

大数据：金融情绪分析的深度学习

大数据杂志

2018年12月, 5:3 | 引用为

萨哈尔·索汉吉尔 (1) 电子邮件作者 (ss)
配置文件 (View OrcID profile)

王定定 (1)

安娜·波美拉涅茨 (2)

Taghi M.Khoshgoftaar (1)

✓ 网页内容翻译完成

显示原文

中英对照

关闭

1. 计算机与电气工程与计算机科学系, 佛罗里达大西洋大学, , 博卡拉顿, 美国
2. 商业学院, 佛罗里达大西洋大学, , 博卡拉顿, 美国

开放存取

研究

第一次在线: 2018年1月25日

收到: 2017年4月20日

接受: 2017年12月28日

12 股份

9.4k 下载

4 引文

下列专题收藏的一部分:

1. 先进的软计算方法及其在社会媒体大数据分析中的应用

摘要

深度学习和大数据分析是数据科学的两个重点。近年来,深度学习模型在语音识别和计算机视觉方面取得了显著的效果。大数据对于需要收集大量数据的组织来说很重要,比如社交网络,而使用深度学习的最大资产之一就是分析大量的数据(大数据)。这一优势使得深度学习成为大数据的一个有价值的工具。深度学习可以用来提取隐藏在大数据中的难以置信的信息。现代股票市场就是这些社交网络的一个例子。他们是一个很受欢迎的地方,以增加财富和创造收入,但根本的问题,什么时候购买或出售股票,或购买哪些股票尚未得到解决。在投资者中,有专业的财务顾问是很常见的,但支持这些人作出决定的最佳资源是什么呢?高盛(GoldmanSachs)、雷曼兄弟(LehmanBrothers)和所罗门兄弟(SalomonBrothers)等投资银行十多年来一直在金融咨询领域占据主导地位。然而,通过互联网和像StockTwits和SeekingAlpha这样的金融社交网络的普及,世界各地的投资者有了新的机会来收集和分享他们的经验。个人专家可以合理准确地预测金融社会网络中股票市场的走势,但这些专家群体对各种股票的看法是什么呢?在本文中,我们试图确定是否可以采用深度学习模型来提高StockTwits的情感分析性能。我们将长期记忆、doc2vec和卷积神经网络等几种神经网络模型应用于股票市场。结果表明,深度学习模型可以有效地用于财务情绪分析,而卷积神经网络是在StockTwits数据集中预测作者情绪的最佳模型。

关键词

我们使用cookie来个性化内容和广告，提供社交媒体功能，并分析我们的流量。我们亦会与我们的社交媒体、广告及分析合作伙伴分享有关阁下使用本网站的资料。隐私声明。您可以在ManageCookies中管理您的首选项。

管理饼干

好的

等等，利用互联网的力量帮助他们的用户做出

网页内容翻译完成

显示原文

中英对照

关闭

此外，还有一些网站让用户有能力与专业人士进行协商，其中一个最受欢迎的话题就是投资。高盛(GoldmanSachs)和雷曼兄弟(LehmanBrothers)等公司有超过150年的投资建议。在互联网时代，世界各地的独立分析师和散户投资者可以通过网络相互合作。“搜寻阿尔法”和“StockTwits”是聚焦于股市的两种常见金融社交媒体平台，为用户提供了一种连接信息和相互联系、扩大投资的途径。^{2]}

金融社交媒体将人、公司和组织聚集在一起，以便他们能够产生想法并与他人分享信息。正是这种媒体提供了大量的非结构化数据(大数据)，可以集成到决策过程中。这样的大数据由于其高频率的创建和低成本地获取，可以被认为是实时估计的一个很好的来源。

情感分析(SA)是一种常用的方法，它越来越多地被用来评估社交媒体用户对某一主题的感受。进行情感分析的最流行方法是使用数据挖掘。我们的中心思想是采用深度学习来根据投资者的信息来确定投资者对股票价格和整个市场的预期。我们之所以选择深度学习方法而不是数据挖掘，是因为在数据挖掘中，识别特征和选择最好的特征是最具挑战性的任务，尤其是在大数据中。

与数据挖掘相比，深度学习模型在学习过程中学习特征。深度学习算法导致抽象表示，因此，它们可以不受输入数据局部变化的影响。此外，借助深度学习可以更好地解决大数据问题，包括语义索引、数据标记和快速信息检索。深度学习为使用更简单的模型来完成复杂的人工智能任务提供了机会。虽然深度学习算法已被应用于一些大数据领域，如计算机视觉^[3, 4, 5, 17, 18, 22]和语音识别^[6, 7, 8, 9, 10, 11]在大数据分析方面，它仍然完好无损。本文对深度学习在财务数据情绪分析中的应用进行了评价。

深度学习算法提供了在高抽象级别提取复杂数据的机会，其方式是将抽象程度较高的高级特征定义为抽象程度较低的低级特征^[12]数据的不同来源(如光、物体形状和图像中的物体材料)可以通过使用深度学习来分离。深层次学习的思想来源于人脑新皮层的初级感觉区^[13]。

卷积神经网络^[14]是各种深度学习模式的一个例子。该模型广泛应用于图像分析中，可以通过卷积层利用数据的内部结构。由于文本文档内部存在的内部结构，CNN在文本数据方面也得到了广泛的关注。CNN用于标记、实体搜索、句子建模等系统。^{15, 16, 17, 18, 19, 20, 21, 22, 23]}。

深度学习算法通常以贪婪的方式学习数据表示，从大数据中学习似乎更有用。^{24, 25]}深度学习可以在大数据分析中提取非线性复杂特征，然后提取特征作为线性模型的输入。

深度学习可以使大数据分析的鉴别任务变得更容易。例如Li等人[26]使用深度学习在大数据中进行搜索。他们使用深度学习来启用语音搜索音频和视频文件。从大量无监督的数据

我们使用cookie来个性化内容和广告，
提供社交媒体功能，并分析我们的流
量。我们亦会与我们的社交媒体、广告
及分析合作伙伴分享有关阁下使用本网
站的资料。隐私声明。您可以在
ManageCookies中管理您的首选项。

☐ 管理饼干

✓ 好的

的一些特点，并指出了大数据在数据分析中面
感分析的方法和深度学习在情感分析中的优势
讨论使用数据挖掘在StockTwits数据集上进行情
种特征选择方法。在“大数据分析中的深度学习”我们将探讨如何将深度学习用于大数据分
析，我们还讨论了深度学习需要克服的一些挑战，以便在大数据领域进行分析；”结果和
讨论“本节将解释我们的实验，并深入探讨如何将深度学习应用于财务情绪分析。我们的
主要发现和结论载于“结论”分段。

✓ 网页内容翻译完成

显示原文

中英对照

关闭

相关工作

特定的大数据领域，包括计算机视觉[29]和语音识别[30]，已经看到了使用深度学习改进
分类建模结果的优势，但是，关于情感分析的深度学习体系结构的研究还很少。2006年
Alexandrescu等人[31]给出一个模型，其中每个单词都表示为特征向量。使用单个嵌入
矩阵来查找所有这些特征。Luong等人[32]使用递归神经网络(RNN)对词的形态结构进行
建模，并学习形态感知的嵌入。2013年Lazaridou等人[33]尝试使用组合分布语义模型来
学习短语的意义。2013年，Chrupala使用一个简单的递归网络(SRN)来学习字符序列的
连续向量表示。他们使用自己的模型来解决字符级别的文本分割和标记任务。利用递归神
经网络可以通过深度学习构造有意义的搜索空间。34]Socher等人，2011年[35]，使用递
归自动编码器[36, 37, 38, 39]为预测情绪分布，提出了一种半监督的方法模型。2012年
Socher等人[40]提出了一种具有学习任意长度句子组成向量表示能力的语义合成模型。
他们提出的模型是矩阵向量递归神经网络模型.[中]提出的递归神经张量网络(RNTN)结
构。41]RNTN使用单词向量和解析树表示短语，然后使用基于张量的组合函数计算较高
节点的向量[42]

关于NLP任务的卷积网络，Colobert等人。15]对于语义角色标注任务，利用卷积神经网
络避免过多的特征工程。在2011年，Colobert使用了类似的网络架构来进行语法分析。
在[43]提出了一种利用字符到句子级信息进行短文情感分析的深卷积神经网络。

本文的实验重点是市场情绪。根据[44]，市场情绪是投资者对预期市场价格发展的普遍态
度。这种态度是各种因素的结合，如世界事件、历史、经济报告、季节性因素等等。市场
情绪是通过情绪分析(又称意见挖掘)发现的。45]，即利用自然语言处理方法从原始材料
中提取作者的态度。

王和桑巴斯文2]在StockTwits数据集上应用市场情绪，使用监督情绪分析，将StockTwits
中的分类信息归类为“看涨”或“看跌”。如果投资者认为股票价格会随着时间的推移而上
涨，并建议购买股票，那么他或她就会被认为是多头投资者。相反，如果投资者看跌，他
或她会预期价格下跌，并会建议出售股票或反对购买。

这一领域最受欢迎的作品之一是拉夫兰和麦克唐纳。[46\]](#)从1994年到2008年，他们使用美国证券交易委员会的门户网站制作了一个金融词典，并手动创建了六个单词的列表。包

我们使用cookie来个性化内容和广告，提供社交媒体功能，并分析我们的流量。我们亦会与我们的社交媒体、广告及分析合作伙伴分享有关阁下使用本网站的资料。隐私声明。您可以在ManageCookies中管理您的首选项。

☐ 管理饼干

✓ 好的

放在“低估股票”这样的词之外，它就会变得积极。

✓ 网页内容翻译完成

显示原文

中英对照

关闭

近年来，深度学习方法由于其相对于其他方法的优势，已经成为大数据中情感分析的有力工具。这些优点之一是，在深度学习过程中，特征是分层学习的，而不是数据挖掘中需要的特征工程。此外，在深度学习方法中，每个单词都被视为句子的一部分。这样，包含在词序、邻近性和关系中的相关信息就不会丢失。此外，深度学习得益于相似模型。与单词模型的包相比，Word嵌入创建了一个具有更低维空间的单词向量表示[\[53, 54\]](#)因此，表示向量空间中相似单词的向量更接近于一起。深度学习算法的另一个主要概念是自动抽取表示(摘要)[\[55\]](#)为了实现这一目标，深度学习使用了大量的无监督数据(大数据)，并自动提取复杂的表示。利用深度学习算法提取抽象表示的优点之一是它们的泛化性。从给定数据集中提取的特征可以成功地用于另一个数据集上的鉴别任务。深度学习是人工智能的一个重要方面，它提供了复杂的大数据表示，并使机器独立于人类知识。

深度学习为图像和视频数据构造了高度抽象的复杂表示。深度学习提供的高级数据表示可以用于大数据的更简单的线性模型。这种表示可以用于图像索引和检索。也就是说，在大数据分析的背景下，深度学习可以用于语义标注的鉴别任务。

方法论

数据集

幸运的是，我们得到了StockTwitsInc.的许可，可以访问他们的数据集。StockTwits是一个金融社交网络，成立于2009年。关于股票市场的信息，如最新的股票价格，价格变动，股票交易历史，买卖建议等等，都提供给StockTwits用户。此外，作为一个社交网络，它提供了在股票市场交易者之间分享经验的机会。通过StockTwits网站，投资者、分析师和其他对市场感兴趣的人可以提供一条关于股票市场的短信，字数限制在140个字符以内。此消息将发布到所有网站访问者都可以看到的公共流中。此外，作者还可以将信息标记为看涨或看跌，以具体说明他们对各种股票的情绪。

在我们的实验中，我们使用了2015年前六个月发布的信息。每条消息包括一个消息ID、一个用户ID、作者的追随者数量、一个时间戳、股票的当前价格以及其他记录保存属性。我们查看了这些帖子，看看未来的股价和用户的情绪是否有关系。换句话说，我们想看看我们是否能够根据许多用户的当前情绪来预测未来的股价。

我们可以使用Pearson相关系数[\[6\]](#)看一只股票的未来价格和用户的情绪之间是否存在线性关系。Pearson相关函数是衡量两个变量间线性相关性最广泛使用的函数之一。如果两个输入变量之间存在完美的正相关，则返回一个， $(-1,1)$ 如果存在完全负相关，则为0。股票

价格与一般用户情绪之间的Pearson相关系数为0.05，这意味着只有53%的用户能够正确预测未来的股票价格。这比随机猜测要好一些，所以如果预测的数量增加，我们将检查这

我们使用cookie来个性化内容和广告，提供社交媒体功能，并分析我们的流量。我们亦会与我们的社交媒体、广告及分析合作伙伴分享有关阁下使用本网站的资料。隐私声明。您可以在ManageCookies中管理您的首选项。

管理饼干

好的

了解顶级作者的情绪，我们可以准确地预测75中只有10%的信息被标记。为了提高股票价格来分析顶级作者的情绪。深度学习有利于面对提供的数据一样。在本文中，我们采用深度学习的方法对顶尖作者进行情感分析。我们认为，利用深度学习可以极大地提高对各种选股情绪分析的正确分类，从而超过目前股票价格预测的准确性。

网页内容翻译完成

显示原文

中英对照

关闭

大数据

自上世纪90年代以来，“大数据”一词就一直在使用。在2012年，Gartner更新了他以前对大数据的定义，并将其定义为：“大数据是高容量、高速度和/或高多样性的信息资产，需要成本效益高、创新的信息处理形式，从而能够增强洞察力、决策和过程自动化”。大数据是指使用传统软件工具和技术难以管理和分析的日益增长的数字数据。大数据通常具有大量的样本、大量的类标签和非常高的维度(属性)。2012年持续移动的大数据的目标大小在几十兆字节到许多千兆字节之间。有四个属性定义了大数据，包括体积、变化、速度和准确性[56]显然，数据量是大数据的主要属性。随着大数据量的增加，数据的复杂性和底层关系也随之增加。大数据系统中的原始数据是无监督和多样化的，尽管它可以包含少量有监督的数据。包括Facebook、Twitter、StockTwits、LinkedIn在内的许多社交媒体公司都拥有大量的数据。随着数据越来越大，深度学习方法在提供大数据分析方面变得越来越重要。

另一件让大数据变得更大的事情是数据的多样性。来自不同来源的大数据比以往任何时候都要多。Web资源(包括社交媒体、点击流和日志)就是这些资源的一些例子。大数据处理的挑战之一是处理各种不同的数据。为了提取数据的结构化表示，大数据需要对非结构化数据进行预处理。

速度是大数据的另一个特征。大数据中的数据生成频率较快。例如，考虑来自StockTwits网站的消息流。速度和大数据的体积和变化特性一样重要。处理输入到可用信息的快速性对于处理与大数据相关的速度非常重要。

准确性是指大数据中数据的可信度。通过增加数据源和类型的数量，对大数据的信任成为一个实际的挑战。除了这四个版本之外，还有许多挑战，包括数据清理、功能工程[57, 58, 59]，大数据分析所面临的高维性和数据冗余。深度学习是用于工业产品，有机会拥有大量的数字数据(大数据)。谷歌使用深度学习算法和互联网上的大数据作为谷歌的翻译器。在一些大数据应用领域，如社交媒体、市场营销和金融数据提要中，使用深度学习算法和体系结构来分析大规模[60, 61]，我们鼓励快速移动的流式数据，但仍然没有通过使用深度学习应用程序来分析大数据。

准确度、精确度、召回率、F-测度、AUC 作出决定和采取行动是至关重要的。大数据有可能给科学和社会的各个方面带来巨大的变化。但从大数据中提取信息并不是一件容易的

我们使用cookie来个性化内容和广告，提供社交媒体功能，并分析我们的流量。我们亦会与我们的社交媒体、广告及分析合作伙伴分享有关阁下使用本网站的资料。隐私声明。您可以在ManageCookies中管理您的首选项。

☐ 管理饼干

✓ 好的

据，这对于分布在不同位置的大数据来说是一需要克服大数据的稀疏性、异构性、不确定性。国创新和经济革命的大交易和基础。甚至在政有效方法。2012年，奥巴马政府宣布了一项“大数据研究和开发计划”，以帮助解决国家面临的一些最紧迫的挑战。

✓ 网页内容翻译完成

显示原文

中英对照

关闭

情感分析

继早期的情绪分析工作之后[63, 64]，我们研究原始材料并应用自然语言处理技术来确定作者对某一主题的态度。一般来说，情感分析是一种将文本文档分类为多个组的形式。大多数情况下，我们只需将文档分为正反两类。[65]此外，在情感分析中，有不同的方法可以帮助我们衡量情绪。这些方法包括基于词汇的方法和有监督的机器学习方法。机器学习模型是比较流行的，因为基于词汇的方法，基于词的语义，使用一个预先定义的积极和消极的词列表来提取新文档的情感。创建这些预定义的列表是费时的，我们不能构建一个独特的基于词法的字典，以便在每个单独的上下文中使用。随着社交媒体的日益普及，评论、博客和社交网络提要的庞大数据集(大数据)不断生成。大数据技术用于我们收集和维持大量数据的应用领域。不断增长的数据、密集的技术和不断增加的数据存储资源发展了大数据科学。大数据分析的主要概念是从大量数据中提取有意义的模式。大数据需要特殊的方法来从大量的数据中提取模式。深度学习提供了解决大量数据(大数据)中存在的学习和数据分析问题的解决方案，而且它们更善于学习复杂的数据模式。还有其他的大数据问题，如域采用和流数据，大数据分析的大规模深度学习模型必须解决这些问题。随着企业、组织和个人寻求更好地利用他们的大数据，情感分析中能够帮助我们从这些领域提取信息的概念和方法变得越来越重要。在接下来的章节中，我们开始研究基于数据挖掘方法的情感分析的性能。

表1

StockTwits数据集上Logistic回归的性能

准确度	精确度	召回	F-测度	AUC
0.7088	0.7134	0.6980	0.7056	0.7088

基于数据挖掘方法的情感分析

王英²使用有监督的数据挖掘方法在StockTwits数据集中查找消息的情感。他们从消息中删除了所有的句号、股票符号和公司名称。他们将地面真相相信自视为训练数据，并测试多

我们使用cookie来个性化内容和广告，提供社交媒体功能，并分析我们的流量。我们亦会与我们的社交媒体、广告及分析合作伙伴分享有关阁下使用本网站的资料。隐私声明。您可以在ManageCookies中管理您的首选项。

☐ 管理饼干

✓ 好的

尖，尖似于无层层的深度学习。在深度学习的换句话说，深度学习算法作为多个特征学习步用于大型数据集。

✓ 网页内容翻译完成

显示原文

中英对照

关闭

我们跟着王²方法和应用Logistic回归^[3]在StockTwits数据集上。表中¹基于不同的性能指标，给出了StockTwits数据的Logistic回归性能。也在图中。¹，我们给出了中华民国的曲线。^{66]}这个模型。

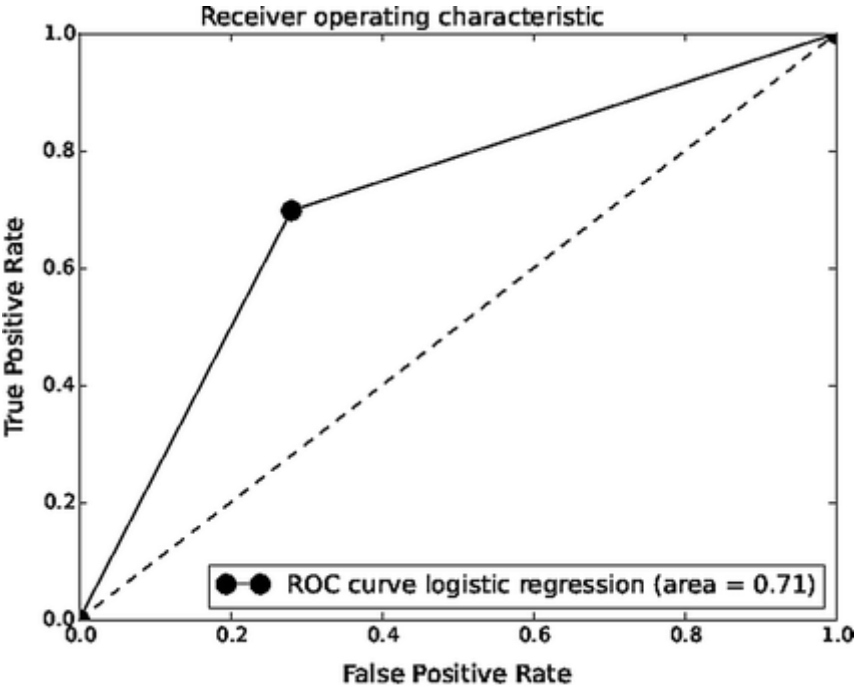


图1

Logistic回归接收机的工作特性

利用特征选择提高精度

阻碍我们对大数据进行精确分类的问题之一是其中发现的噪音。特征选择，包括去除噪声特征和消除无效词汇，使训练和应用分类更加有效。^{67]}现有的寻找适当特征子集的方法可分为两组：特征过滤器和功能包装器^[68]在特征滤波器中，根据这些特征的统计特性选择最终的特征集。使用特性包装器，通过建模工具的结果应用迭代搜索过程。在每次迭代中，在建模工具中使用一组候选特征集，并记录结果。每个步骤都使用上一步的结果，因此生成新的暂定集。这个过程是重复的，直到满足某些特定的收敛准则。在我们的实验中，我们有大量的特性和实例，因此，我们的数据非常稀少。我们尝试了几种特征选择方法，看看它们将如何影响我们的情感分析的准确性。

在特征的准确度、精确度、召回、F-测度器、AUC、X-平方、方差分析和互信息。使用这些特征选择技术的优点在于它们的速度、可扩展性和独立于分类。我们选择这些方法的理

我们使用cookie来个性化内容和广告，提供社交媒体功能，并分析我们的流量。我们亦会与我们的社交媒体、广告及分析合作伙伴分享有关阁下使用本网站的资料。隐私声明。您可以在ManageCookies中管理您的首选项。

☐ 管理饼干

✓ 好的

我们使用了我们的数据集，并逐次减少特征的数量来排名的，这是由Eq决定的。1)在这个方频率。高(X²)Score拒绝了关于术语和类的独

✓ 网页内容翻译完成

[显示原文](#)

[中英对照](#)

[关闭](#)

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

将X-平方应用于我们的数据集，并逐渐减少特征的数量，使我们能够看到它如何影响Logistic回归的性能。分类结果见表2。在某些情况下，减少特征的数量会提高准确性-例如，通过将特征的数量从40,000减少到500，精确度提高了7%。然而，这是一个不规则的数据集，并不意味着X-平方是一种有效的特征选择方法，以提高我们的分类器的准确性。

表2

StockTwits数据集上X平方特征选择的性能

特征	准确度	精确度	召回	F-测度	AUC
55,820	0.7088	0.7134	0.6980	0.7056	0.7088
40,000	0.4796	0.4851	0.6645	0.5608	0.4796
20,000	0.5018	0.5013	0.6879	0.5800	0.5018
4000	0.5274	0.5206	0.6946	0.5951	0.5274
2000	0.5221	0.5190	0.6036	0.5581	0.5221
400	0.5308	0.5278	0.5834	0.5542	0.5308
200	0.5333	0.5280	0.6284	0.5738	0.5333
50	0.5314	0.5232	0.7071	0.6014	0.5314

方差分析

我特征用的准确度特精密度方法召回差F-测度NC AUC特征选择。阿诺瓦[71]用于确定独立组的算术平均值之间是否存在统计上的显著差异。在我们的实验中，通过使用方差分析

我们使用cookie来个性化内容和广告，
提供社交媒体功能，并分析我们的流
量。我们亦会与我们的社交媒体、广告
及分析合作伙伴分享有关阁下使用本网
站的资料。隐私声明。您可以在
ManageCookies中管理您的首选项。

☐ 管理饼干

✓ 好的

$$MS_B = \frac{\sum_{i=1}^m (\bar{x}_i - \bar{x})^2}{m-1}$$

✓ 网页内容翻译完成

显示原文

中英对照

关闭

(3)

在群体内部的可变性， (X_{ij}) 表示ith类中的JTH观测[72].

$$MS_W = \frac{\sum_{i=1}^m \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2}{n-m}$$

(4)

通过基于F检验分数提取更有效的特征，检验了方差分析特征选择是否提高了分类方法的准确性。表中提供的结果3通过方差分析的特征选择，精度没有得到提高，因此不会用于进一步的测试。

表3

StockTwits数据集上方差分析F-检验特征选择的性能

特征	准确度	精密度	召回	F-测度	AUC
55,820	0.7088	0.7134	0.6980	0.7056	0.7088
40,000	0.7094	0.7130	0.7010	0.7070	0.7094
20,000	0.7091	0.7127	0.7007	0.7066	0.7091
4000	0.5274	0.5206	0.6946	0.5951	0.5274
2000	0.7045	0.7048	0.7038	0.7043	0.7045
400	0.6785	0.6638	0.7233	0.6923	0.6785
200	0.6611	0.6378	0.7457	0.6875	0.6611
50	0.6191	0.5863	0.8084	0.6797	0.6191

信息增益

结果表明, 精度、精确度、召回率、F-测度和AUC高分类模型的准确性。在本节中, 我们将讨论互信息特征选择。这是最常用的特征选择方法之一。互信息定义为两个随机变

我们使用cookie来个性化内容和广告, 提供社交媒体功能, 并分析我们的流量。我们亦会与我们的社交媒体、广告及分析合作伙伴分享有关阁下使用本网站的资料。隐私声明。您可以在ManageCookies中管理您的首选项。

☐ 管理饼干

通过使用互信息进行特征选择, 探讨了每个词提取互信息值最高的特征。通过这种方式, 我们得到了包含所有词的大部分信息。在我们的实验中, 用于特征选择的互信息也是无效的。表中提供的结果显示了这一点。4.

☒ 网页内容翻译完成

显示原文

表4

StockTwits数据集上互信息特征选择的性能

特征	准确度	精确度	召回	F-测度	AUC
55,820	0.7088	0.7134	0.6980	0.7056	0.7088
40,000	0.5417	0.5311	0.7115	0.6082	0.5417
20,000	0.5123	0.5087	0.7144	0.5943	0.5123
4000	0.5391	0.5337	0.6190	0.5732	0.5391
2000	0.5406	0.5350	0.6193	0.5741	0.5406
400	0.5665	0.5540	0.6815	0.6112	0.5665
200	0.4713	0.4760	0.5692	0.6126	0.5459
50	0.5077	0.5052	0.7414	0.6009	0.5077

在图中, 2采用卡方、方差分析和信息增益的选择方法减少特征数, 并对Logistic回归的准确性进行比较。结果表明, 特征选择方法不能显著提高Logistic回归的准确性。数据挖掘算法不能提取大数据中存在的复杂非线性模式。提取这些特征后, 深度学习可以使用更简单的线性模型来处理大数据分析任务, 包括分类和预测, 这在我们处理大数据的规模时是非常重要的。

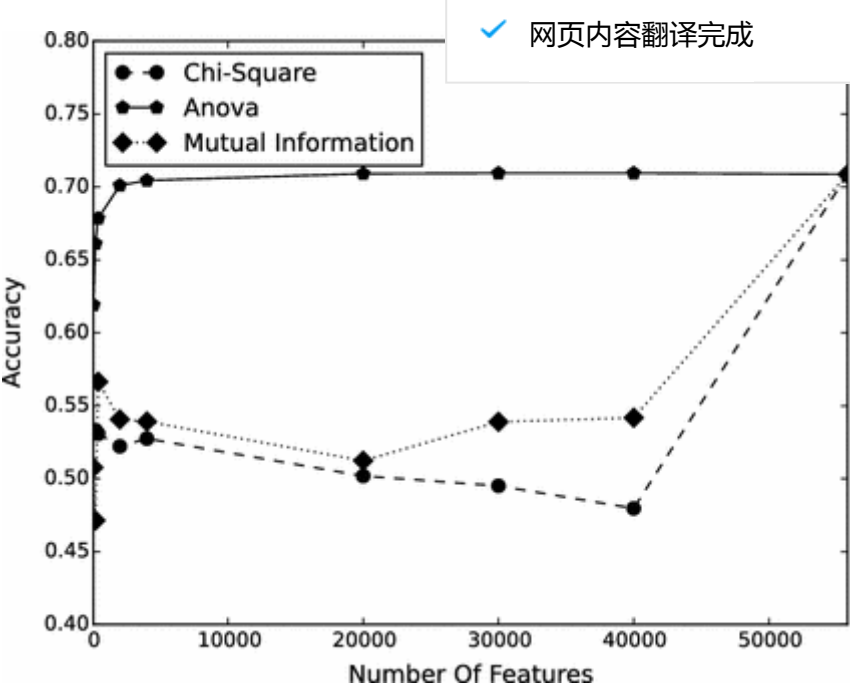
以基于词袋模型的Logistic回归结果为基线，考察了深度学习方法是否能提高大数据中的Logistic回归的准确性。词袋模型不考虑句子中的语序和其他词，而它的词性也是有限

我们使用cookie来个性化内容和广告，提供社交媒体功能，并分析我们的流量。我们亦会与我们的社交媒体、广告及分析合作伙伴分享有关阁下使用本网站的资料。隐私声明。您可以在ManageCookies中管理您的首选项。

[管理饼干](#)

好的

有他们是否能更准确地预测StockTwit用户的T



网页内容翻译完成

显示原文

中英对照

关闭

图2

用特征选择法进行Logistic回归的准确性

大数据分析中的深度学习

在本节中，我们将探讨在大数据分析中使用深度学习算法的优势。此外，我们还将看到一些大数据特性，这些特点对大数据分析中的深度学习提出了挑战。

大数据分析，提供了开发新算法的机会，以解决与大数据相关的一些问题。深度学习算法就是其中之一。例如，深度学习提取的表示可以用于大数据分析。此外，当大数据以更高的抽象形式表示时，大数据分析可以考虑线性建模。使用深度学习算法已经完成了各种各样的工作。

深度学习可以追溯到20世纪40年代。它似乎只是新的，因为在深造流行之前的几年里，深度学习是相对不受欢迎的，因为深度学习经历了许多不同的名称，直到最近才被称为“深度学习”。深度学习被多次改称，反映了不同研究者和不同视角的影响。深入学习史的一些基本背景有助于理解深度学习。深度学习在20世纪40年代至60年代被称为控制论，在1980年代被称为连接主义，目前以“深度学习”的名称重新崛起，始于2006年。[78](#)].

正如我们前面提到的，深度学习算法通过多层次的分层学习来提取大数据的抽象表示。深度学习对于从大数据中提取信息很有吸引力，因为它可以用于从大量未标注的数据中学

我们使用cookie来个性化内容和广告，提供社交媒体功能，并分析我们的流量。我们亦会与我们的社交媒体、广告及分析合作伙伴分享有关阁下使用本网站的资料。隐私声明。您可以在ManageCookies中管理您的首选项。

☐ 管理饼干

深度学习算法可以用未标注大数据中的卷是深度学习的优势之一。由于深度式和资源(大数据中的多样性)的原始数据，并征的需求降到最低。

☒ 网页内容翻译完成

显示原文

然而，大数据具有流化、快速移动等特点，这给大数据采用深度学习带来了挑战。深度学习需要与大量连续的大数据相适应。有一些与深度学习和流媒体大数据相关的作品。例如，[82]说明如何使用深度学习从流数据中学习。周等人[83]描述如何使用深度学习算法对大数据进行特征学习。在大数据中使用深度学习的另一个问题是将深度学习用于大规模模型和海量数据集。在[84]迪安使用数千个CPU核心来训练具有数十亿个参数的深度学习神经网络。Coates等人[85]建议使用GPU服务器集群的强大功能来解决大规模数据集中的深度学习问题。

大数据包含了很多东西，从医学、基因组和生物数据到呼叫中心。要处理与大数据相关的大量输入，需要大规模的深度学习模型。它们可以说明模型参数的最优数目，并克服大数据分析的深度学习挑战。还有其他大数据问题，如域适应和流数据，大数据的大规模深度学习模型需要处理。

大数据的多样性是大数据的另一个特点，它关注的是大数据中输入域和数据类型的变化，因此域的采用问题是大数据分析中深度学习需要解决的另一个问题。有一些研究包括[86, 87]主要集中在学习过程中的域采用。Glorot等人[86]说明深度学习可以分层学习的方式找到中间数据表示，这种表示可以用于其他领域。Chopra等人[87]提出了一种新的领域采用深度学习模型。他们提出的新的深度学习模型考虑了从火车和测试数据之间的分布变化中获得的信息。本文主要集中在信息检索方面，在接下来的章节中，我们总结了深度学习在情感分析中的应用。

基于深度学习方法的情感分析

在前面的章节中，我们讨论了在大数据分析中使用深度学习的一些优点，包括深度学习算法在大数据分析中的应用，以及大数据的具体特性如何导致在大数据分析任务中采用深度学习算法的一些挑战。在这一部分中，我们探讨了使用深度学习算法进行情感分析。在数据挖掘、预测任务中，特征工程是最重要和最困难的技术。在特征工程中所付出的努力是寻找能够自己学习特征的算法的主要原因。深度学习中的层次特征学习提取多层非线性特征，然后一个分类器结合所有特征进行预测。[88]基于支持向量机和决策树等浅层学习的数据挖掘模型不能提取复杂的特征。另一方面，深度学习算法有能力以全局的方式进行推广，生成学习模式，并在大数据中的近邻之外产生关系。[79]为了获得更复杂的特征，深度学习算法首先对图像中的边缘和气泡等特征进行再变换，提取出更多的信息特征来区分分类。这个过程非常接近大脑活动。对特定边缘和气泡敏感的第一层神经元在视觉皮层中接收信息。[89]虽然大脑区域在更远的地方，视觉管道是敏感的更复杂的结构，如面孔。也就是说，深度学习大数据在深层体系结构中的表示，数据经过的层次越多，构造的非线性

向量之后，它们可以有效地用于一项任务，特别是当标记数据受到限制时。分布式的单词包模型忽略了单词上下文作为输入，而是通过从段落中随机抽取样本来预测单词。图中提供了分布式的单词包模型和分布式的段落模型。

我们使用cookie来个性化内容和广告，提供社交媒体功能，并分析我们的流量。我们亦会与我们的社交媒体、广告及分析合作伙伴分享有关阁下使用本网站的资料。隐私声明。您可以在ManageCookies中管理您的首选项。

☐ 管理饼干

✓ 好的

Paragraph →

✓ 网页内容翻译完成

显示原文

中英对照

关闭

Paragraph id

图4
分发的文字袋

为了达到更高精度的目标，在随机梯度下降的每一次迭代中，我们抽取一个文本窗口，并从该窗口中选取一些随机词。在这个过程的最后，基于给定的段落向量，我们将形成一个分类。分布式单词包模型概念简单，不需要存储单词向量，因此需要较少的内存。深度学习算法强大地从各种大数据中提取有用的表示，由深度学习提供的鉴别结果可用于信息检索。[96](#)].

递归神经网络

背后的想法递归神经网络输入数据不是相互独立的。了解以前迭代的数据将提高我们的预测精度。例如，考虑到我们希望在一系列单词中预测下一个单词。掌握前面的单词有助于我们提高预测的准确性。递归神经网络，通过考虑先前的计算，对序列中的每个元素执行相同的任务。换句话说，它有内存来捕获有关到目前为止所计算的内容的信息。但在实践中，消失梯度是深度学习中普遍存在的问题。由于消失梯度问题，RNNS只回顾几个步骤。虽然消失梯度并不是RNN独有的，但它们限制了我们的网络深度小于句子的长度。谢天谢地，有多种方法可以帮助我们解决消失梯度问题。例如，与其使用坦赫或者乙状结肠作为激活函数，我们可以使用relu。然而，我们为我们的工作选择了一个更受欢迎的解决方案-长期短期记忆(Lstm).

LSTM被提议[\[76\]](#)作者声明：Schmidhuber.RNNS和LSTM的主要区别是门控单元。与RNN相比，LSTMS中的门控单元有助于系统存储更多的信息。信息可以存储在单元格中、写入单元格或从单元格中读取。单元格通过打开和关闭门来决定是删除还是存储信息。单元由四个主要元素组成：输入门、具有自循环连接的神经元、遗忘门和输出门。遗忘门是一个元素，它允许单元格记住或忘记以前的状态。例如，假设我们想要捕捉对象的性别。在这种情况下，当看到一个新的主题时，应该忘记前一个主题，以便确定和存储相关信息。

卷积神经网络

最常用的深度学习模型之一是全连接神经网络.虽然完全连接的神经网络被认为是分类任务中的一个很好的解决方案，但这些网络中的大量连接可能会导致问题。这些问题可以在文本处理中进一步放大，因为需要大量的神经元。此外，我们认为，与任何句子中从未接近过的词相比，句子中相近的词更能相互联系。但是，完全连接的神经网络处理输入词时，

距离很远，就像句子中接近的单词一样。深度学习的分层学习过程使得像图像或文本这样窗口内准确度和精密度召回，F-测度，AUC量(大数据的特征之一)时，这种深度学习

我们使用cookie来个性化内容和广告，提供社交媒体功能，并分析我们的流量。我们亦会与我们的社交媒体、广告及分析合作伙伴分享有关阁下使用本网站的资料。隐私声明。您可以在ManageCookies中管理您的首选项。

☐ 管理饼干

✓ 好的

Hinton等人[29]使用深度学习和卷积神经网络。于现有的其他方法。Hinton的团队工作很有价的重要性。迪恩等人[84]使用类似的深度学习建模方法，但使用大规模的软件基础设施作为培训，并在[98]使用视频数据。它们使用层叠和卷积等深度学习方法来学习层次表示。

✓ 网页内容翻译完成

显示原文

中英对照

关闭

结果和讨论

在本节中，我们将解释我们在StockTwits数据集上应用深度学习方法的实验。我们试图观察深度学习模型是否能提高StockTwits信息情感分析的准确性。深度学习试图模仿人脑的分层学习方法。利用深度学习提取特征会给大数据分析带来非线性。下面介绍了三种常用的深度学习方法在自然语言处理中的应用结果。

Doc2vec

表5

doc2vec在StockTwits数据集上的性能

窗口	准确度	精密度	召回	F-测度	AUC
5	0.6202	0.6097	0.6682	0.6376	0.6202
10	0.6723	0.6687	0.6830	0.6757	0.6723

作为第一步，我们将doc2vec模型应用于StockTwits数据集，看看它是否能提高股市作者情绪预测的准确性。这被选择为第一个模型，因为它使用段落作为记忆来保持句子中单词的顺序，并将段落和单词映射到向量。

[例]7.3建议同时使用doc2vec架构来创建段落向量。按照他在实验中的方法，每个段落向量是两个向量的组合-一个是通过分布式内存体系结构(DM)学习的，另一个是用分布式单词包(DBOW)结构学习的。doc2vec模型的精度也可能受到窗口大小的影响；较大的窗口具有更高的精度。为了评估这一点，我们考虑了最常用的窗口大小-5和10。Python中的Gensim库用于实现doc2vec，所有总频率小于两个的单词都被忽略。结果如表所示5.

准确度 精密度的召回 F-测度 AUC 于窗口大小为10的窗口比窗口大小为5的窗口的精度要高。但两者之间的差异是可以忽略不计的。

我们使用cookie来个性化内容和广告，提供社交媒体功能，并分析我们的流量。我们亦会与我们的社交媒体、广告及分析合作伙伴分享有关阁下使用本网站的资料。隐私声明。您可以在ManageCookies中管理您的首选项。

☐ 管理饼干

✓ 好的

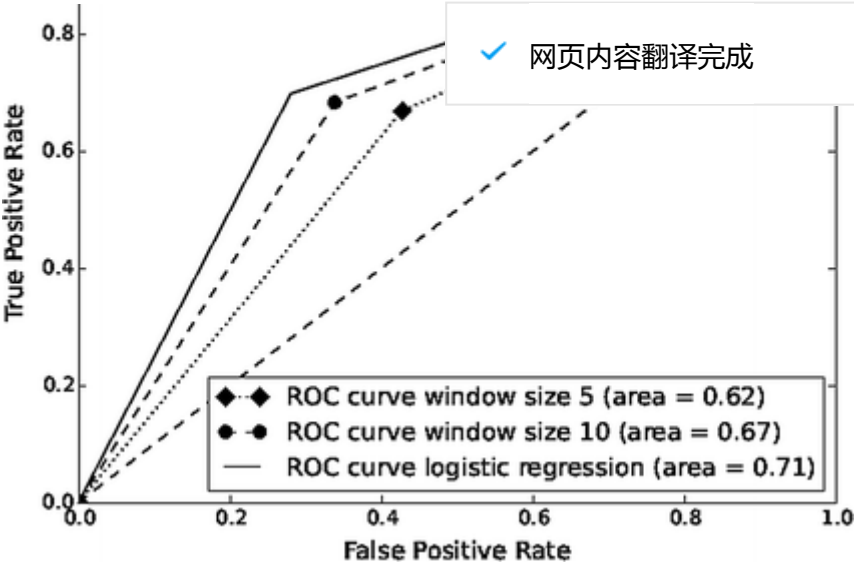


图5

窗口大小为5和10的doc2vec曲线下面积

表6

StockTwits数据集上LSTM的性能

准确度	精确度	召回	F-测度	AUC
0.6923	0.8518	0.6571	0.7419	0.7109

长期短期记忆

基于上一节的研究结果，doc2vec并不是一个很好的预测作者对股市情绪的模型，所以我们继续讨论RNNS。这些是在自然语言处理中使用的其他一些最流行的模型，显示了非常好的效果。采用RNNS，看看它们是否有助于提高StockTwits情绪分析的准确性。虽然我们的实验没有使用实际的RNN，但LSTM[99, 100, 101, 102]可能是一个可行的替代品，因为它具有更深的内存结构。

在我们的实施中，我们使用了Theano103]Python中的库。我们使用平均池作为我们的池方法。对于最后一步，我们将池的结果输入到逻辑回归层，以找到与当前输入序列相关联

我们使用cookie来个性化内容和广告，提供社交媒体功能，并分析我们的流量。我们亦会与我们的社交媒体、广告及分析合作伙伴分享有关阁下使用本网站的资料。隐私声明。您可以在ManageCookies中管理您的首选项。

☐ 管理饼干

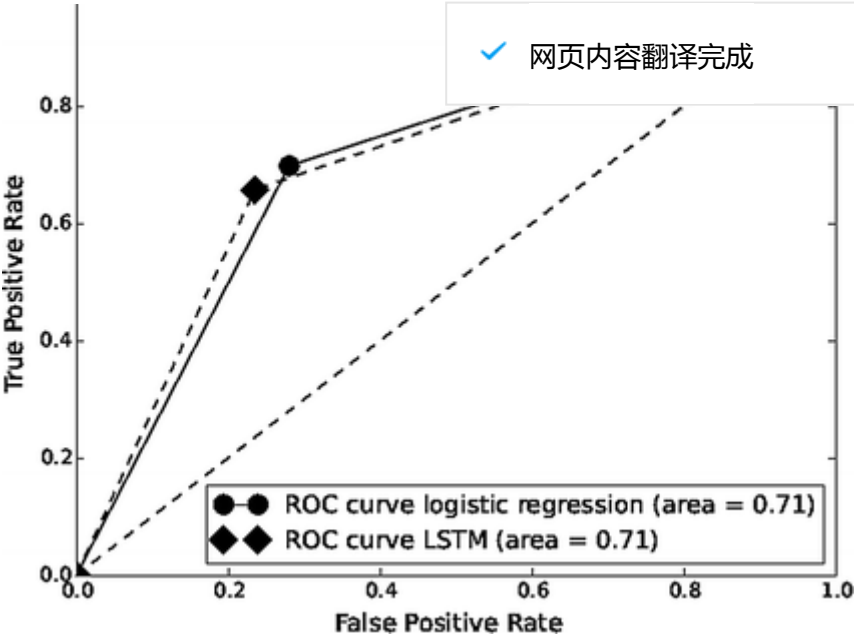


图6

短期记忆曲线下面积

表7

卷积神经网络在StockTwits数据集上的性能

步骤 准确度 精密度 召回 **F-测度** **AUC**

我们使用cookie来个性化内容和广告，提供社交媒体功能，并分析我们的流量。我们亦会与我们的社交媒体、广告及分析合作伙伴分享有关阁下使用本网站的资料。隐私声明。您可以在 **ManageCookies** 中管理您的首选项。

☐ 管理饼干

✓ 好的

6000 0.8651 0.8778 0.8484 0.8629

✓ 网页内容翻译完成

显示原文

中英对照

关闭

8000 0.8891 0.8774 0.9046 0.8908 0.8891

10,000 0.9093 0.9168 0.9004 0.9086 0.9093

70,000 0.9897 0.9909 0.9885 0.9897 0.9897

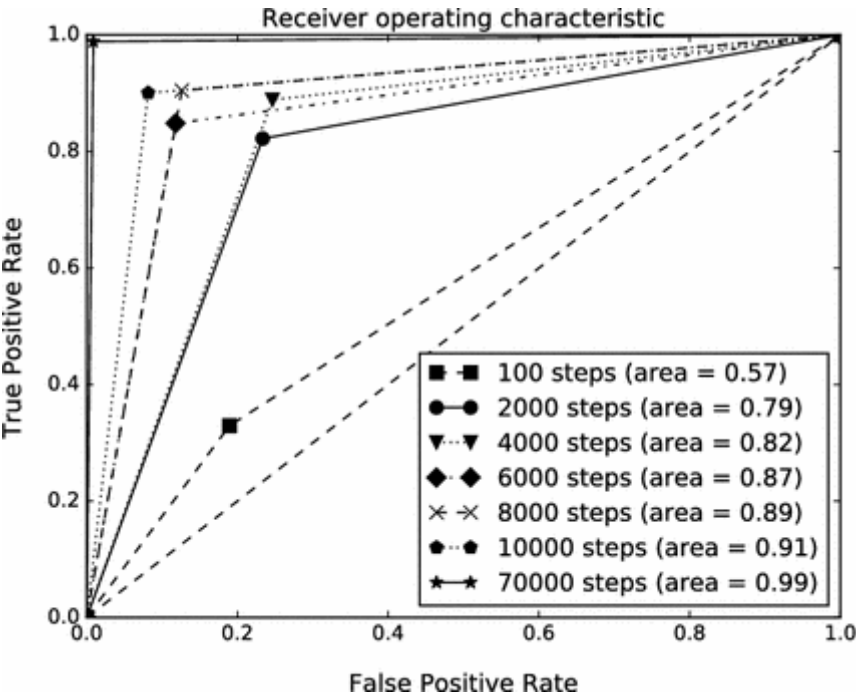


图7

卷积神经网络在ROC曲线下的比较面积

卷积神经网络

在LSTM被发现无效的情况下，我们求助于CNN。虽然CNN在图像处理中非常流行，但是能够找到大数据的内部结构使它成为我们理想的模型。我们雇用CNN，看看它是否可以用TensorFlow来改进我们的情绪分析任务。[104](#)]包在Python中。我们的第一步是将单词嵌入到低维向量中。

然后，模型输入的字向准确度行精密度召回的F-测度算AUC我们的实验中，我们使用了2.1和2.5的过滤器尺寸。然后，在卷积结果上引入最大池，并加入手动正则化。最后，使

我们使用cookie来个性化内容和广告，
提供社交媒体功能，并分析我们的流
量。我们亦会与我们的社交媒体、广告
及分析合作伙伴分享有关阁下使用本网
站的资料。隐私声明。您可以在
ManageCookies中管理您的首选项。

☐ 管理饼干

✓ 好的

LSTM和CNN(10,000步后)的结果。8。在此
据集中预测作者情绪的有效模型，因为它在所

✓ 网页内容翻译完成

显示原文

中英对照

关闭

表8

财务情绪分析中的深度学习模式比较

模型	准确度	精密度	召回	F-测度	AUC
Logistic回归	0.7088	0.7134	0.6980	0.7056	0.7088
Doc2vec	0.6723	0.6687	0.6830	0.6757	0.6723
LSTM	0.6923	0.8515	0.6571	0.7419	0.7109
CNN(10,000步)	0.9093	0.9168	0.9004	0.9086	0.9093

结论

深度学习在自然语言处理等领域具有良好的性能和前景。深度学习有机会解决大数据中的
数据分析和学习问题。与传统的数据挖掘方法相比，深度学习算法通过更多的层次来转换
输入信息。深度学习中的隐藏层通常用于提取特征或数据表示。在深度学习中，这种分层
学习过程提供了寻找单词语义和关系的机会。这些属性使得深度学习成为情感分析最理想
的模型之一。

本文在研究结果的基础上，证明了卷积神经网络可以克服股票情绪分析中的数据挖掘方
法。在文本分类的标准数据挖掘方法中，文档表示为字包向量。这些向量表示文档中出现的
单词，但不考虑句子中单词的顺序。很明显，在某些情况下，语序可以改变句子的情感。
解决这个问题一个办法是，除了使用单位克之外，还使用双克或n克。**86, 105, 106]**不
幸的是，在**(n>1)**无效**107]**使用cnn提供了这个机会，使用n-克来有效地提取文档的情
感。它通过卷积层从文档中存在的数据内部结构中受益，其中每个计算单元对输入数据
的一个小区域进行响应。我们以基于一袋袋单词的Logistic回归作为基线，并将深度学习
应用于Logistic回归的结果进行了比较。基于我们的结果，在情感分析中常用的深度学习方
法中，只有卷积神经网络优于Logistic回归。与其他模型相比，卷积神经网络的精度要好

得多。根据我们的研究结果，我们可以利用CNN从他们的言论中提取作者对股票的看法。在金融社会网络中，有一些人能够正确地预测股票市场。通过CNN预测他们的情绪，我们可以预测股票市场走势。

我们使用cookie来个性化内容和广告，提供社交媒体功能，并分析我们的流量。我们亦会与我们的社交媒体、广告及分析合作伙伴分享有关阁下使用本网站的资料。隐私声明。您可以在ManageCookies中管理您的首选项。

☐ 管理饼干

☒ 好的

SS是王安的作者，她做了所有的研究，并体系看涨和看跌的想法是DW的想法。我们咨询AP利用TMK的经验来改进我们的工作。所有作者读并批准了最后的手稿。

☒ 网页内容翻译完成

显示原文

中英对照

关闭

致谢

不适用。

竞争利益

提交人宣布，他们没有相互竞争的利益。

数据和材料的提供

不适用。

同意发表

不适用。

道德操守认可和参与的同意

不适用。

供资

不适用。

出版商注

斯普林格自然公司对已出版的地图和机构附属机构的管辖要求保持中立。

我们使用cookie来个性化内容和广告，提供社交媒体功能，并分析我们的流量。我们亦会与我们的社交媒体、广告及分析合作伙伴分享有关阁下使用本网站的资料。隐私声明。您可以在ManageCookies中管理您的首选项。

管理饼干

好的

title=Social%2Onetwork%20sites%3and%20scholarship&author=NB.%2diat%20Commun&volume=13&issue=230&publication_year=2007)

网页内容翻译完成

显示原文

中英对照

关闭

2.

王G, 王T, 王B, 桑巴斯文D, 张Z, 郑宏, 赵比. 华尔街的人群: 从社会投资平台、基础和检索的趋势中提取价值. 纽约: ACM; 2014年.

谷歌学者 (http://scholar.google.com/scholar_lookup?title=Crowds%20on%20wall%20street%3A%20extracting%20value%20from%20social%20investing%20platforms%2C%20foundations%20and%20trends%20in%20information%20retrieval&author=G.%20Wang&author=T.%20Wang&author=B.%20Wang&author=D.%20Sambasivan&author=Z.%20Zhang&author=H.%20Zheng&author=BY.%20Zhao&publication_year=2014)

3.

弗里德曼检察官. 统计模型: 理论与实践. 剑桥: 剑桥大学出版社; 2009年.

交叉参考 (<https://doi.org/10.1017/CBO9780511815867>)

数学 (<http://www.emis.de/MATH-item?1167.62001>)

谷歌学者 (http://scholar.google.com/scholar_lookup?title=Statistical%20models%3A%20theory%20and%20opractice&author=D.A.%20Freedman&publication_year=2009)

4.

Krizhevsky A, Sutskever I, Hinton GE. 基于深度卷积神经网络的ImageNet分类. 在: 神经信息处理系统方面的进展; 2012年. 1097-105

谷歌学者 (<https://scholar.google.com/scholar?q=Krizhevsky%20A%2C%20Sutskever%20I%2C%20Hinton%20GE.%20Imagenet%20classification%20with%20deep%20convolutional%20neural%20networks.%20In%3A%20Advances%20in%20neural%20information%20processing%20systems%3B%202012.%20p.%201097%E2%80%93105>)

5.

题名/责任者: Rev.用于释义检测的动态池和展开递归自动编码器. ADV神经处理系统2011年; 24: 801-9.

谷歌学者 (<https://scholar.google.com/scholar?q=Socher%20R%2C%20Huang%20EH%2C%20Pennin%20J%2C%20Manning%20CD%2C%20Ng%20AY.%20Dynamic%20pooling%20and%20unfolding%20recursive%20autoencoders%20for%20paraphrase%20detection.%20Adv%20Neural%20Inf%20Process%20Syst.%202011%3B24%3A801%E2%80%93939>)

6.

皮尔逊K.关于父母双方的回归和继承的注记. PROC R Soc Lond.1895年; 58: 240-2.

交叉参考 (<https://doi.org/10.1098/rspl.1895.0041>)

谷歌学者 (http://scholar.google.com/scholar_lookup?title=Notes%20on%20regression%20and%20inheritance%20in%20the%20case%20of%20two%20parents&author=K.%20Pearson&journal=Proc%20R%20Soc%20Lond&volume=58&pages=240-242&publication_year=1895)

7. Graves A, Mohamed AR, Hinton G.基于深度递归神经网络的语音识别。在：2013年IEEE声学、语音和信号处理国际会议(ICASSP)。2013年。

我们使用cookie来个性化内容和广告，提供社交媒体功能，并分析我们的流量。我们亦会与我们的社交媒体、广告及分析合作伙伴分享有关阁下使用本网站的资料。隐私声明。您可以在ManageCookies中管理您的首选项。

☐ 管理饼干

✓ 好的

网页内容翻译完成

显示原文

中英对照

关闭

title=Phone%20recognition%20with%20restricted%20Boltzmann%20machines&author=A.%20Graves&author=AR.%20Mohamed&author=GE.%20Hinton&journal=Adv%20Neural%20Information%20Processing%20Systems&volume=23&pages=469-477&publication_year=2010)

9. George E, Yu D, 邓L, Acero A.大词汇量语音识别的依赖上下文的深层神经网络.IEEE跨音频语音处理。2012年； 20(1): 30-42。
交叉参考 (https://doi.org/10.1109/TASL.2011.2134090)
谷歌学者 (http://scholar.google.com/scholar_lookup?title=Context-dependent%20pre-trained%20deep%20neural%20networks%20for%20large-vocabulary%20speech%20recognition&author=E.%20George&author=D.%20Yu&author=L.%20Deng&author=A.%20Acero&journal=IEEE%20Transactions%20on%20Audio%20Speech%20Language%20Processing&volume=20&issue=1&pages=30-42&publication_year=2012)

10. Seide F, Li G, Yu D.会话式语音转录，使用上下文依赖的深层神经网络.国际言语传播协会第十二届年会； 2011年。
谷歌学者 (https://scholar.google.com/scholar?q=Seide%20F%20Li%20G%20Yu%20D.%20Conversational%20speech%20transcription%20using%20context-dependent%20deep%20neural%20networks.%20In%3A%20Twelfth%20annual%20conference%20of%20the%20international%20speech%20communication%20association%3B%202011.)

11. 作者声明: Mohamed A, Dahl GE, Hinton G.IEEE跨音频语音处理。2012年； 20(1): 14-22。
交叉参考 (https://doi.org/10.1109/TASL.2011.2109382)
谷歌学者 (http://scholar.google.com/scholar_lookup?title=Acoustic%20modeling%20using%20deep%20belief%20networks&author=A.%20Mohamed&author=GE.%20Dahl&author=G.%20Hinton&journal=IEEE%20Transactions%20on%20Audio%20Speech%20Language%20Processing&volume=20&issue=1&pages=14-22&publication_year=2012)

12. Itamar A, Rose DC, Karnoski TP.深入机器学习-人工智能研究[研究前沿]的一个新前沿。IEEE Comput INTER MAG。2010年； 5(4): 13-8。
交叉参考 (https://doi.org/10.1109/MCI.2010.938364)
谷歌学者 (http://scholar.google.com/scholar_lookup?title=Deep%20machine%20learning%E2%80%94a%20new%20frontier%20in%20artificial%20intelligence%20research%20%5Bresearch%20frontier%5D&author=A.%20Itamar&author=DC.%20Rose&author=TP.%20Karnowski&journal=IEEE%20Comput%20Intell%20Mag&volume=5&issue=4&pages=13-18&publication_year=2010)

13. Najafabadi NM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, ...
IEEE.1998年; 86(11): 2278-324。

我们使用cookie来个性化内容和广告，
提供社交媒体功能，并分析我们的流
量。我们亦会与我们的社交媒体、广告
及分析合作伙伴分享有关阁下使用本网
站的资料。隐私声明。您可以在
ManageCookies中管理您的首选项。

☐ 管理饼干

✓ 好的

IEEE.1998年; 86(11): 2278-324。
交叉参考 (https://doi.org/10.1109/
谷歌学者 (http://scholar.google.com/scholar_lookup?title=Gradient-
based%20learning%20applied%20to%20document%20recognition&author
=Y.%20LeCun&author=L.%20Bottou&author=Y.%20Bengio&author=P.%2
0Haffner&journal=Proc%20IEEE&volume=86&issue=11&pages=2278-
2324&publication_year=1998)

✓ 网页内容翻译完成

显示原文

中英对照

关闭

15. 作者声明: [by]Colobert R, Weston J, Bottou L, Karlen M, Kavukcuoglu
K, Kuksa P.自然语言处理(几乎)从零开始。j马赫学习研究报告, 2011年; 12:
2493-537。
数学 (http://www.emis.de/MATH-item?1280.68161)
谷歌学者 (http://scholar.google.com/scholar_lookup?
title=Natural%20language%20processing%20%28almost%29%20from%20s
cratch&author=R.%20Collobert&author=J.%20Weston&author=L.%20Bott
ou&author=M.%20Karlen&author=K.%20Kavukcuoglu&author=P.%20Kuks
a&journal=J%20Mach%20Learn%20Res&volume=12&pages=2493-
2537&publication_year=2011)

16. Colobert R, Weston J.自然语言处理的统一架构: 具有多任务学习的深层神经网络。第25届国际机器学习会议记录。伦敦: ACM; 2008年。160-7
谷歌学者 (https://scholar.google.com/scholar?
q=Collobert%20R%2C%20Weston%20J.%20A%20unified%20architecture%
20for%20natural%20language%20processing%3A%20deep%20neural%20ne
tworks%20with%20multitask%20learning.%20In%3A%20Proceedings%20
of%20the%2025th%20international%20conference%20on%20machine%20
learning.%20London%3A%20ACM%3B%202008.%20p.%20160%E2%80%99
37.)

17. 高杰, 邓L, 伽蒙, 何X, 潘特尔P.深层神经网络建模兴趣。2014年。美国专利应
用程序14/304 863。
谷歌学者 (https://scholar.google.com/scholar?
q=Gao%20J%2C%20Deng%20L%2C%20Gamon%20M%2C%20He%20X%2C
%20Pantel%20P.%20Modeling%20interestingness%20with%20deep%20ne
ural%20networks.%202014.%20US%20Patent%20App.%2014%2F304%2C
863.)

18. Kalchbrenner N, Grephenstette E, Blunsorp.一个用于句子建模的卷积神经网络.
arxiv预印arXiv:1404.2188 (http://arxiv.org/abs/1404.2188). 2014.

19. 句子分类的卷积神经网络arXiv:1408.5882
(http://arxiv.org/abs/1408.5882). 2014.

20. 沈勇, 何X, 高杰, 邓L, 梅斯尼尔.一种潜在语义模型, 具有卷积池结构, 用于
信息检索。在: 第23届ACM信息和知识管理国际会议记录。纽约: ACM; 2014
年。101-10.

我们使用cookie来个性化内容和广告，提供社交媒体功能，并分析我们的流量。我们亦会与我们的社交媒体、广告和分析合作伙伴分享有关阁下使用本网站的资料。隐私声明。您可以在[Manage Cookies](#)中管理您的首选项。

✓ 好的

关闭

- 24/38

1'0 0/ 1 . 0/ 8 0/ 10/ 111110/ 1 0/ 1 0/ . 1 1 0/

✓ 好的

- 关闭

32. 陆T, 索契R, 曼宁CD. 使用递归神经网络对词法进行更好的表示。温哥华: CoNLL; 2003年。151-156。

我们使用cookie来个性化内容和广告，提供社交媒体功能，并分析我们的流量。我们亦会与我们的社交媒体、广告及分析合作伙伴分享有关阁下使用本网站的资料。隐私声明。您可以在ManageCookies中管理您的首选项。

☐ 管理饼干

✓ 好的

gically%20complex%20words%20in%20the%20context%20of%20sentiment%20analysis%20or=A.%20Lazaridou&author=M.%20Baroni&journal=ACL&volume=1&pages=151-156&publication_year=2013)

✓ 网页内容翻译完成

显示原文

中英对照

关闭

34. 基尔加里夫A, 格里芬斯特特G.介绍特别问题在网络上作为语料库。康普特·林吉斯。2003年; 29(3): 333-47。
MathSciNet (<http://www.ams.org/mathscinet-getitem?mr=2113093>)
交叉参考 (<https://doi.org/10.1162/089120103322711569>)
谷歌学者 (http://scholar.google.com/scholar_lookup?title=Introduction%20to%20the%20special%20issue%20on%20the%20web%20as%20corpus&author=A.%20Kilgarriff&author=G.%20Grefenstette&journal=Comput%20Linguis&volume=29&issue=3&pages=333-347&publication_year=2003)
35. 题名/责任者: Rev.用于预测情绪分布的半监督递归自动编码器。关于自然语言处理经验方法的会议记录。计算语言学协会; 2011年。151-61。
谷歌学者 (<https://scholar.google.com/scholar?q=Socher%20R%20Pennington%20J%20Huang%20EH%20Ng%20AY%20Manning%20CD.%20Semi-supervised%20recursive%20autoencoders%20for%20predicting%20sentiment%20distributions.%20In%3A%20Proceedings%20of%20the%20conference%20on%20empirical%20methods%20in%20natural%20language%20processing.%20Association%20for%20computational%20linguistics%3B%202011.%20p.%20151%E2%80%99361>)
36. Hinton GE, Salakhutdinov RR.用神经网络降低数据维数。科学。2006年; 313(5786): 504-7。
MathSciNet (<http://www.ams.org/mathscinet-getitem?mr=2242509>)
交叉参考 (<https://doi.org/10.1126/science.1127647>)
数学 (<http://www.emis.de/MATH-item?1226.68083>)
谷歌学者 (http://scholar.google.com/scholar_lookup?title=Reducing%20the%20dimensionality%20of%20data%20with%20neural%20networks&author=GE.%20Hinton&author=RR.%20Salakhutdinov&journal=Science&volume=313&issue=5786&pages=504-507&publication_year=2006)
37. 辛顿·葛, 泽梅尔·RS. 自动编码器, 最小描述长度和Helmholtz自由能。ADV神经网络信息处理系统1994年: 3-10。
谷歌学者 (<https://scholar.google.com/scholar?q=Hinton%20GE%20Zemel%20RS.%20Autoencoders%2C%20minimum%20description%20length%20and%20helmholtz%20free%20energy.%20Adv%20Neural%20Inform%20Process%20Syst.%201994%3A3%E2%80%9910>)

38. 动态系统中的信息处理：和谐理论的基础。科罗拉多大学博尔德计算机科学系技术报告；2006年。

我们使用cookie来个性化内容和广告，提供社交媒体功能，并分析我们的流量。我们亦会与我们的社交媒体、广告及分析合作伙伴分享有关阁下使用本网站的资料。隐私声明。您可以在ManageCookies中管理您的首选项。

☐ 管理饼干

✓ 好的

数学 (http://www.emis.de/MATH-i) 谷歌学者 (http://scholar.google.com) ☒ 网页内容翻译完成 显示原文 中英对照 关闭

40. Socher R, Huval B, Manning CD, Ng AY.通过递归矩阵向量空间实现语义合成。2012年自然语言处理和计算自然语言学习经验方法联席会议记录。计算语言学协会；2012年。1201-11。
谷歌学者 (https://scholar.google.com/scholar?q=Socher%20R%2C%20Huval%20B%2C%20Manning%20CD%2C%20Ng%20AY.%20Semantic%20compositionality%20through%20recursive%20matrix-vector%20spaces.%20In%3A%20Proceedings%20of%20the%202012%20joint%20conference%20on%20empirical%20methods%20in%20natural%20language%20processing%20and%20computational%20natural%20language%20learning.%20Association%20for%20computational%20linguistics%3B%202012.%20p.%201201%E2%80%9311.)

41. 书名/作者责任者：by L.用组合向量语法进行分析。ACL.2013年；1：455-65。
谷歌学者 (http://scholar.google.com/scholar_lookup?title=Parsing%20with%20compositional%20vector%20grammars&author=R.%20Socher&author=J.%20Bauer&author=CD.%20Manning&author=AY.%20Ng&journal=ACL&volume=1&pages=455-465&publication_year=2013)

42. Socher R, Lin CC, Manning C, Ng AY.用递归神经网络解析自然场景和自然语言。第28届国际机器学习会议记录(ICML-11)。2011年p.129-36。
谷歌学者 (https://scholar.google.com/scholar?q=Socher%20R%2C%20Lin%20CC%2C%20Manning%20C%2C%20Ng%20AY.%20Parsing%20natural%20scenes%20and%20natural%20language%20with%20recursive%20neural%20networks.%20In%3A%20Proceedings%20of%20the%2028th%20international%20conference%20on%20machine%20learning%20%28ICML-11%29.%202011.%20p.%20129%E2%80%9336.)

43. Colobert R.深度学习，用于高效的鉴别分析。在：第十四届人工智能和统计国际会议记录。2011年p.224-32。
谷歌学者 (https://scholar.google.com/scholar?q=Collobert%20R.%20Deep%20learning%20for%20efficient%20discriminative%20parsing.%20In%3A%20Proceedings%20of%20the%20fourteenth%20international%20conference%20on%20artificial%20intelligence%20and%20statistics.%202011.%20p.%20224%E2%80%9332.)

44. 市场情绪。http://www.investopedia.com/ (http://www.investopedia.com/).

45. 彭B, 李L.意见挖掘和情感分析.发现趋势在检索.2008年; 2: 1-35。
交叉参考 (https://doi.org/10.1186/s40537-017-0111-6)

我们使用cookie来个性化内容和广告，
提供社交媒体功能，并分析我们的流
量。我们亦会与我们的社交媒体、广告
及分析合作伙伴分享有关阁下使用本网
站的资料。隐私声明。您可以在
ManageCookies中管理您的首选项。

☐ 管理饼干

✓ 好的

title=When%20is%20a%20liability%20analysis%2C%20dictionaries&author=J.%20McDonald&journal=J%20Finance&volume=66&pages=35-65&publication_year=2011)

✓ 网页内容翻译完成

显示原文

中英对照

关闭

47. 毛赫, 高P, 王英, 博伦J.从大型中文新闻语料库自动构建金融语义导向词汇。
in: 第七届金融风险国际论坛; 2014年。
谷歌学者 (https://scholar.google.com/scholar?q=Mao%20H%2C%20Gao%20P%2C%20Wang%20Y%2C%20Bollen%20J.%20Automatic%20construction%20of%20financial%20semantic%20orientation%20lexicon%20from%20large%20scale%20Chinese%20news%20corpus.%20In%3A%207th%20Financial%20risks%20international%20forum%3B%202014.)

48. SteinwartI, Christmann A.支持向量机。柏林: 斯普林格; 2008年。
数学 (http://www.emis.de/MATH-item?1203.68171)
谷歌学者 (http://scholar.google.com/scholar_lookup?title=Support%20vector%20machine&author=I.%20Steinwart&author=A.%20Christmann&publication_year=2008)

49. 赛义夫H, 何Y, 阿拉尼H.语义网-ISWC 2012。柏林: 斯普林格; 2012年。508-24
交叉参考 (https://doi.org/10.1007/978-3-642-35176-1_32)
谷歌学者 (http://scholar.google.com/scholar_lookup?title=Semantic%20sentiment%20analysis%20of%20Twitter.%20The%20semantic%20Web-ISWC%202012&author=H.%20Saif&author=Y.%20He&author=H.%20Alani&publication_year=2012)

50. SILVA N, Hruschka E, Hruschka E.DEIS支持系统2014年; 66: 170-9。
交叉参考 (https://doi.org/10.1016/j.dss.2014.07.003)
谷歌学者 (http://scholar.google.com/scholar_lookup?title=Tweet%20sentiment%20analysis%20with%20classifier%20ensembles&author=N.%20Silva&author=E.%20Hruschka&author=E.%20Hruschka&journal=Decis%20Support%20Syst&volume=66&pages=170-179&publication_year=2014)

51. 费西尼E, 梅西纳E, 波齐足协。基于大型中文新闻语料库的金融语义导向词典的自动构建。DEIS支持系统2014年; 68: 26-38。
交叉参考 (https://doi.org/10.1016/j.dss.2014.10.004)
谷歌学者 (http://scholar.google.com/scholar_lookup?title=Automatic%20construction%20of%20financial%20semantic%20orientation%20lexicon%20from%20large%20scale%20Chinese%20news%20corpus&author=E.%20Fersini&author=E.%20Messina&author=FA.%20Pozzi&journal=Decis%20Support%20Syst&volume=68&pages=26-38&publication_year=2014)

ournal=Decis%20Support%20Syst&volume=68&pages=26-38&publication_year=2014)

我们使用cookie来个性化内容和广告，提供社交媒体功能，并分析我们的流量。我们亦会与我们的社交媒体、广告及分析合作伙伴分享有关阁下使用本网站的资料。隐私声明。您可以在 **ManageCookies** 中管理您的首选项。

☐ 管理饼干

✓ 好的

53. Salton G, Buckley C. 自动文本检索系统。1988年；24(5)：513-23。

交叉参考 (<https://doi.org/10.1016/>

谷歌学者 (http://scholar.google.com/scholar_lookup?title=Term-weighting%20approaches%20in%20automatic%20text%20retrieval&author=G.%20Salton&author=C.%20Buckley&journal=Inf%20Process%20Manag&volume=24&issue=5&pages=513-523&publication_year=1988)

✓ 网页内容翻译完成

显示原文

中英对照

关闭

54. Robertson SE, Walker S. 关于概率加权检索的2-Poisson模型的一些简单有效逼近。在：第17届国际ACM SIGIR年度信息检索研究与开发会议记录。纽约：斯普林格公司；1994年。232-41。

谷歌学者 (<https://scholar.google.com/scholar?q=Robertson%20SE%20C%20Walker%20S.%20Some%20simple%20effective%20approximations%20to%20the%202-Poisson%20model%20for%20probabilistic%20weighted%20retrieval.%20In%3A%20Proceedings%20of%20the%2017th%20annual%20international%20ACM%20SIGIR%20conference%20on%20research%20and%20development%20in%20information%20retrieval.%20New%20York%3A%20Springer%20Inc.%3B%201994.%20p.%20232%E2%80%99341>)

q=Robertson%20SE%20C%20Walker%20S.%20Some%20simple%20effective%20approximations%20to%20the%202-Poisson%20model%20for%20probabilistic%20weighted%20retrieval.%20In%3A%20Proceedings%20of%20the%2017th%20annual%20international%20ACM%20SIGIR%20conference%20on%20research%20and%20development%20in%20information%20retrieval.%20New%20York%3A%20Springer%20Inc.%3B%201994.%20p.%20232%E2%80%99341.)

55. Bengio Y, 等人学习人工智能的深层架构。发现趋势马赫学习。2009年；2(1)：1-127。

MathSciNet (<http://www.ams.org/mathscinet-getitem?mr=2480723>)

交叉参考 (<https://doi.org/10.1561/2200000006>)

数学 (<http://www.emis.de/MATH-item?1192.68503>)

谷歌学者 (http://scholar.google.com/scholar_lookup?title=Learning%20deep%20architectures%20for%20AI&author=Y.%20Bengio&journal=Found%20Trends%20Mach%20Learn&volume=2&issue=1&pages=1-127&publication_year=2009)

56. 陈赫, 蒋RHL, 斯托里·VC. 商业情报和分析：从大数据到大影响。错误夸脱。2012年；36：4。

谷歌学者 (http://scholar.google.com/scholar_lookup?title=Business%20intelligence%20and%20analytics%3A%20from%20big%20data%20to%20big%20impact&author=H.%20Chen&author=RHL.%20Chiang&author=VC.%20Storey&journal=MIS%20Quart&volume=36&pages=4&publication_year=2012)

title=Business%20intelligence%20and%20analytics%3A%20from%20big%20data%20to%20big%20impact&author=H.%20Chen&author=RHL.%20Chiang&author=VC.%20Storey&journal=MIS%20Quart&volume=36&pages=4&publication_year=2012)

57. Dal N, Triggs B. 人类检测的定向梯度直方图。在：IEEE计算机学会计算机视觉和模式识别会议，CVPR 2005，第一卷。1.2005年p.886-93

谷歌学者 (<https://scholar.google.com/scholar?q=Dalal%20N%20C%20Triggs%20B.%20Histograms%20of%20oriented%20gradients%20for%20human%20detection.%20In%3A%20IEEE%20computer%20society%20conference%20on%20computer%20vision%20and%20patt>

q=Dalal%20N%20C%20Triggs%20B.%20Histograms%20of%20oriented%20gradients%20for%20human%20detection.%20In%3A%20IEEE%20computer%20society%20conference%20on%20computer%20vision%20and%20patt

ern%20recognition%2C%20CVPR%202005%2C%20vol.%201.%202005.%20p.%20886%E2%80%939393.)

我们使用cookie来个性化内容和广告，提供社交媒体功能，并分析我们的流量。我们亦会与我们的社交媒体、广告及分析合作伙伴分享有关阁下使用本网站的资料。隐私声明。您可以在ManageCookies中管理您的首选项。

☐ 管理饼干

✓ 好的

60. 科茨A, 吴安雅. 稀疏编码和矢量量化. [arXiv:1409.0473](http://arxiv.org/abs/1409.0473) (http://arxiv.org/abs/1409.0473). 2014年9月. **网页内容翻译完成** [显示原文](#) [中英对照](#) [关闭](#)
- 会议记录(ICML-11)。2011年p.921-8。
 [谷歌学者 \(https://scholar.google.com/scholar?q=Coates%20A%2C%20Ng%20AY.%20The%20importance%20of%20encoding%20versus%20training%20with%20sparse%20coding%20and%20vector%20quantization.%20In%3A%20Proceedings%20of%20the%2028th%20international%20conference%20on%20machine%20learning%20%28ICML-11%29.%202011.%20p.%20921%E2%80%93938.\)](https://scholar.google.com/scholar?q=Coates%20A%2C%20Ng%20AY.%20The%20importance%20of%20encoding%20versus%20training%20with%20sparse%20coding%20and%20vector%20quantization.%20In%3A%20Proceedings%20of%20the%2028th%20international%20conference%20on%20machine%20learning%20%28ICML-11%29.%202011.%20p.%20921%E2%80%93938.)
61. Hinton GE, Sri瓦斯塔瓦N, Krizhevsky A, Sutskever I, Salakhutdinov RR. 通过防止特征检测器的协同适应来改进神经网络。arXiv:1207.0580 (http://arxiv.org/abs/1207.0580). 2012.
62. 范杰, 韩飞, 刘海. 大数据分析的挑战。Natl Sci Rev.2014; 1(2): 293-314。
[交叉参考](https://doi.org/10.1093/nsr/nwt032) (https://doi.org/10.1093/nsr/nwt032)
 [谷歌学者 \(http://scholar.google.com/scholar_lookup?title=Challenges%20of%20big%20data%20analysis&author=J.%20Fan&author=F.%20Han&author=H.%20Liu&journal=Natl%20Sci%20Rev&volume=1&issue=2&pages=293-314&publication_year=2014\)](http://scholar.google.com/scholar_lookup?title=Challenges%20of%20big%20data%20analysis&author=J.%20Fan&author=F.%20Han&author=H.%20Liu&journal=Natl%20Sci%20Rev&volume=1&issue=2&pages=293-314&publication_year=2014)
63. Abadi M, Agarwal A, Barham P, Brevdo E. 竖起大拇指还是竖起大拇指? 语义导向应用于无监督的评论分类。POC Assoc Comput Linguis. 2002年; 66: 417-24。
 [谷歌学者 \(http://scholar.google.com/scholar_lookup?title=Thumbs%20up%20or%20thumbs%20down%3F%20Semantic%20orientation%20applied%20to%20unsupervised%20classification%20of%20reviews&author=M.%20Abadi&author=A.%20Agarwal&author=P.%20Barham&author=E.%20Brevdo&journal=Proc%20Assoc%20Comput%20Linguis&volume=66&pages=417-424&publication_year=2002\)](http://scholar.google.com/scholar_lookup?title=Thumbs%20up%20or%20thumbs%20down%3F%20Semantic%20orientation%20applied%20to%20unsupervised%20classification%20of%20reviews&author=M.%20Abadi&author=A.%20Agarwal&author=P.%20Barham&author=E.%20Brevdo&journal=Proc%20Assoc%20Comput%20Linguis&volume=66&pages=417-424&publication_year=2002)
64. 庞B, 李L, 维蒂亚纳坦S. 竖起大拇指? 利用机器学习技术进行情感分类。在: 自然语言处理经验方法会议记录, 第二卷。66; 2002。79-86。
 [谷歌学者 \(https://scholar.google.com/scholar?q=Pang%20B%2C%20Lee%20L%2C%20Vaithyanathan%20S.%20Thumbs%20up%3F%20Sentiment%20classification%20using%20machine%20learning%20techniques.%20In%3A%20Proceedings%20of%20the%20conference%20on%20empirical%20methods%20in%20natural%20language%20processing%2C%20vol.%2066%3B%202002.%20p.%2079%E2%80%9386.\)](https://scholar.google.com/scholar?q=Pang%20B%2C%20Lee%20L%2C%20Vaithyanathan%20S.%20Thumbs%20up%3F%20Sentiment%20classification%20using%20machine%20learning%20techniques.%20In%3A%20Proceedings%20of%20the%20conference%20on%20empirical%20methods%20in%20natural%20language%20processing%2C%20vol.%2066%3B%202002.%20p.%2079%E2%80%9386.)
65. 基利琴科S, 朱X, 穆罕默德S. 情绪分析简短的非正式文本。j Artif Intel Res.2014; 50: 723-62。
 [谷歌学者 \(http://scholar.google.com/scholar_lookup?title=Sentiment%20analysis%20of%20short%20informal%20texts&author=S.%20Kiritchenko&author=X.%20Zhu&author=S.%20Mohammad&journal=j%20Artif%20Intel%20Res.\)](http://scholar.google.com/scholar_lookup?title=Sentiment%20analysis%20of%20short%20informal%20texts&author=S.%20Kiritchenko&author=X.%20Zhu&author=S.%20Mohammad&journal=j%20Artif%20Intel%20Res.)

=J%20Artif%20Intell%20Res&volume=50&pages=723-762&publication_year=2014)

我们使用cookie来个性化内容和广告，提供社交媒体功能，并分析我们的流量。我们亦会与我们的社交媒体、广告及分析合作伙伴分享有关阁下使用本网站的资料。隐私声明。您可以在ManageCookies中管理您的首选项。

☐ 管理饼干

✓ 好的

67. 伯明翰ML, Pong-wong R, Spiliopoulou Campbell H, Wright AF, Wilson JF
特征选择在人类基因组预测中的应用。

✓ 网页内容翻译完成

显示原文

中英对照

关闭

交叉参考 (<https://doi.org/10.1038/srep10312>)

谷歌学者 ([http://scholar.google.com/scholar_lookup?](http://scholar.google.com/scholar_lookup?title=Application%20of%20high-dimensional%20feature%20selection%3A%20evaluation%20for%20genomic%20prediction%20in%20man&author=ML.%20Birmingham&author=R.%20Pong-Wong&author=A.%20Spiliopoulou&author=C.%20Hayward&author=I.%20Rudan&author=H.%20Campbell&author=AF.%20Wright&author=JF.%20Wilson&author=F.%20Agakov&author=P.%20Navarro&author=CS.%20Haley&journal=Sci%20Rep&volume=5&pages=10312&publication_year=2015)

title=Application%20of%20high-dimensional%20feature%20selection%3A%20evaluation%20for%20genomic%20prediction%20in%20man&author=ML.%20Birmingham&author=R.%20Pong-

Wong&author=A.%20Spiliopoulou&author=C.%20Hayward&author=I.%20Rudan&author=H.%20Campbell&author=AF.%20Wright&author=JF.%20Wilson&author=F.%20Agakov&author=P.%20Navarro&author=CS.%20Haley&journal=Sci%20Rep&volume=5&pages=10312&publication_year=2015)

68. Torgo L.与R.Boca Raton的数据挖掘: CRC出版社; 2010。

交叉参考 (<https://doi.org/10.1201/b10328>)

谷歌学者 ([http://scholar.google.com/scholar_lookup?](http://scholar.google.com/scholar_lookup?title=Data%20mining%20with%20R&author=L.%20Torgo&publication_year=2010)

title=Data%20mining%20with%20R&author=L.%20Torgo&publication_year=2010)

69. 生物信息学特征选择技术综述。生物信息学。2007年; 23(19): 2507-17。

交叉参考 (<https://doi.org/10.1093/bioinformatics/btm344>)

谷歌学者 ([http://scholar.google.com/scholar_lookup?](http://scholar.google.com/scholar_lookup?title=A%20review%20of%20feature%20selection%20techniques%20in%20bioinformatics&author=Y.%20Saey&author=I.%20Inza&author=P.%20Larra%20B1aga&journal=Bioinformatics&volume=23&issue=19&pages=2507-2517&publication_year=2007)

title=A%20review%20of%20feature%20selection%20techniques%20in%20bioinformatics&author=Y.%20Saey&author=I.%20Inza&author=P.%20Larra%20B1aga&journal=Bioinformatics&volume=23&issue=19&pages=2507-2517&publication_year=2007)

70. Pearson K.X.关于在变量关联系统情况下, 给定的偏离可能系统的判据, 可以合理地假定它是由随机抽样产生的。题名/作者Rond爱丁堡都柏林菲罗斯·梅格·斯奇。

1900年; 50: 157-75

交叉参考 (<https://doi.org/10.1080/14786440009463897>)

数学 (<http://www.emis.de/MATH-item?31.0238.04>)

谷歌学者 ([http://scholar.google.com/scholar_lookup?](http://scholar.google.com/scholar_lookup?title=X.%20On%20the%20criterion%20that%20a%20given%20system%20of%20deviations%20from%20the%20probable%20in%20the%20case%20of%20a%20corsystem%20of%20variables%20is%20such%20that%20it%20can%20be%20reasonably%20supposed%20to%20have%20arisen%20from%20random%20sampling&author=K.%20Pearson&journal=Lond%20Edinbu)

title=X.%20On%20the%20criterion%20that%20a%20given%20system%20of%20deviations%20from%20the%20probable%20in%20the%20case%20of%20a%20corsystem%20of%20variables%20is%20such%20that%20it%20can%20be%20reasonably%20supposed%20to%20have%20arisen%20from%20random%20sampling&author=K.%20Pearson&journal=Lond%20Edinbu

rg%20Dublin%20Philos%20Mag%20J%20Sci&volume=50&pages=157-175&publication_year=1900)

71. 在球体上分散。PROC R Soc Lond.1953年; 217: 295-305。

MathSciNet (<http://www.ams.org/mathscinet-getitem?mr=56866>)

交叉参考 (<https://doi.org/10.1098/rspa.1953.0064>)

数学 (<http://www.emis.de/MATH-item?0051.37105>)
谷歌学者 ([http://scholar.google.com/scholar_lookup?](http://scholar.google.com/scholar_lookup?title=Efficient%20estimation%20of%20word%20representations%20in%20vector%20space.%20In%3A%20Workshop%20at%20ICLR.%202013.)

我们使用cookie来个性化内容和广告，
提供社交媒体功能，并分析我们的流
量。我们亦会与我们的社交媒体、广告
及分析合作伙伴分享有关阁下使用本网
站的资料。隐私声明。您可以在
ManageCookies中管理您的首选项。

☐ 管理饼干

✓ 好的

q=Le%20Q%20C%20MIKOIOV%20I.%20f%20sentences%20and%20document%20inference%20on%20machine%20learning%20tasks.%20In%3A%20Workshop%20at%20ICLR.%202013.)

✓ 网页内容翻译完成

显示原文

中英对照

关闭

74. Mikolov T, Chen K, Corrado G, Dean J.向量空间中字表示的有效估计。in: 在大湖区问题国际会议上举办的讲习班。2013年。
谷歌学者 (<https://scholar.google.com/scholar?q=Mikolov%20T%2C%20Chen%20K%2C%20Corrado%20G%2C%20Dean%20J.%20Efficient%20estimation%20of%20word%20representations%20in%20vector%20space.%20In%3A%20Workshop%20at%20ICLR.%202013.>)

75. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. in: 在大湖区问题国际会议上举办的讲习班。2013年。
谷歌学者 (<https://scholar.google.com/scholar?q=Mikolov%20T%2C%20Sutskever%20I%2C%20Chen%20K%2C%20Corrado%20GS%2C%20Dean%20J.%20Distributed%20representations%20of%20words%20and%20phrases%20and%20their%20compositionality.%20In%3A%20Workshop%20at%20ICLR.%202013.>)

76. 题名/责任者: Reach.神经康普特。1997年; 9: 1735-80
交叉参考 (<https://doi.org/10.1162/neco.1997.9.8.1735>)
谷歌学者 (http://scholar.google.com/scholar_lookup?title=Long%20short-term%20memory&author=S.%20Hochreiter&author=J.%20Schmidhuber&journal=Neural%20Comput&volume=9&pages=1735-1780&publication_year=1997)

77. Cellapilla K, Puri S, Simard P.高性能卷积神经网络用于文档处理。第十届笔迹识别前沿国际研讨会。西雅图: Suvisoft; 2006年。
谷歌学者 (<https://scholar.google.com/scholar?q=Chellapilla%20K%2C%20Puri%20S%2C%20Simard%20P.%20High%20performance%20convolutional%20neural%20networks%20for%20document%20processing.%20In%3A%20Tenth%20international%20workshop%20on%20frontiers%20in%20handwriting%20recognition.%20Seattle%3A%20Suvisoft%3B%202006.>)

78. 古德费尔一世, 本吉奥Y, 考维尔A.深度学习。麻省理工学院出版社; 2016年。
<http://www.deeplearningbook.org> (<http://www.deeplearningbook.org>).

79. Bengio Y, LeCun Y, 等。向AI扩展学习算法。大型核马赫。2007年; 34(5): 1-41。
谷歌学者 (http://scholar.google.com/scholar_lookup?title=Scaling%20learning%20algorithms%20towards%20AI&author=Y.%20Bengio&author=Y.%20LeCun&journal=Large%20Scale%20Kernel%20Machine%20Learning&volume=34&issue=5&pages=1-41&publication_year=2007)

80. 古维尔A, 文森特P.代表性学习: 回顾和新的观点。IEEE横贯模式肛门马赫INTEL。2013年; 35(8): 1798-828。

我们使用cookie来个性化内容和广告，提供社交媒体功能，并分析我们的流量。我们亦会与我们的社交媒体、广告及分析合作伙伴分享有关阁下使用本网站的资料。隐私声明。您可以在ManageCookies中管理您的首选项。

☐ 管理饼干

✓ 好的

q=Bengio%20Y.%20Deep%20learning%20forward.%20In%3A%20International%20language%20and%20speech%20processing.%20Berlin%3A%20Springer%3B%202013.%20p.%201%E2%80%93337.)

✓ 网页内容翻译完成

显示原文

中英对照

关闭

82. Calandra R, Raiko T, Deisenroth MP, Pouzols调频。从非平稳流中学习深层信念网络。国际人工神经网络会议。柏林: 斯普林格; 2012年。379-86
谷歌学者 (https://scholar.google.com/scholar?q=Calandra%20R%2C%20Raiko%20T%2C%20Deisenroth%20MP%2C%20Pouzols%20FM.%20Learning%20deep%20belief%20networks%20from%20non-stationary%20streams.%20In%3A%20International%20conference%20on%20artificial%20neural%20networks.%20Berlin%3A%20Springer%3B%202012.%20p.%20379%E2%80%93386.)

83. 周G, Sohn K, LeeH.带去噪自动编码器的在线增量特征学习。人工智能和统计学。2012年p.1453-61
谷歌学者 (https://scholar.google.com/scholar?q=Zhou%20G%2C%20Sohn%20K%2C%20Lee%20H.%20Online%20incremental%20feature%20learning%20with%20denoising%20autoencoders.%20In%3A%20Artificial%20intelligence%20and%20statistics.%202012.%20p.%201453%E2%80%93361.)

84. 张学文, 张晓华, 等.中华人民共和国科学院院士, 北京: 中国人民大学学报, 2004, 2000(2): 38-15.class='class 3'>, 等人.大规模分布式深网络。在: 神经信息处理系统方面的进展。2012年p.1223-31
谷歌学者 (https://scholar.google.com/scholar?q=Dean%20J%2C%20Corrado%20G%2C%20Monga%20R%2C%20Chen%20K%2C%20Devin%20M%2C%20Mao%20M%2C%20Senior%20A%2C%20Tucker%20P%2C%20Yang%20K%2C%20Le%20QV%2C%20et%20al.%20Large%20scale%20distributed%20deep%20networks.%20In%3A%20Advances%20in%20neural%20information%20processing%20systems.%202012.%20p.%201223%E2%80%93331.)

85. Coates A, Huval B, Wang T, Wu D, Catanzaro B, Ng A.国际机器学习会议。2013年p.1337-45
谷歌学者 (https://scholar.google.com/scholar?q=Coates%20A%2C%20Huval%20B%2C%20Wang%20T%2C%20Wu%20D%2C%20Catanzaro%20B%2C%20Ng%20A.%20Deep%20learning%20with%20cots%20hpc%20systems.%20In%3A%20International%20conference%20on%20machine%20learning.%202013.%20p.%201337%E2%80%93345.)

86. Glorot X, Bordes A, Bengio Y.大规模情感分类的域适应: 一种深度学习方法。在: 第28届国际机器学习会议记录(ICML-11); 2011年。513-20

谷歌学者 (<https://scholar.google.com/scholar?q=Glorot%20X%2C%20Bordes%20A%2C%20Bengio%20Y.%20Domain%201%20challenges%20in%20representation%20learning%20CVPR%202013>)

我们使用cookie来个性化内容和广告，提供社交媒体功能，并分析我们的流量。我们亦会与我们的社交媒体、广告及分析合作伙伴分享有关阁下使用本网站的资料。隐私声明。您可以在ManageCookies中管理您的首选项。

☐ 管理饼干

✓ 好的

3A%20deep%20learning%20for%20comparing%20between%20domains.%20n%20challenges%20in%20representation%20learning%20CVPR%202013.)

✓ 网页内容翻译完成

显示原文

中英对照

关闭

88. 题名/责任者: by L.H, Louradour J, Louradour J, L.Louradour J, Lamblin P.j马赫学习研究报告2009; 10: 1-40。
数学 (<http://www.emis.de/MATH-item?1235.68168>)
谷歌学者 (http://scholar.google.com/scholar_lookup?title=Exploring%20strategies%20for%20training%20deep%20neural%20networks&author=H.%20Larochelle&author=Y.%20Bengio&author=J.%20Louradour&author=P.%20Lamblin&journal=J%20Mach%20Learn%20Res&volume=10&pages=1-40&publication_year=2009)

89. Olshausen AB, Field DJ.过完备基集稀疏编码: v1所采用的策略? 远景决议 1997; 37(23): 3311-25。
交叉参考 ([https://doi.org/10.1016/S0042-6989\(97\)00169-7](https://doi.org/10.1016/S0042-6989(97)00169-7))
谷歌学者 (http://scholar.google.com/scholar_lookup?title=Sparse%20coding%20with%20an%20overcomplete%20basis%20set%203A%20a%20strategy%20employed%20by%20v1%3F&author=AB.%20Olshausen&author=DJ.%20Field&journal=Vision%20Res&volume=37&issue=23&pages=3311-3325&publication_year=1997)

90. Hinton G, Salakhutdinov R.通过学习深度生成模型发现文档的二进制代码。话题Cogn Sci。2011年; 3(1): 74-91。
交叉参考 (<https://doi.org/10.1111/j.1756-8765.2010.01109.x>)
谷歌学者 (http://scholar.google.com/scholar_lookup?title=Discovering%20binary%20codes%20for%20documents%20by%20learning%20deep%20generative%20models&author=G.%20Hinton&author=R.%20Salakhutdinov&journal=Topics%20Cogn%20Sci&volume=3&issue=1&pages=74-91&publication_year=2011)

91. Semantic Hinton G.INT J Approx Reas.2009年; 50(7): 969-78。
交叉参考 (<https://doi.org/10.1016/j.ijar.2008.11.006>)
谷歌学者 (http://scholar.google.com/scholar_lookup?title=Semantic%20hashing&author=R.%20Salakhutdinov&author=G.%20Hinton&journal=Int%20J%20Approx%20Reas&volume=50&issue=7&pages=969-978&publication_year=2009)

92. Le QV利用大规模无监督学习建立高层次特征。2013年IEEE声学、语音和信号处理国际会议(ICASSP)。2013年p.8595-8
谷歌学者 (<https://scholar.google.com/scholar?q=Le%20QV.%20Building%20high-level%20features%20using%20large%20scale%20unsupervised%20learning.%20In%203A%202013%20IEEE%20international%20conference%20on%20>

acoustics%2C%20speech%20and%20signal%20processing%20%28ICASSP%29.%202013.%20p.%208595%E2%80%9338.)

我们使用cookie来个性化内容和广告，提供社交媒体功能，并分析我们的流量。我们亦会与我们的社交媒体、广告及分析合作伙伴分享有关阁下使用本网站的资料。隐私声明。您可以在ManageCookies中管理您的首选项。

☐ 管理饼干

✓ 好的

94. Ranzato M, Szummer M.半监督学习：学习会议记录。纽约：ACM；2008年。谷歌学者 (https://scholar.google.com/scholar?q=Ranzato%20M%2C%20Szummer%20M.%20Semi-supervised%20learning%20of%20compact%20document%20representations%20with%20deep%20networks.%20In%3A%20Proceedings%20of%20the%2025th%20international%20conference%20on%20machine%20learning.%20New%20York%3A%20ACM%3B%202008.%20p.%20792%E2%80%9399.)

✓ 网页内容翻译完成

显示原文

中英对照

关闭

95. Mikolov T, Chen K, Corrado G, Dean J.向量空间中字表示的有效估计 arXiv:1301.3781 (http://arxiv.org/abs/1301.3781). 2013.

96. Mikolov T, Chen K, Corrado G, Dean J.向量空间中字表示的有效估计 arXiv:1301.3781 (http://arxiv.org/abs/1301.3781). 2013.

97. 用于句子分类的卷积神经网络。在：EMNLP会议记录。2014年。谷歌学者 (https://scholar.google.com/scholar?q=Kim%20Y.%20Convolutional%20neural%20networks%20for%20sentence%20classification.%20In%3A%20Proceedings%20of%20EMNLP.%202014.)

98. 乐QV, 邹伟, 杨锡, 吴安忆.学习具有独立子空间分析的动作识别的层次不变时空特征。2011年IEEE计算机视觉和模式识别会议(CVPR)。2011年p.3361-8 谷歌学者 (https://scholar.google.com/scholar?q=Le%20QV%2C%20Zou%20WY%2C%20Yeung%20SY%2C%20Ng%20AY.%20Learning%20hierarchical%20invariant%20spatio-temporal%20features%20for%20action%20recognition%20with%20independent%20subspace%20analysis.%20In%3A%202011%20IEEE%20conference%20on%20computer%20vision%20and%20pattern%20recognition%20%28CVPR%29.%202011.%20p.%203361%E2%80%9338.)

99. 施密德胡伯J, 康明斯F.学习遗忘：LSTM的持续预测。神经康普特。2000年；12：2451-71。交叉参考 (https://doi.org/10.1162/089976600300015015) 谷歌学者 (http://scholar.google.com/scholar_lookup?title=Learning%20to%20forget%3A%20continual%20prediction%20with%20LSTM&author=F.%20Gers&author=J.%20Schmidhuber&author=F.%20Cunhamins&journal=Neural%20Comput&volume=12&pages=2451-2471&publication_year=2000)

100. Graves A.基于递归神经网络的有监督序列标记。海德堡：斯普林格；2012年。交叉参考 (https://doi.org/10.1007/978-3-642-24797-2) 数学 (http://www.emis.de/MATH-item?1235.68014)

谷歌学者 (http://scholar.google.com/scholar_lookup?title=Supervised%20sequence%20labelling%20with%20recurrent%20neural%20networks%20for%20sentiment%20classification&hl=en)

我们使用cookie来个性化内容和广告，提供社交媒体功能，并分析我们的流量。我们亦会与我们的社交媒体、广告及分析合作伙伴分享有关阁下使用本网站的资料。隐私声明。您可以在ManageCookies中管理您的首选项。

☐ 管理饼干

✓ 好的

102. BERGSTRA J, Breuleux O, Bastien F. 表达式编译器。代表科学计算会议的Python。谷歌学者 (<https://scholar.google.com/scholar?q=Bergstra%20J%2C%20Breuleux%20O%2C%20Bastien%20F%2C%20Bengio%20Y.%20Theano%3A%20a%20CPU%20and%20GPU%20math%20expression%20compiler.%20In%3A%20Python%20for%20scientific%20computing%20conference.%202012>.)
103. Bergstra J, Breuleux O, 巴斯蒂安F, 兰布林P.竖起大拇指？情绪分类使用机器学习技术，Python在科学，第二卷。9.2015年。谷歌学者 (<https://scholar.google.com/scholar?q=Bergstra%20J%2C%20Breuleux%20O%2C%20Bastien%20F%2C%20Lambelin%20P.%20Thumbs%20up%3F%20Sentiment%20classification%20using%20machine%20learning%20techniques%2C%20Python%20in%20science%2C%20vol.%209.%202015>.)
104. Pang B, Lee L, Vaihyathan S. TensorFlow: 异构分布式系统上的大规模机器学习。在：初步白皮书，第二卷。9.2015年。谷歌学者 (<https://scholar.google.com/scholar?q=Pang%20B%2C%20Lee%20L%2C%20Vaithyanathan%20S.%20TensorFlow%3A%20large-scale%20machine%20learning%20on%20heterogeneous%20distributed%20systems.%20In%3A%20Preliminary%20white%20paper%2C%20vol.%209.%202015>.)
105. 书名/责任者：by L.传记，宝莱坞，繁荣盒和搅拌机：领域适应情绪分类。ACL.2007年；7：440-7。谷歌学者 (http://scholar.google.com/scholar_lookup?title=Biographies%2C%20bollywood%2C%20boom-boxes%20and%20blenders%3A%20domain%20adaptation%20for%20sentiment%20classification&author=J.%20Blitzer&author=M.%20Dredze&author=F.%20Pereira&journal=ACL&volume=7&pages=440-447&publication_year=2007)
106. 王S, 曼宁CD.基线和大写：简单，良好的情感和主题分类。在：计算语言学协会第50届年会议事录：短文，第二卷。2.计算语言学协会；2012年。90-4。谷歌学者 (<https://scholar.google.com/scholar?q=Wang%20S%2C%20Manning%20CD.%20Baselines%20and%20bigrams%3A%20simple%2C%20good%20sentiment%20and%20topic%20classification.%20In%3A%20Proceedings%20of%20the%2050th%20annual%20meeting%20of%20the%20association%20for%20computational%20linguistics%3A%20short%20papers%2C%20vol.%202.%20Association%20for%20computational%20linguistics%3B%202012.%20p.%2090%E2%80%9394>.)
107. 陈国明, 王英芳, 李政德.使用缩略图来加强文本分类。通知流程经理。2002年；38(4): 529-46。

交叉参考 ([https://doi.org/10.1016/S0306-4573\(01\)00045-0](https://doi.org/10.1016/S0306-4573(01)00045-0))
数学 (<http://www.emis.de/MATH-item?1052.68611>)

我们使用cookie来个性化内容和广告，
提供社交媒体功能，并分析我们的流
量。我们亦会与我们的社交媒体、广告
及分析合作伙伴分享有关阁下使用本网
站的资料。隐私声明。您可以在
ManageCookies中管理您的首选项。

☐ 管理饼干

✓ 好的

(C)提交人2018年

✓ 网页内容翻译完成

显示原文

中英对照

关闭

开放存取本文以Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>) (<http://creativecommons.org/licenses/by/4.0/>)), 它允许在任何媒体中不受限制地使用、分发和复制, 只要您对原始作者和源给予适当的信任, 提供到Creative Commons License的链接, 并指示是否进行了更改。

关于这篇文章

引用这一条如下:
Sohangir, S., Wang, D., Pomeranets, A. et al. J Big Data (2018) 5: 3. <https://doi.org/10.1186/s40537-017-0111-6>
DOI (Digital Object Identifier) <https://doi.org/10.1186/s40537-017-0111-6>
出版者名称 斯普林格国际出版社
在线ISSN 2196-1115

[关于这本日记](#)
[重印和权限](#)

个性化建议

1. Aspect-based Financial Sentiment Analysis with Deep Neural Networks

Published online: 17 August 2018

我们使用cookie来个性化内容和广告，提供社交媒体功能，并分析我们的流量。我们亦会与我们的社交媒体、广告及分析合作伙伴分享有关阁下使用本网站的资料。隐私声明。您可以在ManageCookies中管理您的首选项。

☐ 管理饼干

✓ 好的

18 (2018)

✓ 网页内容翻译完成

显示原文

中英对照

关闭

3. A Novel Lexicon-Based Approach Using Learning Automata

Sarigiannidis, Antonios... Pragidis, Ioannis-Chrysostomos
Internet Science (2018)

Want recommendations via email? [Sign up now](#)

Powered by: Recommended 

SPRINGER NATURE

(2017 年)瑞士斯普林格自然公司。部分斯普林格自然。

未登录 无关联 183.135.116.240

沪icp备15051854号-2