大数据 产业发展

价格、需求、销售情况监测预测

刘晶, 和述群, 朱清香,等. 基于深度学习的线上农产品销量预测模型研究[J]. 计算机应用研究, 2017, 34(8):2291-2293.针对线上农产品销售存在的信息不对称问题,提出一种结合深度学习算法优势和涉农电商销售数据特点的皇冠模型(ICM)。首先建立因素评价指标,将销量分为四个类别;其次采用两层自编码网络提取样本特征,并生成新的特征向量;然后利用带标签样本集训练分类器并对无标签训练样本分类;最后利用BP微调整个网络参数得到使损失函数值达到最小的最优参数,实现线上农产品的销量分类预测。经仿真分析,验证了ICM的分类准确率高达88%,明显高于其他未将数据进行特征学习的浅层分类器,证明了ICM具有较好的增量自学习能力和层次认知能力。

周成. 基于Hadoop的卷烟需求预测系统设计与实现[D]. 南昌大学, 2016.十八届三中全会指出“全面深化改革”战略方针,强调资源配置在市场行为中的重要作用和主导地位。以此为背景,2014年3月国家烟草专卖局提出卷烟营销市场化取向改革,全国卷烟营销市场化取向改革由此开始。2014年改革工作在京、津、冀三地试点成功,总结试点经验,2015年改革工作在众多省市推广,改革试点工作的第一项任务就是探索需求预测方法并落实预期预测的指导作用。1.明确需求预测周期。因为卷烟销量具有明显的季节性和周期性,所以需求预测的周期一般为年、季度和月度。2.完善需求预测模型。对于现有需求预测模型存在的不足进行完善。需求预测一般是地市级单位和县级单位,综合考虑各方面因素确定和系统内部的数据,选择和优化预测模型,从而提高预测准确度。3.发挥需求预测作用。对于预测出来的结果,需要用到实际生产行为中去,指导生产和销售,达到优化资源的目的。本文探索基于大数据的卷烟销量预测系统的设计与实现。年度销量预测方面:改进了原有的一元线性规划算法,用包含经济、社会、人口三大因素,GDP、社会消费品总额、城市人均收入、农村人均收入、市人口总数五项指标做线性多元线性规划。月度预测方面提出了基于灰度模型和综合时间序列模型的混合预测模型。预测准确度比用灰度模型或者综合时间序列模型明显提高。最后用用JAVA语言实现了这个混合预测模型。最后,本文做了一项工作就是把基于Hadoop的卷烟销售预测系统与现有的正在运行的信息化系统做了集成,卷烟销售预测系统很好的与生产系统无缝对接,同时保证不影响现有的运行系统的业务流程。

章旭. 基于时间序列分析的汽车销量预测研究[D]. 合肥工业大学, 2017.中国已经成为全球最大的汽车生产和汽车消费市场。汽车工业的发展带动了诸多周边产业的发展,对于拉动内需和促进出口增长都具有很大的积极意义。因此准确地预测汽车销量有助于政策制定者从宏观上整体掌控市场发育与成长态势,有助于汽车制造厂商从微观上研究市场行情以制定营销策略。现有汽车销量预测研究大多面向整体汽车市场而非单一汽车品牌,而且缺乏对网络大数据的深入挖掘,忽略了大数据对于汽车销量的预测能力。针对上述问题,提出一种基于网络大数据和传统统计学时间序列分析的考虑品牌情感的汽车销量预测BOAR模型,对于单一品牌,在考虑该品牌不同时间窗历史销量基础上,结合用户在线评论中挖掘的用户情感值进行汽车销量预测。对多个汽车品牌的实验表明,综合历史同期销量、前期销量和品牌情感的BOAR模型平均预测误差为5.93%,比自回归模型降低8.59个百分点,BOAR模型可以准确预测单一汽车品牌的销量,具有更好的稳定性。与此同时,我们意识到BOAR模型及传统的时间序列分析模型都存在以下问题:(1)汽车销量影响因素众多,不同品牌的汽车其销量的影响因素本身可能也是不一致的,对于不同的汽车品牌采用一致的影响因素作为预测的解释变量是不合理的(2)需要事先假定影响因素和销量之间呈线性关系,对于高度非线性的关系难以准确用数学模型拟合。因此进一步提出一种通用的销量预测MISF模型,基于MARS变量选择过程和BP神经网络相结合的方式预测单一汽车品牌的月度销量。实验结果表明MISF模型的预测误差平均为4.04%,比BOAR模型进一步降低了1.49个百分点。这也验证了变量选择和神经网络在汽车销量预测研究中具有重要作用。本文提出的BOAR模型和MISF模型准确预测了单一汽车品牌粒度的月度销量,研究结果不仅对于汽车领域的销量预测研究具有一定的借鉴意义,同时可以为汽车制造厂商生产规划和控制提供更有效的决策支持。

李敏波, 王海鹏, 陈松奎,等. 工业大数据分析技术与轮胎销售数据预测[J]. 计算机工程与应用, 2017, 53(11):100-109.工业大数据是在工业领域信息化应用中所产生的海量数据,作为决策问题服务的大数据集、大数据技术和大数据应用的总称.首先分析工业大数据4V特性与工业数据的特有特征,以及工业大数据来源;从多源异构工业数据集成与数据融合方法、工业大数据计算架构、大数据带来的信息安全等三方面论述工业大数据面临的挑战与潜在价值.探讨了工业大数据分析与挖掘方法,提出了工业大数据平台的计算架构与大数据处理平台,构建轮胎企业大数据资源中心、大数据分析与决策应用系统.从销售数据分析和宏观数据趋势两个层面进行轮胎销售大数据分析与预测.采用多个不同领域的销售数据源来解决销售预测历史数据特征空间稀疏的问题,使用LASSO(The Least Absolute Shrinkage and Selectionator Operator)方法的多任务学习方法来解决高维样本空间的缺点,实验数据验证能够提升轮胎销售预测的准确率

崔东佳. 大数据时代背景下的品牌汽车销量预测的实证研究——以网络搜索数据为例[D]. 河南大学, 2014.从宏观上讲，汽车工业是我国国民经济的支柱产业，准确预测汽车销量不仅对汽车制造商和经销商在制定战略和目标有着巨大的作用和意义，从长远角度来说，对我国汽车市场的发育和成长态势及国民经济的快速发展也起到了一定的影响作用。

　　改革开放以来，随着我国国民经济的快速发展，我国居民的消费需求也在逐渐的发生改变，“汽车消费”正在替代“住房消费”成为现阶段我国部分居民的主要消费需求，消费者在购买汽车过程中，首先会有需求准备期，在需求准备的这段时间内，消费者会根据自身的喜好、汽车的价格和性能、以及自身的经济状况来获取潜在购买相关汽车的信息。

　　随着网络技术的快速发展和互联网的普及，颠覆了消费者获取信息的传统方式，使得消费者更加青睐与从互联网上搜索相关数据和信息，互联网成为用户获取资料和信息最快捷的途径之一，根据中国互联网信息中心的统计数据显示，截止到2013年底，80％左右的网民使用搜索引擎获取相关信息。由于网络搜索数据记录了数以亿计的搜索关注与需求，反映了市场主体的行为趋势与规律，为研究社会经济行为提供了必要数据基础。论文的研究建立在大数据时代背景的前提下，通过消费者在互联网搜索相关信息而留下来的记录，论文从我国汽车的低、中、高端市场分别选取了奇瑞、大众、宝马三个品牌的汽车作为研究对象，以消费者购买汽车的行为过程出发点，建立相关理论框架模型，揭示了网络搜索数据与汽车销量之间的相关关系，分别通过百度搜索引擎搜索与奇瑞、大众及宝马三个品牌汽车相关的关键词，采用综合赋权和错位逐步合成方法对搜索到的关键词进行合成，得出相应的网络搜索指数，然后利用网络搜索指数与相应的品牌汽车历史销量数据建立回归预测模型，进而进行协整分析和Granger因果检验。分析与检验结果表明，相比传统的汽车销量预测方法，该方法具有很高的预测精度（处于低端市场的奇瑞汽车拟合度相对较差，拟合度为70.0％，平均绝对误差值为8.51％，处于中端市场的大众汽车的拟合效果远高于奇瑞汽车的拟合效果，拟合度为95.2％，预测的平均绝对误差值为4.85％，而处于高端市场的宝马汽车的拟合度在三个品牌汽车中为最优，拟合度为97.7％，预测效果也最佳，平均绝对误差值仅为2.78％。），预测效果也随着品牌汽车档次的提高而提高，同时该研究方法可以对汽车销量实施监控，且比传统统计部门发布的数据要提前一个月左右。

虽然有关网络搜索数据的社会性与经济性的相关性研究还处于初级阶段，但随着信息技术的不断发展，及网络搜索数据的不断丰富与全面，该研究方法在理论研究方面将会形成一套完整的理论体系，同时也会有更加广阔的实际应用领域，例如:房地产行业、旅游行业等。

杨树新, 董纪昌, 李秀婷. 基于网络关键词搜索的房地产价格影响因素研究[J]. 新疆财经大学学报, 2013(3):5-12. 影响房地产价格的因素众多,而在当前房地产市场“刚性需求”和“投资需求”相互交织、难以区分的情况下,人们的预期就成为影响房地产价格的一个关键因素,这种预期在互联网已相当普及的今天可以通过网络搜索引擎反映出来.本文尝试构建网络搜索数据与房地产价格指数之间的关系框架,通过皮尔逊相关系数法,选取了与房屋销售价格指数相关系数在0.55以上的9个关键词,即房价、出租信息、二手房产网、二手房信息、开盘、户型装修、房产网、房源和装修材料；并通过实证检验,得出房屋销售价格指数与提前5个月搜索指数的相关性最大,即对房价感兴趣的购房者通常都会提前5个月左右搜索相关信息

王希晶. 基于网络搜索的中国区域房价预测模型及应用研究[D]. 南京大学, 2016. 随着互联网对人们日常生活的渗透不断深入,人们在进行经济决策时往往会借助网络进行必要的信息检索,因此人们在互联网的搜索行为也就反过来暴露了其自身的经济活动的意向。基于这样的假设基础,对互联网搜索数据的充分挖掘可以使得研究者提前捕捉到相关消费领域里潜在消费者的购买意向,并以此来对该领域未来一段时间内由供求关系决定的价格水平进行一定程度的预测。谷歌最早于2004年推出了关键词搜索量数据查询服务,因此国外研究者早在2009年便开始了利用互联网搜索数据预测社会经济领域相关指标的研究。国内最大的搜索引擎百度,于2011年开始提供关键词搜索指数的数据服务,因此从2013年起国内学者也开始利用百度指数进行相关领域的预测研究,但总体而言不论是从研究数量还是研究成熟度上而言都有待提高。本文在梳理已有研究成果的基础上,以北京、上海、广州、深圳、南京、杭州6个城市为研究对象,探索了从关键词选择到复合指数合成再到模型构建及检验的一整套程序和方法,得到了6个城市二手房和新房网络搜索复合指数,并在此基础上利用1000次2折交叉验证法构建了基于网络搜索的二手房和新建商品房价格预测模型。研究结果显示,时差相关分析法筛选并用主成分分析法合成的网络搜索复合指数对房价指数具有较好的拟合和预测效果。但对基于网络搜索的房价预测模型与房价预测基本模型的预测效果进行比较后,笔者发现加入网络搜索指数后的预测模型对基本预测模型的改进作用有限,并对其原因进行了分析,对提升和挖掘网络搜索数据的预测能力提出了进一步完善的方向。

杨树新, 董纪昌, 李秀婷. 基于网络关键词搜索的房地产价格影响因素研究[J]. 新疆财经大学学报, 2013(3):5-12. 影响房地产价格的因素众多,而在当前房地产市场“刚性需求”和“投资需求”相互交织、难以区分的情况下,人们的预期就成为影响房地产价格的一个关键因素,这种预期在互联网已相当普及的今天可以通过网络搜索引擎反映出来.本文尝试构建网络搜索数据与房地产价格指数之间的关系框架,通过皮尔逊相关系数法,选取了与房屋销售价格指数相关系数在0.55以上的9个关键词,即房价、出租信息、二手房产网、二手房信息、开盘、户型装修、房产网、房源和装修材料；并通过实证检验,得出房屋销售价格指数与提前5个月搜索指数的相关性最大,即对房价感兴趣的购房者通常都会提前5个月左右搜索相关信息.

董倩. 基于网络搜索数据的房地产价格预测[C]// the 5th statistics annual conference csac2014. 2014.房地产业是我国国民经济发展的重要支柱产业,房地产价格走势一直是人们 关注的热点,但政府统计部门发布的数据往往具有一定的时滞,不能满足大众的 需求。为解决这一问题,本文以北京、上海、广州、南京、沈阳和西安 6 个大中 城市的二手房价格和新房价格为研究对象,以来自我国最大搜索引擎的百度搜索 指数为数据基础,首先选出了对二手房价格变动影响最大的 12 个关键词和对新 房价格变动影响最大的 8 个关键词;然后采用交叉验证技术,运用线性回归、回 归树、随机森林、Bagging、m-Boosting、神经网络、支持向量机和混合线性回归 8 种模型分别对 6 个城市的二手房价格和新房价格进行了拟合和预测;最终在 8 种模型中选择的最优模型成功地预测了 6 个城市的二手房和新房销售价格指数。 结果显示:在采用的 8 种模型中,支持向量机和随机森林模型的预测效果最佳, 回归树模型的预测效果最差;在二手房和新房价格的网络搜索关键词中,宏观经 济形势和房产政策是关注的重点。可见,网络搜索数据不但能够较好地预测房价 指数,同时能够分析得出经济主体行为的趋势与规律,而且具有一定的时效性, 预测的月度房地产价格指数能够比官方数据发布提前约两周

韩琳, 吴华瑞, 顾静秋. 基于K-Means聚类的农产品价格异常数据检测[J]. 计算机系统应用, 2017, 26(3):139-143. 使用K-means 技术解决根据已有海量的农产品市场价格数据来检索异常价格和计算价格边界的问题。采用的数据为农产品市场价格数据，具体包括副食品、瓜果、粮棉油糖、蔬菜和水产品等类别, 全国各个省市地区历年的农产品 市场价格数据, 例如茄子、生姜、芹菜、山药、大葱、 鸡蛋、牛肉、白鲢活鱼、粳米、菠菜、苹果、和菠萝 等 11914 个品种的 1300 万条市场价格数据。采用数据挖掘聚类技术K-means 技术进行分析, 改进传统 K-means 算法并提出了基于 K-means 的农产品市场价格异常数据检测方法. 实验 结果表明, 改进的算法可以很好地筛选出离群点数据, 找到异常价格元素从而获得农产品市场价格边界. 目 前该算法已经应用于农业智能搜索引擎市场价格分析 模块, 效果显著

王文亮. 基于网站搜索数据的电子商务交易预测的研究——以某网站为例[D]. 中国科学院研究生院, 2012. 讨论了现有的网络关键词搜索、电子商务交易量两方面相关的研究，给出了一种通过网络关键词对电子商务交易量的预测方法。本文主要工作和特色如下:1）首先在建立网络关键词搜索和电子商务交易量理论的基础上，建立了理论分析框架，以揭示网络搜索数据与电子商务交易量之间存在的相关关系及先行滞后关系;2）其次通过逐步合成方法合成得到进入回归模型的关键词合成指数;3)然后实证分析中国某电子商务网站的搜索数据和交易量与相应合成指数之间的关系;4）进行预测对比试验。电子商务网站的搜索数据与交易量之间具有较高的相关性;加入合成的搜索指数后，模型拟合度达到0.901，模型的预测效果明显提高。

　　 传统预测方法的数据来源于相关统计部门或咨询公司的调研报告，但这些数据的收集和公布往往存在滞后性，且数据量通常难以达到模型的预测要求，这极大地制约了预测的有效性。我们通过对电子商务网站搜索数据进行过滤，应用逐步合成法进行预测交易量。实验结果证明模型具有较强的时效性，可以更为及时地预测电子商务交易量，进而帮助电子商务企业完善库存管理、订货计划等。

李新. 基于海量数据的投资者关注与国际原油价格研究[D]. 中国科学院大学, 2015.国际原油价格的分析和预测一直是学界、企业及投资者密切关注的热点问题，目前国内外的研究主要集中在构建基于传统统计数据的分析框架，而很少考虑使用海量互联网数据。随着大数据时代的到来，基于互联网所生成的海量数据的分析及应用逐渐得到学术界和业界的重视。有别于传统统计数据，海量数据大多由互联网用户直接产生，能够一定程度上反映互联网用户对当前热点问题的关注及趋势。大数据的应用给很多传统行业带来变革，这意味着传统的国际原油市场的研究需要重要的理论与方法创新。因此，在大数据时代背景下，研究海量数据在国际原油市场的分析及应用具有重要的理论与现实意义。  
　　本文研究了大数据时代下，基于海量互联网数据分析及建模的理论基础、研究框架及关键模型、技术和方法。在研究框架的指导下，本文构建反映投资者对国际原油市场关注度的衡量指标，探讨投资者关注与国际原油市场上不同类型的投资者交易持仓关系。进一步构建经济计量模型分析投资者关注对国际原油价格及价格波动的影响程度。最后，综合计量经济模型、数据挖掘技术与集成预测等方法提出了基于海量数据的国际原油价格预测模型，并对模型进行综合评价。本文的主要特色及创新工作包括以下几个方面:  
　　(1)提出了针对互联网海量数据的指数构建及预测的理论框架，包含对互联网海量数据的获取、分析、合成、预测及评价等部分。以互联网数据作为基础，通过网络文本挖掘、高维特征选择，指数构建方法从海量数据中挖掘并构建综合指数;再将传统的计量经济模型和人工智能技术等综合集成起来，形成一个系统的针对海量数据的分析及预测框架。在这个框架下，可构造多种不同的分析及预测方法。这个研究框架不仅可以将互联网海量数据应用到国际原油市场投资者及原油价格波动的分析及预测中，也可以扩展到其他研究领域如宏观经济及其他能源价格的研究中，为传统领域的分析预测问题提供大数据背景下新的研究方法及技术支持。  
　　(2)构建了基于互联网海量数据的反映国际原油市场的投资者关注指数。在理论研究框架的指导下，采用网络挖掘-特征选择-广义动态因子模型(WM-FS-FHLR)提取互联网搜索关键词，进行特征选择筛选变量，并构建反映互联网关注的“一致指数”。在对投资者关注指数的评价上，首次引入生物信息领域的因果分析方法Partial Directed Coherence(PDC)，从而分析投资者关注指数与国际原油现货及期货价格的因果关系，从而对所构建的指数进行综合评价。这一个基于互联网数据所构建的综合指数可用于更全面地衡量原油市场中部分投资者的关注，更及时且准确的表征原油市场的动态变化，所提出的WM-FS-FHLR技术是一个重要的方法创新。  
　　(3)建立计量经济模型分析了投资者关注指数与原油市场投机活动的关系，并讨论了机构投资者和个体投资者的交易持仓与投资者关注指数的差异。数据来源丰富，分别采用美国商品期货交易委员会（Commodity Futures TradingCommission, CFTC）发布的非商业投资者、商业投资者与散户的交易持仓数据、非商业投资者净持仓比例、Working's投机指数、原油库存、原油期货及相关调查数据等构建的衡量指标等多种来源的数据。构建多组VAR模型来研究投资者关注指数对非商业投资者、商业投资者及散户的持仓关系，从而分析投资者关注指数与投机活动的领先性。首次从原油市场投机活动为出发点，构建模型探讨投资者关注指数的“内涵”，以及关注指数对机构及个人投资者持仓的影响，研究成果是对已有文献的重要补充。  
　　(4)提出了基于海量数据的投资者关注对原油价格的“非对称”影响关系的马尔科夫状态转移模型。首先建立单变量时间序列模型分析投资者关注对国际原油价格的影响，并且通过“象限图”更为直观的表明投资者关注指数对国际原油价格的影响并不是一致的，且该影响与原油价格本身的趋势相关。进一步，构建了两状态下的马尔科夫状态转移模型研究投资者关注对国际原油价格的影响，实证结果表明，投资者关注对原油价格的影响在原油价格上升和下降两个不同的阶段表现是不对称的，并且通过引入其他解释变量和改变样本区间等方法对研究结果进行稳健性检验。本文所提出的投资者关注对价格的非对称影响关系模型可扩展到其他研究领域如宏观经济及金融市场中，从而探讨在不同经济周期阶段（如繁荣、衰退、萧条、复苏）中，基于互联网的海量数据对于经济及金融指标的影响，是重要的研究方法及研究成果创新。  
　　(5)建立了投资者关注与国际原油价格收益及波动关系的EGARCH模型。首先计算原油价格收益及其波动序列，并建立包含投资者关注与原油价格收益的VAR模型，并通过分析VAR模型的脉冲响应函数研究了投资者关注与原油价格收益序列之间的相互影响方向及幅度的不同。在价格波动研究方面，通过建立EGARCH模型研究投资者关注对原油价格波动的关系，着重探讨了投资者关注指数对国际原油市场价格收益及波动的影响程度，以及投资者关注指数是否会对价格波动产生不对称影响。这一研究成果为后续原油价格波动率的分析及预测提供新的且重要的数据及模型支撑。  
　　(6)提出了基于海量数据的国际原油价格集成预测模型。通过集成时间序列模型、多元回归模型、和具有不同核函数的支持向量回归模型等方法研究如何在原有数据的基础上，通过引入海量数据构建的投资者关注来提高国际原油价格预测精度，通过样本外滚动预测等方法综合评价该方法的预测精度，并探讨了该框架在其他研究问题中的可扩展性等。本文所提出的基于海量数据的预测模型为传统原油价格预测引入新的数据，并为更及时和准确的原油价格预测提供多种方法和技术。  
　　综上所述，本文采用互联网的海量数据构建原油市场的投资者关注度，提出了基于海量数据研究的分析和预测框架，并对投资者关注与不同类型的投资者交易持仓、国际原油价格及波动关系进行深入探讨。本文所研究成果为已有的国际原油价格的分析及预测贡献有价值的结论。本文的研究紧扣大数据时代的研究热点趋势，不仅分析了互联网生成的海量数据与国际原油市场的关系，而且对于原油市场价格分析、投资决策具有指导意义。本文所提出的研究框架对其他研究领域如宏观经济和其他能源行业的分析及价格预测也具有重要的借鉴意义。

徐振敬. 基于情感分析的国际原油价格走势预测研究[D]. 北京化工大学, 2016.石油价格的增长或者下跌对于世界经济的发展会产生重大的影响,因此,石油价格的预测显得尤为重要。但是,由于石油这种商品的复杂性和不规律性,使得石油价格预测成为一个非常困难的问题。石油价格基本是由石油出口国的供给和石油进口国的需求平衡共同决定的,但是由政治事件或者经济因素导致的供给的不规律性往往会导致石油价格的不规律性。在这种情况下,石油价格上涨或者下跌的预测对于决策者就显得非常有价值。同时,随着互联网和大数据技术的发展,随之产生了大量新闻数据。一些新闻内容代表了对金融市场未来趋势的实时评估,新闻的内容将会影响金融投资者的投资行为,进而影响金融市场。如果能充分利用这些数据,那么将有助于石油价格趋势的预测。在这种背景下,本文提出了基于情感分析的国际原油价格走势预测模型。该模型主要基于石油相关新闻的分析,通过采用领域关键词词典的方法,得到新闻的情感序列,再通过格兰杰因果检验的方法,得到情感序列和石油价格序列的相关性和滞后期,最后通过机器学习的方法(支持向量机、决策树、逻辑回归和神经网络)预测石油价格的走势。同时,为了验证新闻情感对于石油价测具有预测能力,选取了美国西得克萨斯轻质原油(WTI)和路透社原油新闻作为研究对象进行案例分析。结果表明,新闻情感和原油价格之间确实存在着格兰杰因果关系,即新闻情感的变化会引起石油价格的变化,这说明新闻情感对于石油价格的走势具有预测能力。实验结果还表明,对于大部分预测模型而言,新闻情感的引入一般能极大地提高石油价格走势的预测准确率。

Lee W, On B W, Lee I, et al. A big data management system for energy consumption prediction models[C]// Ninth International Conference on Digital Information Management. IEEE, 2014:156-161.

In this work, we develop a prototype about a big data management system for storing, indexing, and searching for huge-scale energy usage data. Rather than existing, commercial relational databases such as Oracle and IBM-DB2, this system is able to provide us with high availability and performance at low cost. It is also able to manage unstructured data and store big data in distributed environment. In addition, using data access APIs, target data is quickly retrieved from our proposed system. To utilize our prototype system, we also propose an energy consumption prediction model based on penalized linear regression-based map/reduce algorithms. Then, we exploit discriminate features with respect to time stamp. Finally, given a time stamp (e.g., 2014-01-05 12:01:08), our proposed learning model will give us a predicted value about the energy usage (e.g., 90 watt) at that time. According to our experimental results obtained from about 7.5 million records, each of which consists of an energy usage and time stamp during three months in 2014, it turns out that our prediction model can predict real values that are very close to actual energy usage at that time, and is about 1.72 times faster than in a single machine.