎用户的行为日志数据记录人们搜索内容、搜索频率以及位置等信息。在某种程度上,人们在使用搜索引擎时输入的搜索关键词义反映了人们在现实世界里的需求、兴趣、和关注,而搜索频率的高低则显示需求、兴趣、和关注的强弱。于是,有研究利用此类数据把网络搜索活动与现实世界的社会经济活动联系起来。

2.1 网络搜索数据预测产品销售情况 在美国,根据雅虎网页查询日志记录的搜索量可预测电影票房收入和游戏销售变化^[5]。谷歌趋势把关键词的搜索频率通过特定算法转化为关键词标准搜索指数,并根据搜索关键词词义进行分类。其中,“汽车和零部件”类别指数可预测实际汽车销售,“房地产中介”类别指数可预测同期房屋销售^[6]。来自以色列、智利等国家的研究人员也发现谷歌搜索指数可用于汽车、房产、家具等商品的销售预测^[7,8]。

2.2 网络搜索数据预测宏观经济趋势 通常情况下,政府统计依赖于传统的调查方式来建立各种宏观经济指数,这些宏观数据往往滞后于政府宏观调控政策制定或企业经营决策需求。于是,对于即时信息的迫切需求使网络搜索高频数据成为预测宏观经济趋势变化的重要数据来源。谷歌趋势中有关“工作”和“福利和失业”类别搜索指数可以帮助预测美国首次申领失业救济人数^[9]。研究表明谷歌趋势中与就业相关的词条查询或招聘查询指数可预测德国、以色列、土耳其、意大利、美国的失业率趋势。而且,用谷歌求职搜索指数预测美国的失业率的研究表明该创新模型优于基于专业预测人士调查的全国失业率预测模

型^[8,10-13]。

2.3 网络搜索数据预测宏观经济各个层面的趋势变化 谷歌趋势中金融相关关键词搜索量可代表投资者搜索股市信息行为密集度。研究发现金融相关关键词的谷歌搜索量在道琼斯指数涨之前下跌的规律,据此规律建立炒股策略比传统炒股策略收益高^[14]。此外,从谷歌趋势中提取数据构建的消费指数不仅可以预测美国总体个人消费水平变化,其预测精度比基于调查数据构建的消费者信心指数预测精度要高^[15]。同时,“房地产”等关键词的谷歌搜索指数与常用的Case-Shiller房屋价格指数有显著关联^[16]。

2.4 网络搜索数据预测中国社会经济活动 从谷歌趋势抽取“汽车评级”“汽车4S店”等关键词搜索数据合成综合指数可以预测不同价位的汽车销量^[17]。用个股公开发行人前谷歌搜索量衡量的投资者关注度对于个股市场热销程度、首日超额收益和长期表现有一定的解释力^[18]。也有研究用与谷歌趋势类似的百度指数构建投资者关注度,发现其对股票市场有正向价格影响^[19]。另外,用百度指数和谷歌趋势数据构建中国“失业”相关关键词搜索指数与一系列就业相关的宏观指数显著相关,并可预测这些宏观指数变化^[20]。

2.5 网络搜索数据预测社会经济研究中存在的问题

首先,还没有有效的理论指导如何从巨量搜索词条中挑选与所研究的社会经济现象相关词条,也没有相应的技术支持精确快速提取特定目标词汇组合。现有的研究一般是主观性的选择关键词开展研究,其研究结果的客观性和稳健性有待考察。

其次,拥有网民搜索行为大数据的公司一般不对外公布数据,或是只公开少量数据,限制了搜索数据应用于社会经济预测的范围。同时,不断调整和改进的搜索趋势指数算法也会影响数据的连贯性和可用性。

最后,用户反馈性的搜索行为会影响利用搜索趋势指数进行社会经济预测结果。例如,有研究发现“谷歌流感趋势”在2011年到2013年期间有100周的预测结果比实际流感发生率偏高,原因是各类媒体报道“谷歌流感趋势”后,人们更多的搜索相关新闻,从而影响该服务的统计数据结果^[21]。如何在这种动态反馈的社会经济系统中挖掘有效信息,合理的处理分析数据,并进行较准确的趋势预测是一项亟待展开的课题。

3 基于舆情信息数据的社会经济预测研究

在互联网上,论坛、博客、微博是人们获取、发布、交流信息的重要渠道,具有全民公开、观点和情感集中、以及即时更新的特性,通过计算机技术对这些平台

上的信息进行观点挖掘和情感分析,可以探究公众心理和衡量群体预期,开展舆情研究进行社会经预测。

3.1 论坛舆情的股市预测研究 股票专题论坛集中了投资者感兴趣的话题和讨论,论坛里有关股票收益的讨论比专业分析师的预测更能体现收益预期,根据这种非正式信息进行操盘的收益也更高^[22]。同时,股票论坛里异常的发帖活动和观点变化与股票异常成交量和回报率有显著关系^[23]。用投资者网络讨论信息构建投资者关注度、乐观情绪与意见分歧指标的研究也显示投资者情绪与IPO抑价之间有显著关系^[24]。从股票论坛中提取中小投资者观点,构建的投资者情绪指标与传统指标存在一定关联性^[25]。

3.2 微博舆情预测投资市场研究 近些年,以推特和微博为代表的网络社交平台上产生用户生成数据具有空前的规模性和群体性,其中的舆情与社会经济活动也紧密关联。例如,通过挖掘微博文本信息构建的舆情指数可有效预测美国股票价格指数变化^[26]。研究还发现包含观点和情感的微博舆情模型比非舆情微博模型能更准确的预测股票价格变化,熊市预测比牛市预测更精确^[27]。在中国,“牛市”,“利空”等股市相关词汇在微博里出现的频率可以短期预测上证指数变化,而负面词汇有更强的预测能力^[28]。一个时期内微博中正面与负面词汇所建立的公众情绪指数不仅与消费者信心指数有显著相关性,还能提高其预测精度^[29]。

3.3 应用舆情信息于社会经济预测应关注的问题 论坛和微博等网络公开信息平台已经成为互联网用户观点挖掘和网络情感分析的重要数据源之一。当前中国已经出现大量网络舆情预测与预警研究,然而这些研究侧重于社会民主政治、治安事件、危机处理等社会治理应用,较少关注民生和经济研究应用,今后的研究应更多关注后者。

同时,将情感分析应用于社会经济预测应注意以下一些问题:

首先,社会经济主题的微博舆论信息量大且结构复杂;文本词义表达方式多样;微博里的情绪表达形式多变且更新较快。如何从微博中挖掘出与研究主题密切相关的有效舆情信息,较为准确且全面的量化公众预期和情感倾向,需要多学科共同开发研究。

其次,即使网络信息的收集具有即时性,预测是否应该且能够达到即时性则需要根据预测目标确定。由于情感因素对社会活动影响程度和机制不同,情感分析结果作为辅助依据和作为主要依据的预测方法应该有所区别。

最后,应探索建立社会经济主题舆情收集和研判机制,在众多舆情信息中进行甄别和筛选,对社会经济

生活中突发重大事件或政府实施政策从内容、范围、程度等方面进行及时分析,给决策者提出针对性建议。

4 总结与展望

通过梳理国内外一些大数据经济应用于社会经济预测研究,本文总结几点研究现状及展望未来研究方向。

第一,网络大数据提供了预测社会经济的新工具。网络用户使用终端记录用户网络行为信息,同时各类网络平台还包含用户主动发布的互动信息,互联网技术的发展使得获得即时海量数据的成本大大降低,数据的可获得性和丰富性拓展了社会经济形势研究的前沿。

第二,网络大数据的背后是各种人群的社会经济活动、心理活动、情感取向信息,社会经济情报工作者迫切需与计算机学等其他领域的学者密切合作,从心理学、经济学、信息科学等不同学科领域共同探讨网络数据的产生和传播的基本规律和应用方法。

第三,未来的研究应特别重视基于大数据的网络行为机理识别;网络用户在线活动大数据生成及社会化网络中的信息传播机理;用户互动创造信息对经济个体和总体趋势影响机理;以及网络化的经济态势研判与预警决策研究。

第四,如何将网络数据与传统方法收集的信息结合起来,分析社会经济政策与社会经济现象的相互影响关系,以及如何在预测的精准性和实时性之间寻求平衡是大数据环境下社会经济形势研究的新课题。

最后,基于大数据的社会经济预测研究也面临诸多问题。大量随机产生的网络数据具有多源异构、交互性、时效性、社会性、突发性和高噪声等特点,存在信息冗余以及价值密度低的缺陷。现有的网络大数据研究大多立足于信息科学,侧重于大数据的获取、存储、处理、挖掘和信息安全等方面研究,而采用大数据分析社会经济形势刚刚起步,未来仍有很大的发展空间。

参考文献

- [1] Manyika J,Chui M,Brown B,et al. Big data:The Next Frontier for Innovation, Competition, and Productivity [R]. McKinsey Global Institute,2011.
- [2] Schroeck M, Shockley R, Smart J, et al. Analytics: The Real-World Use of Big Data [R]. IBM's Institute for Business Value (IBV) and the University of Oxford,2012.
- [3] Wef Big Data,Big Impact:New Possibilities for International Development [R]. Geneva, Switzerland: World Economic Forum, 2012.
- [4] Letouzé E. Big Data for Development:Challenges & Opportunities [R]. New York,USA:UN Global Pulse,2012.
- [5] Goel S,Hofman J M,Lahaie S,et al. What Can Search Predict?

- [R]. New York: Yahoo! Research, 2010.
- [6] Choi H, Varian H. Predicting the Present with Google Trends [J]. Economic Record, 2012, 88 (Supplement s1): 2-9.
- [7] Carrière - Swallow Y, Labbé F. Nowcasting with Google Trends in an Emerging Market [J]. Journal of Forecasting, 2011, 32 (4): 289-298.
- [8] Suhoy T. Query Indices and a 2008 Downturn: Israeli Data [R]. Bank of Israel Discussion, 2009. 06.
- [9] Choi H, Varian H. Predicting Initial Claims for Unemployment Benefits [R]. Google Inc., 2009.
- [10] Askatas N, Zimmermann K F. Google Econometrics and Unemployment Forecasting [J]. Applied Economics Quarterly, 2009, 55(2): 107-120.
- [11] Chadwick M, Sengul G. Nowcasting Unemployment Rate in Turkey: Let's ask Google [R]. 1891-599X. Central Bank of the Republic of Turkey, 2010.
- [12] D'amuri F, Marcucci J. "Google it!" Forecasting the US Unemployment Rate with a Google Job Search Index [R]. the Economic and Social Research Council Working Paper 421, 2009.
- [13] D'amuri F. Predicting Unemployment in Short Samples with Internet Job Search Query Data [R]. Munich Personal RePEc Archive, 2009.
- [14] Preis T, Moat H S, Stanley H E. Quantifying Trading Behavior in Financial Markets Using Google Trends [J]. Scientific reports, 2013, 3 (1684): 1-6.
- [15] Vosen S, Schmidt T. Forecasting Private Consumption: Survey-Based Indicators vs. Google Trends [J]. Journal of Forecasting, 2011, 30 (6): 565-578.
- [16] Wu L, Brynjolfsson E. The Future of Prediction: How Google Searches Foreshadow Housing Prices and Quantities [C]. Proceedings of the the International Conference on Information System 2009. ICIS; 2009.
- [17] 袁庆玉, 彭 赓, 刘 颖, 等. 基于网络关键词搜索数据的汽车销量预测研究 [J]. 管理学家: 学术版, 2011(1): 12-24.
- [18] 宋双杰, 曹 晖, 杨 坤. 投资者关注与 IPO 异象——来自网络搜索量的经验证据 [J]. 经济研究, 2011(增1): 145-155.
- [19] 俞庆进, 张 兵. 投资者有限关注与股票收益——以百度指数作为关注度的一项实证研究 [J]. 金融研究, 2012(8): 152-165.
- [20] Su Z. Chinese Online Unemployment-Related Searches and Macroeconomic Indicators [J]. Frontier Economics of China, 2014, 9 (2): 347-376.
- [21] Lazer D, Kennedy R, King G, et al. Big data. The Parable of Google Flu: Traps in Big Data Analysis [J]. Science, 2014, 343 (6176): 1203-1205.
- [22] Bagnoli M, Beneish M D, Watts S G. Whisper Forecasts of Quarterly Earnings per Share [J]. Journal of Accounting and Economics, 1999, 28 (1): 27-50.
- [23] Tumarkin R, Whitelaw R F. News or Noise? Internet Postings and Stock Prices [J]. Financial Analysts Journal, 2001: 41-51.
- [24] 林振兴. 网络讨论、投资者情绪与 IPO 抑价 [J]. 山西财经大学学报, 2011(2): 23-29.
- [25] Das S R, Chen M Y. Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web [J]. Management Science, 2007, 53 (9): 1375-1388.
- [26] Bollen J, Mao H, Zeng X. Twitter Mood Predicts the Stock Market [J]. Journal of Computational Science, 2011, 2(1): 1-8.
- [27] Oh C, Sheng O R L. Investigating Predictive Power of Stock Micro Blog Sentiment in Forecasting Future Stock Price Directional Movement [C]. Proceedings of the Thirty Second International Conference on Information Systems, 2011.
- [28] Zhou S, Shi X, Sun Y, et al. Stock Market Prediction Using Heat of Related Keywords on Micro Blog [J]. Journal of Software Engineering and Applications, 2013(6): 37-41.
- [29] O'connor B, Balasubramanyam R, Routledge B R, et al. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series [C]. Proceedings of the the International AAAI Conference on Weblogs and Social Media, 2010.

(责编: 贺小利)