

DOI: 10.16538/j.cnki.fem.2016.07.007

## 大数据在劳动力市场研究中的应用与展望

史珍珍, 曾湘泉

(中国人民大学劳动人事学院, 北京 100872)

**摘要:** 互联网为劳动经济学研究提供了全新的数据收集渠道。基于互联网的劳动力市场大数据具有动态性、及时性等特点, 能用来分析传统调查数据无法分析的问题。本文在文献分析的基础上, 讨论了劳动力市场大数据的产生过程、内涵及应用, 并在此基础上总结了数据收集方式、测量指标, 以及可能存在的问题与解决方法, 最后对国内学者未来可以基于大数据对劳动力市场开展的研究进行了展望, 以期推动大数据在国内劳动经济学研究中的应用。

**关键词:** 大数据; 劳动力市场; 基于互联网的劳动力市场大数据; 搜寻与匹配

**中图分类号:** F270 **文献标识码:** A **文章编号:** 1001-4950(2016)07-0096-17

### 一、引言

劳动力市场搜寻与匹配是劳动经济学研究的重要问题。自1962年斯蒂格勒在《劳动力市场中的信息》一文中提出劳动力市场信息理论之后, 学者们构建了工作搜寻理论, 并基于调查数据对劳动者的工作搜寻行为进行了大量实证研究。相比之下, 有关雇主搜寻策略和行为的研究仍显不足(Barron等, 1985; Villena-Roldán, 2012)。雇主搜寻研究的滞后在一定程度上是缺乏数据导致的(DeVaro, 2005, 2008; Horton, 2010; Pellizzari, 2011)。在我国, 重视劳动供给研究、轻视劳动需求研究的倾向存在已久(蔡昉, 2002)。国内权威的劳动需求调查、职位空缺调查缺位是造成这一现象的关键原因。同时, 传统的问卷调查只能获取搜寻与匹配结果数据, 相关的过程数据仍较为缺乏, 导致劳动力市场动态研究进展缓慢。而且, 现有的劳动力市场动态研究基于问卷调查的“回忆性”填答(Burdett和Cunningham, 1998)获取的有关搜寻过程的数据具有测量误差, 导致模型估计结果有偏。

收稿日期: 2016-02-07

基金项目: 国家社会科学基金重点项目(13AZD005)

作者简介: 史珍珍(1983—), 女, 中国人民大学劳动人事学院博士研究生, 中国人民大学中国就业研究所助理研究员;

曾湘泉(1955—), 男, 中国人民大学劳动人事学院教授、博士生导师, 中国人民大学中国就业研究所所长。

随着互联网在劳动力市场搜寻与匹配中应用规模的不断扩大，基于互联网的劳动力市场大数据应运而生。劳动力市场大数据不仅具有海量样本的特征，还具有动态性和即时性等显著特征，因此能够用来分析那些传统问卷调查数据无法分析的问题（Kuhn，2014）。近年来，采用劳动力市场大数据的实证研究陆续出现（Kureková等，2015），并陆续刊登在《经济学季刊》（*Quarterly Journal of Economics*）、《计量经济学》（*Econometrica*）、《经济学视角》（*Journal of Economic Perspective*）、《劳动经济学》（*Journal of Labor Economics*）等核心期刊上。本文基于这些实证研究，在文献分析的基础上，讨论劳动力市场大数据的产生过程、内涵及在劳动力市场研究中的应用，并在此基础上，总结劳动力市场大数据的收集方法、测量指标，以及数据中可能存在的问题及解决方法。最后，本文展望了国内学者采用此类数据开展劳动力市场研究的可能性，以期为推动大数据在国内劳动经济学研究中的应用提供参考。

## 二、劳动力市场大数据的产生过程及内涵

目前，对于大数据的概念，企业界和学术界尚未形成公认的准确定义（冯芷艳等，2013），然而，这并不影响社会各界对大数据的探讨。美国国家科学基金会（National Science Foundation）将大数据定义为：基于仪器、传感器、互联网交易、电子邮件、视频等数据源所生成的大规模、多元化、综合的、追踪性的分布式数据集<sup>①</sup>。我们将以该定义为基础，讨论和总结劳动力市场大数据的产生过程及内涵。

劳动力市场大数据的产生源于互联网对劳动力市场运行方式的变革。信息技术和互联网的发展改变了劳动者与雇主之间的搜寻与匹配机制（Autor，2001）。Freeman（2002）这样描述了互联网时代劳动力市场的搜寻与匹配机制：求职者通过互联网了解到更多有关工作机会（offer）的信息；雇主通过互联网收集更多、更详细的求职者信息；最为关键的是，求职者和雇主不仅能在短时间内收集和浏览大量的信息，而且信息收集过程不再受时间和地域的限制，信息的更新速度更快。因此，互联网逐渐成为应用广泛的劳动力市场中介，甚至可能取代其他中介，成为未来劳动力市场搜寻与匹配的主要渠道（Parry和Wilson，2009）。在此背景下，网络所记录的有关工作搜寻、雇主搜寻的过程与结果的数据（McLaren和Shanbhogue，2011；Askitas和Zimmermann，2015），形成了劳动力市场搜寻与匹配的客观观测数据集，受到了研究者的重视。例如，Chan和Swatman（2000）尝试性地在1999年5月—2000年1月，从AustralianJob Search.com上收集了有关“电子商务”类岗位的招聘信息，并通过简单的描述统计，阐述了电子商务类岗位的工作内容和职责，以期为该学科的毕业生提供就业指导。Wade和Parent（2002）在分析雇主对网站管理员（webmaster）的技能要求时，同时在两家贸易杂志和五家招聘网站上共收集了800个网站管理员招聘广告<sup>②</sup>。他们采用内容分析法（content analysis）所进行的分析显示，雇主对网站管理员技能的要求是多层面的，包括专业技能（technical skills）和组织技能（organizational skills），而且对专业技

<sup>①</sup> 参见<http://www.cccblog.org/2012/03/29/obama-administration-unveils-200m-big-data-rd-initiative/>。

<sup>②</sup> 这两家贸易杂志为*Computerworld*和*Informationweek*。五家在线工作搜索网站为[www.monster.com](http://www.monster.com)；[www.computerjobs.com](http://www.computerjobs.com)；[www.jobengine.com](http://www.jobengine.com)；[www.jobs-online.com](http://www.jobs-online.com)；[www.peoplesearch.com](http://www.peoplesearch.com)。

能的要求高于对组织技能的要求。

通过梳理国外文献,我们发现劳动经济学家并没有对劳动力市场大数据进行严格的定义。Autor(2001)认为招聘网站(internet job board)不仅能提供某一时点的劳动力市场“快照”(snapshot)信息,而且能提供劳动供需双方的交易历史(transaction history)数据。与此类似,Kuhn(2014)认为通过招聘网站(online job board)可以获得某空缺岗位收到的求职申请数量、求职者发送简历的数量及发送地等数据,并将这种数据命名为基于互联网的劳动力市场数据。除了招聘网站,互联网搜索引擎,如Google,也是人们搜索工作信息的工具之一。因此,研究者根据人们在Google上的工作检索行为来预测失业率,并将其命名为“谷歌预测器”(Google predictor)(Askitas和Zimmermann,2009)或者“谷歌指数”(Google index,GI)(D'Amuri和Marcucci,2009)。已有学者采用这种数据预测了美国(D'Amuri和Marcucci,2009)、德国(Askitas和Zimmermann,2009)、法国(Fondeur和Karamé,2013)、英国(McLaren和Shanbhogue,2011)、意大利(D'Amuri,2009)、以色列(Suhoy,2009)等国家的失业率。因此,从现有实证研究可以发现,劳动力市场大数据是基于互联网收集的有关搜寻与匹配的数据集。这里的互联网既包括互联网搜索引擎(如Google),也包括招聘网站、企业网站等媒介。另外,有学者将通过互联网进行问卷调查获取的数据(如Kureková等,2015)或者通过互联网进行实验获取的数据(如Edelman,2012)也列入大数据行列。然而,本文认为这些数据并非自然观测数据,而只是借由互联网这个媒介发放问卷或进行实验所获得的数据,因此不在本文讨论的大数据范畴。

### 三、劳动力市场大数据在实证研究中的应用

早期,学者们采用大数据对劳动力市场的研究,集中在根据招聘广告内容分析雇佣标准方面。随着大数据的发展,学者们逐渐将其应用在搜寻与匹配过程、搜寻持续时间及雇主偏好等方面的研究中。

#### (一) 雇佣标准研究

20世纪90年代初,随着互联网的普及,雇主对互联网相关高技能工人的需求显著增长(Autor等,1998)。在此背景下,学者们尝试从网站收集招聘广告,并根据广告的内容,来分析信息技术(information technology,IT)相关岗位的雇佣标准。如前所述,Wade和Parent(2002)分别从雇主和劳动者角度分析了网站管理员职位对任职者组织技能和专业技能的要求。他们对从杂志和招聘网站收集的800个网站管理员职位招聘广告的分析显示,雇主对专业技能的要求高于对组织技能的要求。他们进一步对232名网站管理员进行的问卷调查显示,这些从业人员认为组织技能和专业技能对工作同等重要。Capiluppi和Baravalle(2010)通过一个计算机程序(web spider)在Monster.com上收集了48 000个招聘广告,数据分析结果显示,英国雇主对IT技能工人存在较大需求。他们进一步将雇主IT技能需求的内容与10所大学IT专业的课程进行了对比分析,认为大学的培养内容与方案不能完全满足雇主对IT技能的需求,即现行人才培养机制与劳动需求之间存在错配(mismatch)。

随着高等教育的大众化,大学毕业生在劳动力市场中的比例不断提高,高技能工人与低技能工人之间的替代效应(拉齐尔,2000)或者大学毕业生对低技能工人的挤出效

应 (crowding out) 受到了学者的关注。由于劳动力市场大数据包含了高技能岗位和低技能岗位的信息, 学者们基于相关数据能够从雇主需求的角度对上述效应进行分析。Kureková和Zilinčíková (2015) 使用斯洛伐克招聘网站 (Profesia. sk) 的数据, 根据任职资格要求, 将岗位分为要求应聘者为大学毕业生的岗位 (student positions) 和不要求应聘者为大学毕业生的岗位 (non-student positions), 并在此基础上, 从劳动需求角度分析了这两类岗位对求职者技能的要求。他们发现, 没有接受大学教育的工人 (低技能工人) 与大学毕业生 (高技能工人) 不是竞争关系, 而是以一种互补的方式满足雇主不同层次的技能需求。Beblavý等 (2016) 采用相同的网站数据, 不同的岗位分类方法, 分析了中等技能岗位和低技能岗位在学历要求方面的差异。他们将技能岗位划分为高、中、低三个等级, 并通过内容分析法和简单统计模型分析发现, 中等技能岗位和低技能岗位对学历及其他技能的要求存在差异。这再次说明高技能工人与低技能工人之间存在互补性。

学者们还基于不同国家招聘网站的数据, 进行了雇佣标准的国际比较分析。在以往的研究中, 跨国性的职位空缺调查存在较大难度, 使得学者们无法基于调查数据分析雇佣标准的国际差异。然而, 从一个多国使用的网站上收集的数据有利于进行这种国际比较研究。欧洲就业服务网 (European Employment Service, EURES) 是一个标准化网站, 在欧盟以及EEA成员国<sup>①</sup>之间通用。而且, 该网站发布的空缺岗位数量占欧洲劳动力市场空缺岗位总量的30%—40% (Ackers, 2012)。Kureková等 (2015) 采用该网站的数据, 分析了雇主对劳动者认知技能和非认知技能要求的国际差异。结果显示, 不同国家对这种技能组合 (skill-mix) 的要求不同, 这表明国家内在因素对雇主需求的形成存在影响。

## (二) 搜寻与匹配过程研究

基于网络大数据, 学者们能够针对互联网中介在搜寻与匹配中的作用进行深入分析。1962年斯蒂格勒在其论文《劳动力市场中的信息》中指出, 劳动者在首次进入劳动力市场时, 面对大量的潜在雇主, 如何获取相关的工资、就业稳定性等信息是工作搜寻的关键。因此, 分析劳动力市场中介在搜寻与匹配中的作用是学者们研究的重要问题。Stanton和Thomas (2014) 利用从oDesk. com上收集的2008年8月1日—2009年12月28日的劳动者求职数据以及雇主雇佣数据研究发现, 互联网作为中介向雇主发出的有关劳动者低技能、工作经验少的信号是高质量的。对于劳动者而言, 在职业生涯初期, 与中介建立联系则找到工作的概率更高, 得到的工资也更高。这表明, 在劳动力市场搜寻与匹配过程中, 中介在人员配置方面具有较高的效率。

网络大数据所具有的即时性和动态性等特征 (Autor, 2001) 使其可用于劳动力市场动态研究。问卷调查提供的是劳动力市场搜寻与匹配的结果数据, 相比之下, 学者们通过网络不仅能够获取搜寻与匹配的结果数据, 还能收集状态变化数据。因此, 网络大数据有助于推动劳动力市场动态研究的发展。例如, 雇主在搜寻过程中, 为了降低搜寻成本, 可以公布高工资以降低低技能工人的求职申请比例, 或者提高自身的筛选、甄别能力。通过问卷调查获取的数据由于只能提供雇主搜寻的结果信息 (van Ours和

<sup>①</sup> EEA成员国包括欧盟国家、挪威、冰岛和列支敦士登。

Ridder, 1991), 因此无法用来分析雇主在面临这两种替代性选择 (Stigler, 1962) 时的决策。然而, 基于通过互联网收集的雇主搜寻过程动态信息 (Brenčič, 2009), 学者们得以对上述问题进行分析。Brenčič和Norris (2010a) 在Monster.com上收集了121 381个空缺岗位在2周内的状态变化数据, 研究发现, 雇主变更招聘广告内容的行为, 如降低或提高工资水平、改变任职要求, 与雇佣成本存在正相关关系, 即当雇主的搜寻成本发生变化时, 雇主会主动改变招聘广告中的任职要求。此外, 基于招聘广告数据, 学者们还直接从雇主的角度分析了雇主实行某种工资制度的决策机制和影响因素。Brenčič和Norris (2010b) 从美国Monster.com上收集了250 000个空缺岗位数据, 来分析雇主支付绩效工资的决定因素。结果发现, 当空缺岗位涉及多任务、质量控制或团队合作时, 雇主为该岗位提供绩效工资的可能性较低。

在经济学研究中, 学者们往往关注历史中的自然实验, 以替代实验室实验来对某些经济问题或现象进行研究。基于互联网的劳动力市场大数据, 能够即时反映外部冲击产生的影响, 因而更加容易形成自然实验数据。以2008年爆发的金融危机为例, 为了缓解危机期间“就业难”的形势, 美国政府将领取失业救济金的时间从26周延长至99周 (Marinescu, 2014)。该政策是否达到了预期目的呢? 由于传统的 (或官方的) 统计数据具有发布期滞后、无法即时反映经济结构变化等弊端, 因此无法利用其进行及时的政策评估 (Askatas和Zimmermann, 2009)。而网络大数据所包含的庞大且连续的信息使这一难题迎刃而解。Marinescu (2014) 在CareerBuilder.com上收集了2007—2011年间的工作搜寻相关数据。由于包含了经济危机前后的信息, 该数据形成了自然实验数据。Marinescu研究发现, 经济危机期间推出的积极政策具有就业抑制效应: 失业救济金领取时间延长10%, 使得州一级的求职人数降低1%。从以上分析可见, 传统调查数据的滞后性使其更适用于评估政策的长期效果, 而大数据则更有利于对政策短期效果的评估。从这个层面来看, 大数据是调查数据的一种补充。

### (三) 搜寻持续时间研究

目前, 有关工作搜寻持续时间或失业持续时间的研究所使用的数据, 主要是基于问卷调查获得的, 即根据被调查者报告的工作搜寻开始时间和结束时间来计算持续时间。而在招聘网站上, 学者们通过观测劳动者首次和末次投递简历的时间, 便可直接测量工作搜寻持续时间。Faberman和Kudlyak (2014) 在美国SnagAJob.com上进行了相关数据的收集。他们在观测期 (2010年9月—2011年9月) 内, 观测劳动者投递第一份求职申请的时间和投递最后一份求职申请的时间, 并根据这两个时间点来测算工作搜寻持续时间。他们以此数据分析求职者工作搜寻强度与搜寻持续时间之间的关系。研究发现, 随着工作搜寻时间的推移, 求职者发送求职申请的数量越来越少, 即工作搜寻强度逐渐降低; 劳动者通过互联网进行工作搜寻的时间越长, 在整个持续时间内, 其工作搜寻强度就越大。

雇主搜寻持续时间或者岗位空缺持续时间研究是从动态角度对搜寻成本 (Holzer, 1987) 进行分析。学者们通过在招聘网站上观测空缺岗位的状态变化及相应时间, 可以直接测量空缺持续时间。Brenčič和Norris (2009, 2010b, 2012) 在2005年4月30日—7月7日期间, 在Monster.com上以选定的空缺信息为对象, 连续16周监测这些空缺信息的变

化及其时间点，最后以空缺信息的下线时间和上线时间之差来测量空缺持续时间。基于此类数据，学者们分析了空缺持续时间的影响因素。Nivalainen（2014）在2002—2003年从芬兰公共就业服务网站（Public Employment Service, PES）上收集了空缺持续时间数据，以评估就业政策的效果。研究显示，2002年10月芬兰公共就业服务机构推出的雇主网络招聘政策，有效缩短了雇主搜寻的持续时间，但是这种促进效应具有显著的区域差异：对城市地区的雇主而言，促进效应更大。

#### （四）雇主偏好研究

根据网站的招聘广告内容，学者们分析了劳动力市场搜寻与匹配过程中的雇主偏好问题。Kuhn和Shen（2009）分别在2008年5月16日—7月29日以及2008年12月17日—2009年2月28日期间，在Zhaopin.com上收集了633 664条招聘信息。通过对这些招聘广告内容的分析，他们发现雇主对求职者性别、年龄、身高和外貌存在一定的偏好。而且，雇主的性别偏好与企业的发展阶段存在一定的联系。Helleseter等（2014）分别在“xmzyjs.com”“xmrc.com.cn”“zhaopin.com”和“computrabajo.com”四个招聘网站上收集了141 188、39 727、1 051 038和90 487条招聘广告信息。他们研究发现，雇主在企业的不同发展阶段，对女性外貌和男性领导力的重视程度不同，导致其具有不同的性别偏好。

传统的问卷调查只能获取被某一岗位雇用的劳动者的信息，而无法获取没有被该岗位雇用的劳动者的信息。然而，网络大数据却同时包含这两类信息。基于此类数据，经过对比分析，可以得出有关雇主偏好的结论。例如，根据大数据中劳动者的个体特征、工作搜寻历史以及所在地信息（Rand，2012），学者们分析了雇佣过程中雇主对劳动者所在地的偏好。在线劳动力市场（如freelancer.com和oDesk.com）为劳动供需双方提供了劳动交易的互联网平台（Hong和Pavlou，2012；Pallais，2014）：劳动者免费在网站上注册、雇主免费在网站上发布工作内容和要求。劳动者选择合适的工作并向雇主发送申请，雇主决定由哪位劳动者来完成工作。被雇主选定的劳动者进行远程办公。工作完成后，劳动者通过网站将工作成果交予雇主。雇主在对工作成果进行评估后，将工作报酬支付给网站。在扣除一定比例的费用后，网站将剩余的报酬支付给劳动者（Mill，2011；Agrawal等，2012）。由于劳动者在网站上注册时，填写的个人信息比较少，雇主只能根据劳动者注册时所在地区的特征来判断劳动者的劳动生产率，进而做出是否雇用的决策。Mill（2011）从freelancer.com上收集了申请同一岗位被雇用的与没有被雇用的劳动者的数据，研究发现，居住地对劳动者是否被雇用具有显著影响。他将这种地区偏好归为统计性歧视。与此研究类似，Agrawal等（2012）在oDesk.com上随机选择劳动者，分析劳动者被雇用的概率的区域差异。结果显示，在欠发达地区的求职者申请发达地区的工作时，存在惩罚效应（penalty）：在控制了学历、工作经验和保留工资等变量的情况下，发达地区的雇主雇用欠发达地区求职者的概率较低。

与地区偏好类似的是户籍偏好。学者们基于招聘广告的内容，分析了雇主对求职者户籍的偏好。Kuhn和Shen（2014）在2010年7月—2011年7月期间，在xmrc.com.cn上收集了237 128份求职申请数据，以及是否得到雇主反馈（feedback）的数据。他们基于这些数据分析了雇主对求职者户籍的偏好。结果显示，私营企业更倾向于雇用非本地求职者，而且，对于低技能岗位而言这一点更为突出。这是因为非本地劳动者的工资低于本地劳动者；而且，为了提高工作收入，非本地劳动者的工作努力程度更高、延长工作时间的意愿更强烈。所以，他们认为，对于追求利润最大化的雇主而言，非本地劳动者的

这些特征能够满足雇主的偏好。

表1对采用了劳动力市场大数据的实证研究文献进行了汇总。

表1 采用劳动力市场大数据的实证研究文献汇总

研究问题	研究者（时间）	数据	研究发现
雇佣标准研究	Wade和Parent（2002）	Monster.com等5家招聘网站刊登的800个网站管理员职位招聘广告数据；232名网站管理员问卷调查数据	雇主对专业技能的要求高于对组织技能的要求，但是网站管理员群体认为组织技能和专业技能对于本工作同等重要
	Capiluppi和Baravalle（2010）	在Monster.com上收集了48 000个招聘广告	英国雇主对IT技能工人存在较大需求，然而现行人才培养机制与劳动需求之间存在错配
	Kureková和Zilinčíková（2015）	在Profesia.sk网站上收集了2009—2014年的招聘广告信息，共收集到493 853条信息	没有接受大学教育的工人（低技能工人）与大学毕业生（高技能工人）不是竞争关系，而是以一种互补的方式满足雇主不同层次的技能需求
	Beblavý等（2016）	Profesia.sk网站招聘广告信息	中等技能岗位和低技能岗位对学历及其他技能的要求存在差异
	Kureková等（2015）	欧洲就业服务网站招聘广告数据	雇主对求职者认知技能与非认知技能要求的国际差异来自于国家内在因素
搜寻与匹配过程研究	Stanton和Thomas（2014）	2008年8月1日—2009年12月28日期间从oDesk.com上共收集了124 801条求职信息	互联网作为中介向雇主发出的有关劳动者低技能、工作经验少的信号是高质量的；在职业生涯初期，与中介建立联系的劳动者找到工作的概率更高，得到的工资也更高
	Brenčič和Norris（2010a）	在Monster.com上收集了121 381个空缺岗位在2周内的状态变化数据	雇主变更招聘广告内容的行为，如降低或提高工资水平、改变任职资格要求，与雇佣成本存在正相关关系
	Brenčič和Norris（2010b）	从Monster.com上收集了250 000个空缺岗位数据	当空缺岗位涉及多任务、质量控制或团队合作时，雇主为该岗位提供绩效工资的可能性较低
	Marinescu（2014）	从CareerBuilder.com上收集了2007—2011年间共2 397条工作搜寻信息	2008年金融危机期间推出的积极就业政策具有就业抑制效应：失业救济金领取时间延长10%，使得州一级的求职人数降低1%
搜寻持续时间研究	Faberman和Kudlyak（2014）	2010年9月—2011年9月在SnagAJob.com上观测了800万名求职者向181万个空缺岗位发送的4 615万条申请	随着工作搜寻时间的推移，工作搜寻强度逐渐降低；劳动者通过互联网进行工作搜寻的时间越长，在整个持续时间内，其工作搜寻强度就越大
	Brenčič和Norris（2009，2010b，2012）	2005年4月30日—7月7日在Monster.com上连续16周监测空缺信息从网站下线的时间及状态变化数据	岗位空缺持续时间受到岗位对求职者学历及工作经验的要求、雇主对工作申请过程介绍的详细程度等因素的显著影响
	Nivalainen（2014）	2002—2003年从芬兰公共就业服务网站收集了90 000条空缺岗位信息	2002年10月芬兰公共就业服务机构推出的雇主网络招聘政策，有效缩短了雇主搜寻持续时间，但是这种促进效应具有显著的区域差异：对城市地区的雇主而言，促进效应更大

续表

研究问题	研究者（时间）	数据	研究发现
雇主偏好研究	Kuhn和Shen（2009）	2008年5月16日—7月29日以及2008年12月17日—2009年2月28日在Zhaopin.com上共收集了633 664条招聘信息	雇主对求职者的性别、年龄、身高和外貌存在一定的偏好
	Helleseeter等（2014）	分别在“xmzyjs.com”“xmrc.com.cn”“zhaopin.com”和“computrabajo.com”上收集了141 188、39 727、1 051 038和90 487条招聘广告信息	雇主的性别偏好与企业发展阶段存在相关性：雇主在企业的不同发展阶段，对女性外貌和男性领导力的重视程度不同，导致其具有不同的性别偏好
	Mill（2011）	从freelancer.com上收集了336 556条求职信息以及是否被雇主雇用的信息	研究者认为居住地对劳动者是否被雇用具有显著影响，并将这种地区偏好归为统计性歧视
	Agrawal等（2012）	在oDesk.com上收集了2011年9月和12月的35 997条工作申请信息	欠发达地区的求职者申请发达地区的工作时存在惩罚效应（penalty）：在控制了学历、工作经验和保留工资等变量的情况下，发达地区的雇主雇用欠发达地区求职者的概率较低
	Kuhn和Shen（2014）	在xmrc.com.cn上收集了237 128份求职申请数据，以及是否得到雇主反馈的数据	私营企业更倾向于雇用非本地求职者，而且，对于低技能岗位而言这一点更加突出。这是因为与本地劳动者相比，非本地劳动者的工资水平更低、工作努力程度和延长工作时间的意愿更高

#### 四、劳动力市场大数据的收集方式与测量指标

##### （一）收集方式

传统的调查数据通过问卷设计、问卷发放与问卷回收的流程来收集，基于互联网的数据其收集方式则与此有显著区别。通过梳理现有文献，我们将劳动力市场大数据的收集方式归结为直接观测、跟踪监测和网站档案数据库收集三种。

第一种是通过直接观测收集数据。在使用这种方法时，研究者在某一时点，在既定的招聘网站上，直接观测劳动者的求职申请、雇主发布的招聘广告以及双方形成的匹配信息，获得横截面数据。Edelman（2012）认为经济学家通过互联网收集数据的步骤包括：确定目标网站、借由计算机程序在网站上收集数据以及进行数据分析。例如，Kuhn和Shen（2013）分别在2008年5月19日—6月22日、2009年1月19日—2月22日、2009年5月18日—6月21日以及2010年1月18日—2月21日，在Zhaopin.com上收集招聘广告数据。在收集数据之前，他们设计了一个计算机应用程序。在数据收集期间，从每天上午11:30到次日00:00，该程序会自动收集该网站上的招聘广告。而且，在当天收集的招聘广告中，程序会标记出哪些是新刊登的，哪些是前一天就存在的。他们共收集了1 051 706条招聘广告。对这些招聘广告内容的分析显示了雇佣过程中雇主的性别偏好现象。

第二种是通过跟踪监测收集数据。在使用这种方法时，研究者在招聘网站上连续监测选定的求职者或雇主，观测他们的搜寻行为，以获得短期或长期跟踪数据（longitudinal data）。例如，Brenčić和Norris（2009，2010b，2012）按照以下三个步骤在Monster.com上收集岗位空缺持续时间数据：第一，列出一些关键词，如学历要求、

技能要求等。第二，将包含关键词的招聘广告作为监测对象，并进行为期16周的跟踪观测。每周定时监测一次，其目的是观察这些空缺信息在什么时间失效，并记录具体的失效时间<sup>①</sup>。第三，利用空缺信息失效的时间减去该信息开始刊登的时间，获得空缺持续时间数据。在通过自然观测获取的数据中，持续时间的测量单位可以精确到天，比问卷调查更为准确。在传统的职位空缺调查中，被调查雇主以“回忆性”方式报告招聘起止时间，或者招聘活动的持续时间。在这个过程中，雇主往往无法提供具体的时间，只报告大概的时间。例如，在1982年美国雇主调查（Employer Opportunity Pilot Project, EOPP）获取的数据中，空缺持续时间集中在30天、60天和90天。这种测量误差导致估计结果有偏（Burdett和Cunningham, 1998）。此外，有些雇主仅报告招聘的起止月份，学者只能根据这些月份信息近似计算空缺持续时间。van Ours（1989）在利用荷兰1986年职位空缺调查数据（How do firms recruit?）测量空缺持续时间时，对于招聘在一个月完成的岗位，将其空缺持续时间确定为15天，反之，首先计算招聘持续的月数，再将一个月定义为30天，进而计算空缺持续的天数。

第三种是通过网站档案（internet archive）收集二手数据。网站档案包含很多历史数据，利用这些二手数据能够分析网站使用者的行为方式和特征（Edelman, 2012）。在现有的劳动力市场研究中，学者们主要基于Google储存的工作检索（unemployment-related search）信息收集数据，并以此预测和分析失业率。2008年夏季，Google推出了一项名为“Google Insights”的工具，通过该工具学者们可以收集某一搜索关键词在Google上被检索的频率和相关统计数据（McLaren和Shanbhogue, 2011）。Askitas和Zimmermann（2009）采用从Google Insights中获取的劳动者工作检索数据，分析了德国的失业率。首先，他们列出4个与工作搜索相关的关键词：失业服务机构（unemployment office or agency）、失业率（unemployment rate）、人事咨询（personnel consultant）、德国使用最多的工作搜索引擎（most popular job search engines in Germany）。其次，在Google Insights中收集第M月第1周和第2周以及第M-1月第3周和第4周这些关键词的检索频率数据。最后，基于这些数据对第M月的失业率进行预测<sup>②</sup>。

表2总结了劳动力市场大数据的收集方式。

## （二）测量指标

就工作搜寻而言，劳动者通过鼠标在网页上的“点击”来完成信息浏览、简历投递，因此，劳动者的“鼠标点击”行为轨迹和次数及相应时间，是工作搜寻的测量指标。以CareerBuilder.com为例，每条岗位空缺信息都有一个“立即申请”按钮，劳动者想要申请某个岗位，用鼠标点击该岗位的这一按钮，并按照流程完成相关步骤即可。一般而言，招聘网站规定同一个岗位劳动者只能申请一次。因此，在既定时间内，“工作申请数量”是工作搜寻强度的测量指标。Marinescu（2014）利用该网站的数据分析了工作搜寻行为的特征。他基于连续3年的横截面数据分析发现，工作搜寻强度受失业救济金政策的影响显著：失业救济金领取时间延长1周，工作搜寻强度降低0.4%。此外，在招聘网站上，劳动者搜寻的起止时间差是工作搜寻持续时间的测量指标。Fberman和

<sup>①</sup> 在Monster.com上，当某条岗位空缺信息出现“抱歉，此空缺信息已不存在”的提示时，就表示该空缺信息已经失效。

<sup>②</sup> 在德国，对第M月失业率的估计，是依据第M-1月第3周和第4周、第M月第1周和第2周的失业数据进行的(Askitas和Zimmermann, 2009)。

表2 劳动力市场大数据的收集方式

收集方式	举 例			
	研究者（时间）	网站/工具	数据收集过程	数 据
直接观测	Kuhn和Shen（2013）	Zhaopin.com	设计一个计算机程序。在数据收集期间，每天上午11:30到次日00:00，该程序自动收集该网站上的招聘广告	1 051 706条招聘广告
跟踪监测	Brenčič和Norris（2012）	Monster.com	首先，列出一些关键词，如学历要求、技能要求等。其次，将包含关键词的招聘广告作为监测对象，进行为期16周的跟踪观测。每周定期监测一次，并记录其失效时间。最后，利用空缺信息失效的时间减去该信息开始刊登的时间，获得空缺持续时间数据	379 310个空缺岗位信息
通过网站档案收集二手数据	Askitas和Zimmermann（2009）	Google Insights	首先，列出4个与工作搜索相关的关键词：失业服务机构、失业率、人事咨询、德国使用最多的工作搜索引擎。其次，在Google Insights中收集第M月第1周和第2周以及第M-1月第3周和第4周这些关键词的检索频率数据	2004—2008年每月在Google上进行工作搜寻的数据

Kudlyak（2014）在2010年9月—2011年9月的观测期内，在SnagAJob.com上收集了劳动者发送第一份求职申请和最后一份求职申请的时间，并以这两个时间的差值测量工作搜寻持续时间。

就雇主搜寻而言，在空缺岗位产生之后，雇主在招聘网站上发布空缺信息、接受求职申请，并对求职者进行筛选。因此，雇主在招聘网站上发布的空缺信息的数量（Young，2012）以及空缺信息获得求职者“鼠标点击”的次数（Prabhakar等，2014）是雇主搜寻结果及效率的测量指标。Marinescu和Wolthoff（2012）在CareerBuilder.com上观测了2011年1月期间芝加哥和华盛顿的雇主刊登的招聘广告信息，收集了有关空缺岗位求职者数量的数据。他们对求职者的学历和工作经验的分析显示，空缺岗位的名称可以解释80%的学历和工作经验等任职资格要求差异。此外，雇主刊登招聘广告的起止时间是雇主搜寻持续时间的测量指标。根据雇主搜寻理论，空缺持续时间是雇主开始寻找劳动者与找到合适的劳动者（van Ours和Ridder，1991）或被雇用的劳动者开始工作的时间之差（Burdett和Cunningham，1998；Davis等，2014）。学者们将招聘网站上空缺信息发布时间与失效时间之差定义为空缺持续时间（Brenčič和Norris，2009，2012；Nivalainen，2014）。

就匹配结果而言，雇主通过招聘网站的“面试通知服务”可通知求职者是否参加面试，因此，求职者是否收到面试通知既是工作搜寻结果的衡量指标，也是雇佣歧视的测量指标。Kuhn和Shen（2014）在xmrc.com.cn上收集了237 128份求职申请数据，其中包括是否得到雇主反馈的信息，他们基于这些数据对雇佣过程中的户籍歧视进行了实证分析。此外，在在线劳动力市场中，求职者是否收到雇主的录用通知也是匹配结果的测量指标，也可用于分析雇佣过程中的歧视问题。Mill（2011）在freelancer.com上观察了向同一个岗位投递简历的求职者的个人信息，以及是否收到雇佣通知的信息，并基于这些数据，对雇佣过程中求职者的居住地歧视进行了实证分析。

表3对劳动力市场大数据的测量指标进行了汇总。

表3 劳动力市场大数据的测量指标汇总表

一级指标	二级指标	含义	举例
工作搜寻	搜寻强度	劳动者的“鼠标点击”行为轨迹和次数及相应时间	Marinescu (2014)
	搜寻持续时间	求职者在招聘网站上首次发送简历与最后一次发送简历的时间之差	Faberman和Kudlyak (2014)
雇主搜寻	搜寻强度	雇主在招聘网站上发布招聘信息数量	Young (2012)
	搜寻效率	一定时间内, 招聘广告获得求职者“鼠标点击”的次数	Prabhakar等 (2014)
	搜寻持续时间	招聘广告在网站上的刊登时间与下线时间之差	Brenčič和Norris (2009, 2012); Nivalainen (2014)
匹配结果	招聘网站面试通知	求职者是否收到雇主发出的面试通知	Kuhn和Shen (2014)
	在线劳动力市场雇佣结果通知	在在线劳动力市场上, 求职者是否收到雇主的录用通知	Mill (2011)

### 五、劳动力市场大数据的代表性及其他问题

互联网虽然降低了数据收集成本、解决了区域限制问题 (Shin等, 2012; Cook, 2014; Askitas和Zimmermann, 2015), 但是, 现阶段劳动力市场大数据仍处于发展初期, 在收集和提供有关劳动力市场运行的高质量动态数据方面, 还存在一些问题 (Kuhn, 2014)。其中, 大数据的代表性 (representativeness) 是学者们讨论较多的问题。此外, 变量与统计模型选择问题、数据涉及商业和个体隐私问题也是大数据使用过程中可能存在的问题。

#### (一) 样本代表性问题

由于大数据的采集不依赖随机抽样 (Mayer-Schonberger和Cukier, 2013), 因此数据的代表性受到不同程度的质疑。此外, 虽然互联网逐渐成为劳动力市场的主要中介, 但是传统的广告、招聘会等搜寻渠道并未完全被取代, 而这些渠道产生的搜寻与匹配数据并不包含在劳动力市场大数据中 (Capiluppi和Baravalle, 2010; Nivalainen, 2014), 这在一定程度上也导致数据的代表性问题受到了学者们的关注。

学者们通常采用两种方法来解决大数据样本代表性问题。第一种方法是将大数据与调查数据做比较: 将大数据的结构与大型问卷调查数据或普查数据的结构进行对比, 进而判断大数据的代表性。Štefánik (2012) 在采用斯洛伐克招聘网站 (Profesia.sk) 的空缺岗位数据和简历数据 (CVs) 分析高技能工人的劳动力市场隔离时, 首先认为在互联网群体 (internet population) 中, 大学生所占比例较高, 因此该数据能够代表以大学生为主体的高技能工人的需求与供给匹配情况。然后, Štefánik进一步对比了公共服务部门的技术岗和建筑业的技术岗在网络数据和全国劳动力调查 (National Labor Force Survey) 数据中的比例。在网络数据中, 公共服务部门技术岗占空缺岗位数量的6.7%, 占求职者申请岗位的4.8%; 在调查数据中, 该岗位占劳动力市场空缺岗位数量的5.7%。对于建筑业的技术岗而言, 这三个比例分别为1.1%、1.4%和1.6%。通过对比不难发现, 这两类岗位在网络数据和调查数据中的比例相差不大。进而, Štefánik认为大数据

具有较好的代表性。第二种方法是将大数据中空缺岗位数量与经济中空缺岗位总量做比较,如果前者所占比例较高,则可推定大数据的代表性较好。Marinescu(2014)在2011年1月从美国CareerBuilder.com收集岗位空缺数据,分析了延长失业救济金领取时间对失业率的影响。Marinescu将此数据与同期美国的岗位空缺和劳动力流动调查数据(Job Openings and Labor Turnover Survey, JOLTS)进行了对比,发现当月该网站发布的岗位空缺数量占美国同期岗位空缺总数的35%。而且,从岗位空缺的地理分布来看,网站数据与调查数据具有高度的一致性。基于该结论,Marinescu认为该网站的大数据对美国经济有非常好的代表性。

### (二) 变量与模型选择问题

传统问卷调查数据的收集过程是按照科学研究的方法进行的,即先提出理论假设,再收集数据,因此,问卷调查数据涉及的变量完全符合研究计划的要求。相比之下,目前基于互联网的劳动力市场大数据涉及的有关搜寻与匹配的变量则是既定的。一般而言,网站招聘广告数据包含雇主基本信息(如地址、规模、行业等)、空缺岗位基本信息(如岗位名称与类别、合同类型等)以及任职资格要求信息(如学历要求、工作经验要求、所需技能名称与等级等);求职者数据包含性别、年龄、学历、工作经验、技能(语言技能、办公软件使用、计算机技能等)、意欲进入的行业和岗位及地区等信息(Štefánik, 2012)。此外,在不同的网站上,变量的特征也不同。有的网站公布交易的绝对数,而有的网站则根据交易量公布交易排名(Edelman, 2012)。在这种情况下,研究者不仅在变量选择方面要更加慎重(Varian, 2014),而且在使用数据之前,要对变量的含义、内容等进行深入分析,以保证变量满足研究要求。

### (三) 数据涉及商业及个人隐私问题

在互联网时代,人们在网络上的检索、交易等行为会被网站记录下来。这些记录不仅能够揭示人们在日常生活中的内在选择和逻辑决策,而且,其中的一些记录在使用上是开放的(McLaren和Shanbhogue, 2011)。此外,基于网络的大数据的收集不同于问卷调查,不需要得到受访者的许可(Askitas和Zimmermann, 2015)。这些都使得大数据的收集和使用过程涉及的商业和个人隐私问题受到了关注(Edelman, 2012)。对于这一问题的处理,一方面,需要建立互联网背景下的大数据收集和使用监管体制。例如,保护隐私的责任由个体转到数据使用者(Mayer-Schonberger和Cukier, 2013),即数据使用者承担保护个人隐私的责任。另一方面,在数据使用过程中,尽量对涉及隐私的内容进行处理(Edelman, 2012)。例如,当个体姓名、联系方式等信息不是研究的核心内容时,应在数据中予以删除。

## 六、启示与展望

互联网改变了劳动者与雇主之间的搜寻—匹配机制,显著减少了劳动力市场摩擦,因而逐渐成为主要的劳动力市场中介。在此背景下,基于互联网的劳动力市场大数据应运而生。这种数据包括从搜索引擎、招聘网站、在线劳动力市场等网络主体上获得的有关搜寻与匹配的信息。由于大数据在描绘现状、预测未来等方面具有显著的优越性(Cook, 2014),学者们已将其用于失业率预测、雇佣标准分析、搜寻与匹配动态研

究、搜寻持续时间分析以及雇主偏好分析等方面的研究中。在相关实证研究中，研究者通过直接观测、跟踪监测、网站档案来收集横截面数据或追踪数据。劳动力市场大数据不仅降低了数据收集成本，也提高了数据的客观性。因此，劳动力市场大数据是调查数据的一种补充（Kuhn，2014）。值得注意的是，有关雇主搜寻行为的大数据是劳动需求研究的良好数据来源（Brenčić和Norris，2012），有助于促进劳动需求研究（Horton，2012；Davis等，2014）。然而，劳动力市场大数据仍处于起步阶段，样本代表性、变量与模型选择以及数据涉及商业或个人隐私等都是使用数据的过程中应当注意的问题。

基于上述分析，我们不难发现，大数据的出现和应用在一定程度上改变了有关劳动力市场的现代劳动经济学研究。首先，研究内容更加深入。大数据是客观观测数据，有利于推进学者对雇佣过程的分析，即推动动态研究（Autor，2001）。其次，对现行劳动力市场运行的把握更加准确。传统的劳动力市场现状描述主要基于随机样本问卷调查数据以及定期的普查数据。然而，随机样本问卷调查难以体现整体状况，定期的普查数据本质上是对历史状况的描述。相比之下，基于互联网的大数据所具有的覆盖全体、及时更新、动态变化等特征，使得其对现行劳动力市场运行状况的反映更加及时和全面。再者，实证研究方法得以扩展。传统的实证研究方法是提出假设、收集数据、进行统计分析并得出结论，而在互联网背景下，大数据是关于现行劳动力市场运行的客观观测数据，而依据已有理论对相关客观现象进行解释可能是未来研究的方向之一。最后，研究工具更加广泛和复杂。研究者利用互联网获得全体数据，那么，原有的基于随机样本数据分析的统计方法，其适用性或许就要大打折扣，更加灵活的计量模型（Varian，2014）也许是更好的选择。

自20世纪90年代中期以来，我国的互联网普及率不断攀升。中国互联网信息中心发布的数据显示，截至2015年12月，全国平均互联网普及率（不含港澳台地区）已达到50.3%，网民数达到6.88亿<sup>①</sup>。我国互联网的快速发展为学者们收集大数据奠定了坚实的基础。国内学者已采用大数据对劳动力市场进行了探索性研究。例如，中国就业研究所基于Zhaopin.com上注册的求职人数、发布的岗位空缺数量，以求职申请人数与招聘需求人数的比值定义了“劳动力市场景气指数”（CIER index）。这个指数不仅能够及时、灵敏地反映劳动力市场供求匹配状况，对就业与宏观经济变量的关系也具有较强的解释力（耿林和丁大建，2012）。Su（2014）使用Baidu.com和Google.com上人们的工作检索行为数据，构建了失业相关的在线检索强度（intensity of online unemployment-related searches）测量指标。研究结果显示，该指标能够提高宏观经济指标的预测能力。可见，国内学者对劳动力市场大数据的应用还集中在工作搜寻方面，对雇主搜寻行为的研究仍显不足。今后，研究者可以基于招聘网站大数据，以及本文介绍的雇主搜寻强度和持续时间测量指标，对国内雇主搜寻行为的特征进行分析。同时，随着“互联网+”概念的提出，互联网在生产要素配置方面所发挥的作用将进一步凸显。在此背景下，互

<sup>①</sup> 数据来源于中国互联网信息中心2016年1月发布的《中国互联网发展状况统计报告》。其中，网民是指过去半年内使用过互联网的6周岁及以上中国居民。

联网能够为学者们分析劳动力市场运行、劳动供给与需求、劳动力流动等问题,提供更客观、更直接、更具动态性的数据,有助于推进国内的劳动经济学动态研究。由此可见,网络大数据与抽样调查数据相互补充,不仅能够推动我国劳动经济学研究,尤其是劳动需求研究,而且有助于提高政府部门把握劳动力市场运行状况以及就业与失业状态的准确程度。

### 主要参考文献

- [1]蔡昉. 中国人口与劳动问题报告[M]. 北京: 社会科学文献出版社, 2002.
- [2]冯芷艳, 郭迅华, 曾大军, 等. 大数据背景下商务管理研究若干前沿课题[J]. 管理科学学报, 2013, (1): 1-9.
- [3]耿林, 丁大建. 劳动力市场的自然观察与CIER指数研究[EB/OL]. <http://www.ems86.com/touzi/html/?623.html>, 2012.
- [4] (美) Lazear E P著, 刘昕, 董克用, 译. 人事管理经济学[M]. 北京: 北京生活·读书·新知三联出版社, 2000.
- [5] (英) Mayer-Schonberger V, Cukier K著, 盛杨燕, 周涛, 译. 大数据时代: 生活、工作与思维的大变革[M]. 杭州: 浙江人民出版社, 2013.
- [6]Ackers D. The experience of EURES: Improving access to labour market information for migrants and employers[R]. Paper presented at High Level Conference, European Commission DG Employment Social Affairs and Inclusion Unit C4, 2012.
- [7]Agrawal A, Lacetera N, Lyons E. How do online platforms flatten markets for contract labor?[R]. Working Paper, 2012.
- [8]Askitas N, Zimmermann K F. Google econometrics and unemployment forecasting[J]. Applied Economics Quarterly, 2009, 55 (2): 107-120.
- [9]Askitas N, Zimmermann K F. The internet as a data source for advancement in social sciences[J]. International Journal of Manpower, 2015, 36 (1): 2-12.
- [10]Autor D H. Wiring the labor market[J]. Journal of Economic Perspectives, 2001, 15 (1): 25-40.
- [11]Autor D H, Katz L F, Krueger A B. Computing inequality: Have computers changed the labor market?[J]. The Quarterly Journal of Economics, 1998, 113 (4): 1169-1213.
- [12]Barron J M, Bishop J, Dunkelberg W C. Employer search: The interviewing and hiring of new employees[J]. The Review of Economics and Statistics, 1985, 67 (1): 43-52.
- [13]Beblová M, Kureková L M, Haita C. The surprisingly exclusive nature of medium- and low-skilled jobs: Evidence from a Slovak job portal[J]. Personnel Review, 2016, 45 (2): 255-273.
- [14]Brenčič V. Employers' hiring practices, employment protection, and costly search: A vacancy-level analysis[J]. Labour Economics, 2009, 16 (5): 461-479.
- [15]Brenčič V, Norris J B. Employers' online search: An empirical analysis[J]. Industrial Relations, 2009, 48 (4): 684-709.
- [16]Brenčič V, Norris J B. Do employers change job offers in their online job ads to facilitate search?[J]. Economics Letters, 2010a, 108 (1): 46-48.
- [17]Brenčič V, Norris J B. On-the-job tasks and performance pay: A vacancy-level analysis[J]. ILR Review, 2010b, 63 (3): 511-544.
- [18]Brenčič V, Norris J B. Employers' on-line recruitment and screening practices[J]. Economic Inquiry, 2012, 50 (1): 94-111.
- [19]Burdett K, Cunningham E J. Toward a theory of vacancies[J]. Journal of Labor Economics, 1998, 16 (3): 445-478.
- [20]Capiluppi A, Baravalle A. Matching demand and offer in on-line provision: A longitudinal study of monster.com[A]. Proceedings of the 12th IEEE International Symposium on Web Systems Evolution[C]. Timisoara: IEEE, 2010: 13-21.
- [21]Chan E S K, Swatman P M C. Electronic commerce careers: A preliminary survey of the online marketplace[A]. Proceedings of the 13th Bled Electronic Commerce Conference[C]. Bled, Slovenia: BECC, 2000: 19-21.

- [22]Cook T D. “Big data” in research on social policy[J]. *Journal of Policy Analysis and Management*, 2014, 33 ( 2 ) : 544–547.
- [23]D’Amuri F, Marcucci J. “Google it!” Forecasting the US unemployment rate with a Google job search index[R]. ISER Working Paper Series, No.2009–32, 2009.
- [24]D’Amuri F. Predicting unemployment in short samples with internet job search query data[R]. MPRA Paper No.18403, 2009: 1–17.
- [25]Davis S J, Röttger C, Warning A, et al. Job recruitment and vacancy durations in Germany[R]. University of Regensburg Working Papers in Business, Economics and Management Information Systems No.481, 2014.
- [26]DeVaro J. Employer recruitment strategies and the labor market outcomes of new hires[J]. *Economic Inquiry*, 2005, 43 ( 2 ) : 263–282.
- [27]DeVaro J. The labor market effects of employer recruitment choice[J]. *European Economic Review*, 2008, 52 ( 2 ) : 283–314.
- [28]Edelman B. Using internet data for economic research[J]. *Journal of Economic Perspectives*, 2012, 26 ( 2 ) : 189–206.
- [29]Faberman R J, Kudlyak M. The intensity of job search and search duration[R]. The Federal Reserve Bank of Richmond Working Paper No.14–12, 2014.
- [30]Fondeur Y, Karamé F. Can Google data help predict French youth unemployment?[J]. *Economic Modelling*, 2013, 30: 117–125.
- [31]Freeman R B. The labour market in the new information economy[J]. *Oxford Review of Economic Policy*, 2002, 18 ( 3 ) : 288–305.
- [32]Hellesester M D, Kuhn P, Shen K L. Employers’ age and gender preferences: Direct evidence from four job boards[R]. Humboldt-Aarhus-Xiamen Workshop, 2014.
- [33]Holzer H J. Hiring procedures in the firm: Their economic determinants and outcomes[R]. NBER Working Paper No. 2185, 1987.
- [34]Hong Y L, Pavlou P A. Are global online labor markets truly “flat” ? Global frictions and global labor arbitrage[R]. SSRN Working Paper, 2012.
- [35]Horton J J. Online labor markets[A]. Saberi A ( Ed. ) . *Internet and network economics*[C]. Berlin Heidelberg: Springer, 2010: 515–522.
- [36]Horton J J. Computer-mediated matchmaking: Facilitating employer search and screening[EB/OL]. [2012–10–31]<http://sole-jole.org/13473.pdf>.
- [37]Kuhn P. The internet as a labor market matchmaker: How effective are online methods of worker recruitment and job search?[J]. *IZA World of Labor*, 2014, 18 ( 5 ) : 1–10.
- [38]Kuhn P, Shen K L. Employers’ preferences for gender, age, height and beauty: Direct evidence[R]. NBER Working Paper No.15564, 2009.
- [39]Kuhn P, Shen K L. Gender discrimination in job ads: Evidence from China[J]. *The Quarterly Journal of Economics*, 2013, 128 ( 1 ) : 287–336.
- [40]Kuhn P J, Shen K L. Do employers prefer undocumented workers? Evidence from China’s Hukou system[R]. IZA Discussion Paper No.8289, 2014.
- [41]Kureková L M, Beblavý M, Haita C, et al. Employers’ skill preferences across Europe: Between cognitive and non-cognitive skills[J]. *Journal of Education and Work*, 2015a, doi:10.1080/13639080.2015.1024641.
- [42]Kureková L M, Beblavý M, Thum-Thysen A. Using online vacancies and web surveys to analyse the labour market: A methodological inquiry[J]. *IZA Journal of Labor Economics*, 2015b, 4 ( 1 ) : 18.
- [43]Kureková L M, Zilínčíková Z. Low-skilled jobs and student jobs: Employers’ preferences in Slovakia and the Czech

- Republic[R]. IZA Discussion Paper No.9145, 2015.
- [44]Marinescu I. The general equilibrium impacts of unemployment insurance: Evidence from a large online job board[R]. Chicago: University of Chicago, 2014.
- [45]Marinescu I E, Wolthoff R P. Opening the black box of the matching function: The power of words[A]. Proceedings of The Annual Meeting of the Society of Economic Dynamics[C]. Chicago: the University of Chicago, 2012: 1–43.
- [46]McLaren N, Shanbhogue R. Using internet search data as economic indicators[J]. Bank of England Quarterly Bulletin, 2011, 51 ( 2 ) : 134–140.
- [47]Mill R. Hiring and learning in online global labor markets[R]. NET Institute Working Papers No. 1117, 2011.
- [48]Nivalainen H. Internet-based employer search and vacancy duration: Evidence from Finland[J]. Labour, 2014, 28 ( 1 ) : 112–140.
- [49]Pallais A. Inefficient hiring in entry-level labor markets[J]. American Economic Review, 2014, 104 ( 11 ) : 3565–3599.
- [50]Parry E, Wilson H. Factors influencing the adoption of online recruitment[J]. Personnel Review, 2009, 38 ( 6 ) : 655–673.
- [51]Pellizzari M. Employers' search and the efficiency of matching[J]. British Journal of Industrial Relations, 2011, 49 ( 1 ) : 25–53.
- [52]Prabhakar A M, Oklu R, Harvey H B, et al. The radiology job market: Analysis of the ACR jobs board[J]. Journal of the American College of Radiology, 2014, 11 ( 5 ) : 507–511.
- [53]Rand D G. The promise of mechanical Turk: How online labor markets can help theorists run behavioral experiments[J]. Journal of Theoretical Biology, 2012, 299: 172–179.
- [54]Shin E, Johnson T P, Rao K. Survey mode effects on data quality: Comparison of web and mail modes in a U. S. National Panel Survey[J]. Social Science Computer Review, 2012, 30 ( 2 ) : 212–228.
- [55]Stanton C, Thomas C. Landing the first job: The value of intermediaries in online hiring[R]. CEP Discussion Paper No. 1316, 2014.
- [56]Štefánik M. Internet job search data as a possible source of information on skills demand ( with results for Slovak university graduates ) [A]. Building on skills forecasts—Comparing methods and applications[C]. Luxembourg: Publications Office of the European Union, 2012: 246–260.
- [57]Stigler G J. Information in the labor market[J]. Journal of Political Economy, 1962, 70 ( 5 ) : 94–105.
- [58]Su Z. Chinese online unemployment-related searches and macroeconomic indicators[J]. Frontiers of Economics in China, 2014, 9 ( 4 ) : 573–605.
- [59]Suhoy T. Query indices and a 2008 downturn: Israeli data[R]. Bank of Israel Discussion Paper No. 2009.06, 2009.
- [60]van Ours J C. Durations of Dutch job vacancies[J]. De Economist, 1989, 137 ( 3 ) : 309–327.
- [61]van Ours J, Ridder G. Cyclical variation in vacancy durations and vacancy flows: An empirical analysis[J]. European Economic Review, 1991, 35 ( 5 ) : 1143–1155.
- [62]Varian H R. Big data: New tricks for econometrics[J]. Journal of Economic Perspectives, 2014, 28 ( 2 ) : 3–28.
- [63]Villena-Roldán B. Aggregate implications of employer search and recruiting selection[R]. University of Chile Center for Applied Economics Working Paper, No. 271, 2012.
- [64]Wade M R, Parent M. Relationships between job skills and performance: A study of webmasters[J]. Journal of Management Information Systems, 2002, 18 ( 3 ) : 71–96.
- [65]Young S K. Applicant behavior on internet job boards: The effects of content and style[D]. Chicago, Illinois: DePaul University, 2012.

# The Application of Big Data in Labor Market Research and Prospects

Shi Zhenzhen, Zeng Xiangquan

(School of Labor and Human Resources, Renmin University of China, Beijing 100872, China)

**Abstract:** The internet has provided a novel data collecting tool for labor economics research. Because of the key advantages of internet-sourced labor market data like dynamic and timeliness, it can be used to answer questions which are difficult to study with traditional data. Based on a literature review, this paper discusses the origin process, connotation and application of the internet-sourced labor market data. Then it summarizes methods for data collecting, measuring indicators, and potential problems as well as the corresponding solutions. At last, it gives prospects for using big data to study national labor market for the purpose of promoting the application of big data in domestic labor economics research.

**Key words:** big data; labor market; internet-sourced labor market data; searching and matching

(责任编辑: 苏宁)

---

(上接第95页)

## Star Employees in Organizations: A Literature Review and Prospects

Sha Kaiqing, Yang Zhong

(School of Business, Nanjing University, Nanjing 210093, China)

**Abstract:** The survival and development of organizations are closely related to human capital owned by organizations themselves. However, the distribution of employees' ability and performance is not normal. Instead, there are usually a few excellent employees whose performance is much higher than other employees; they shine as bright as stars in organizations, have crucial, even decisive, effects on organization performance, and are the important part of organizational human capital. The research of organization high-end talents in the academia has just started. This paper reviews related literature of star employees in organizations, analyzes the characteristics and attributes of star employees, elaborates the role of star employees in organizations, discusses the management of star employees and makes prospects for future research.

**Key words:** star employee; high performer; organizational high-end talent

(责任编辑: 苏宁)