基于大数据的京沪人口流动流量、流向新变化

周晓津，姚阳
广州市社会科学院，广东 广州 510410

摘要
位置大数据为人口流动流量、流向及其变化研究提供了条件，大数据人口推断比人口普查更加精准且成本大幅度降低。基于大数据的人口流动分析表明，2015年末北京、上海外来人口规模稳中有降，外来人口来源构成与2010年全国第六次人口普查相比也发生了较大变化，来自于邻近省份的外来人口增加。在加快实施国家大数据战略的背景下，应加快共享公共数据，推进人口流动大数据研究，以尽快构建基于大数据的国家人口基础信息库。

关键词
人口流动大数据；国家人口基础信息库；国家大数据战略

中图分类号：C921 文献标识码：A doi: 10.11959/j.issn.2096-0271.2016030

Population inflow and outflow of Beijing and Shanghai based on big data analysis

ZHOU Xiaojin, YAO Yang
Guangzhou Academy of Social Sciences, Guangzhou 510410, China

Abstract
With the help of LBS big data analysis, more about population floating and its changing is know while the outcome is more accurate and its cost less than traditional population investigation or census. Analysis shows that the population size from outside province of Beijing and Shanghai remains stable with a slight decline by the end of 2015. Compared with national population census of 2010, the inflow population structure has seen a great change, the inflow population from the neighboring provinces increased a lot. To speed up the implementation of the national big data strategy, the government should promote public data sharing, attract more scholars to engage in big data analysis and help to build the country's basic population database.

Key words
big-data of population floating, national basic database of population, national big-data strategy

1 引言

虽然国内外独立成篇的有关人口流量、流向研究的学术文献较少，但是有关人口流动的可搜索学术文献数量却非常巨大，其中大部分是1980年以来的著述。在中国知网上全文检索“人口流动”就有1 270 595 条结果，以“人口流动”作为关键词检索有8 034条结果；在谷歌粉丝联合建立谷粉搜索中全文检索“人口流动”时，则可搜索出40余万篇文献（类似谷歌学术搜索）。基于同样的检索条件，以“上海人口流动”进行全文检索时有3 718条结果，而以“北京人口流动”进行全文检索时有3 709条结果，表明学术界对北京市和上海市研究的热度大致相当。尽管有关人口流动文献数量巨大，但基础人口数据来源却相当有限：来自人口普查和全国性大型人口抽样调查所占份额最大，以地区人口迁移或人口流动为专题的抽样调查甚至普查所占份额次之，而受成本制约的学者们小范围专题人口流动调查份额较少，但数据最为真实可信。

传统的人口流动研究按数据来源可分为两大类：一类是以全国的人口流动为研究对象，数据主要来源于全国人口普查和大型人口抽样调查；另一类是大城市与各地区的流动人口调查研究，通常以调查报告的形式出现。国家人口计生委流动人口服务管理司首次发布《中国流动人口发展报告2010》，截至2014年已累计出版了5本报告，其有关的调查数据现已经向国内高校和科研机构免费开放。国家卫生和计划生育委员会的调查数据包括中国大陆的所有县、区，但其最大的缺陷在于按地区均衡抽样，在外来人口聚集区域的样本偏少，外来人口比例较低区域的样本相对过多，导致抽样调查效率较低，以此推断的全国人口流动总量就会失真。即使从上海的情况来看，面对全市超千万的外来人口，只有区区8 000个样本，显然这种调查推断的上海外来人口总量会有较大的偏差。由于中国流动人口规模巨大，数据繁杂纷乱，境外学者文献数量稀少，且更多地引用中国大陆学者的数据和结论。尽管有人口普查，但从学术界到政府再到社会公众，目前为止对我国有多少流动人口等基本问题都缺乏统一、明确的答案，相同年份的流动人口数量差异极大，同一年份不同来源的数据之间差距有些也大得惊人[1]。笔者[2]以人口流动研究为出发点，系统地研究和分析了国内跨省人口流动流量、流向情况。研究结果表明，北京流动人口规模大致与上海相当，宏观经济周期波动、政策调整与政治事件对流动人口的影响也大致相同。

2 基于大数据的人口流动流量、流向研究及其进展

大数据研究主要集中在欧美发达国家和地区，相关文献主要来源于美、英、德等国以及信息技术发达的韩国、日本等国，中国是唯一挤入大数据研究阵营的发展中国家。国际上对大数据的研究主要集中在数据挖掘、可视化分析、云计算和信息检索等方面，研究内容涉及生物学、传播学等不同学科领域，由于国外人口流动多以旅游、商务等短期性流动为主，而国内则以就业性的人口流动为主，且在时间跨度、距离跨度和数量方面都远胜全球任何一个国家。因此，国内有关人口流动大数据的实证研究基本上与国外保持同步，甚至领先。胡巧玲等人[3]利用改进算法进行大数据统计的人口迁移量预测，以提高人口迁移预测的准确度。王峰等人[4]通过数据分析和数据挖掘，分析了城市人口的时空分布及

动态迁移情况。赵时亮等人[5]指出，利用手机与移动通信基站之间的广播机制，可以分析诸如住宅小区空置率和城市人口通勤的规模和流向等。李红娟[6]对大数据时代我国人口信息管理及应用进行了探索性研究。刘瑜等人[7]探讨了解释所观测移动模式的模型构建方法。童大焕①首次利用QQ大数据分析北上广深一线城市的人口流动情况。张强等人[8]利用移动通信总量数据对国内主要特大城市人口进行估计，其结果与北京、上海等城市最新调查人口相当一致。与传统依赖人口普查或人口调查的人口数据不同的是，基于大数据的人口流动研究更多地来自信息技术领域的专业人士，而传统人口学领域的研究成果将有助于大数据分类、聚类、回归以及关联等分析和判断的有效性。中国社会科学院人口与劳动经济研究所的王广州研究员认为，就目前的情况来看，我国的人口大数据的来源主要是人口普查、人口信息系统和行政登记大数据。王广州[9]根据人口数据的收集方式的不同，将中国人口大数据划分为全员人口大数据和特定人群/亚人口大数据。王广州认为，全员人口大数据主要是人口普查信息和户籍管理信息，理论上覆盖全国所有人口，是最具有权威性和长期历史积累的大数据。虽然并不认可这种人口大数据的划分方式②，但基于早前年份的大数据缺失，人口普查数据仍然不失为重要的比较研究数据来源。

社交网络大数据中，对腾讯公司QQ用户实时登录和微信用户的分析，同样可以得到比较准确的人口分布及流动数据。由于QQ用户年龄主要在18~50岁，该年龄段也与外来人口的年龄结构基本一致。因此，通过分析春节期间大规模QQ登录地域的变动，可以推算城市区域该年龄段人口流动情况。童大焕首次利用QQ大数据分析北京、上海、广州、深圳一线城市的人口流动情况。童大焕认为，包含瞬间流动人口在内，北京、上海、广州、深圳2013年底的实际人口数量并非官方公布的6 930万，而是高达16 476万。童大焕的错误在于简单地将QQ用户与人口相对应，却忽略了这样的一个关键事实：QQ活跃用户一方面可以通过电脑登录，另一方面更多地通过手机等移动用户端进行登录，而北京、上海、广州、深圳这样的一线城市该年龄段人均拥有1.5部手机。将这些关键因素考虑之后，4个一线城市18~50岁的人口估计为6 414万人，再加上这4个一线城市户籍人口中该年龄之外的人口以及外来流入人口在该年龄段之外的人口，才是这4个城市的总人口。

除了必然考虑人口结构外，利用QQ登录进行城市人口估计时，必须考虑商旅人员在某些特定地区频繁地登录现象。例如，最新汇集的QQ登录数据中（共13.87亿个登录账号），北京市东城区QQ登录的新用户记录在2015年就达到5 212万，登录的用户总数占北京全市的66.7%，如**图1**所示。上海市黄浦区登录的用户总数占上海全市的34.9%。众所周知，天安门地区日均人流量平时也有40万~50万人，节假日、国

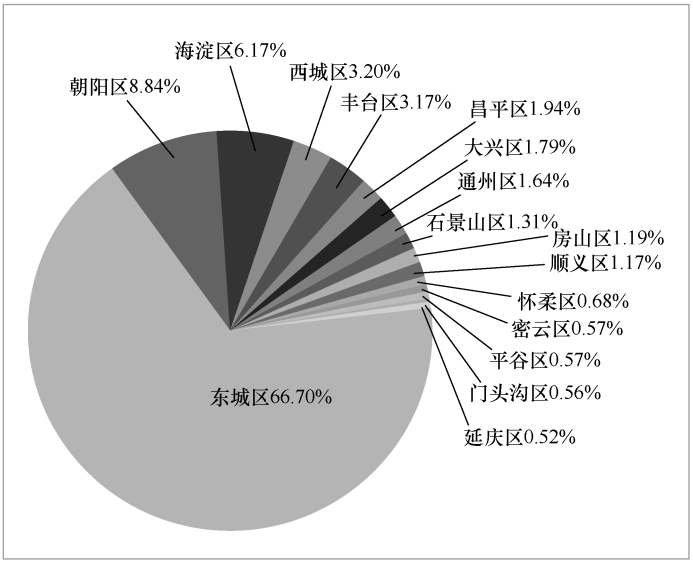

图1 北京市QQ用户在各区分布情况（2016年更新）③

① http://dajia.qq.com/blog/385280074101218.html

② 随着现代信息技术在人口普查中的广泛应用，人口普查数据越来越多地具有大数据特性，称之为“大数据化”数据

③ http://www.8ad.com/product/province/34.html

庆高峰人流量通常在100万左右，几乎所有外来游客都会到天安门。上海市黄浦区也是上海国内旅游和商务人员的首选地，广州市广交会也为广州带来大量的商旅人员。

童大焕认为，2014年春节北京、上海、广州、深圳有1 070万人（用户）永久逃离，但实际并非是逃离，因为这4个城市是中国的旅游大市，其所对应的QQ登录地址变化也在情理之中，仅通过QQ登录地址变化尚无法判断究竟有多少人永久离开这4个城市。与此相类似，春节过后大量的QQ新用户登录这4个城市，也并不意味着大量人口来大城市寻求工作机会，原因是春节期间旅游人口规模更大。例如，2015年春节7天假期，外埠来京游客142.5万人次，比2014年同期增长7.1%。2015年腾讯公司发布的报告表明，春节期间北京、上海、广州、深圳四大城市QQ用户登录变动只有1%，表明四大城市人口流动流量、流向变化基本稳定，QQ登录新增用户主要来自假日旅游人口。相较2014年春节QQ登录数量而言，1%新QQ用户登录这4个城市并不意味着外来人口的增加，其他途经的估计表明，这4个城市外来人口并非增加了，而是减少了，意味着这4个城市农民工流失的速度大于高校毕业生流入一线城市的速度。

3 人口流动大数据来源与处理方法

中国春运无疑是全球范围内最大规模的人口迁移活动，也是研究国内人口流动流量、流向变化的最佳时期。早在2014年1月25日晚间，中央电视台与百度公司合作，启用百度地图定位可视化大数据播报春节人口迁徙情况，该项目利用百度公司LBS（location based service）数据进行计算分析，展现春节前后人口大迁徙轨迹与特征。利用用户产生的位置大数据来绘制地图的方法并非百度公司独有。如Facebook公司绘制了其10亿用户全球分布地图，腾讯公司绘制了其QQ在线用户的分布图，新浪公司绘制了全球新浪微博全球签到用户位置图。绘制迁徙地图涉及空间和时间两个至关重要的因素及其变化，百度迁徙地图以地级市为最大分辨率，能够较为宏观地呈现中国春运期间人口的迁徙位置及其变化趋势，由于运算量极其巨大，2014年的百度春节人口大迁徙将时间分辨率固定为8 h。由于春节期间人们的迁徙路径空间跨度较大，有分析认为，百度迁徙地图所采用的8 h不能较为接近地反映全国人口的迁徙流动动向。

“百度迁徙”技术功能包括几方面，第一个是全国迁徙的区域带，第二个是热门线路分析，包括迁入迁出和热省分析、选定城市分析、时间维度分析。2015年更新版“百度地图春节人口迁徙大数据”上线，功能升级后包含了人口迁徙、实时航班、机场热度和车站热度四大板块。百度迁徙的动态图包含春运期间全国人口流动的情况与排行、实时航班的详细信息以及全国火车站、飞机场的分布和热度排行，通过百度迁徙动态图能直观地确定迁入人口的来源和迁出人口的去向。本文利用百度公司提供的2015年春运以及春季全国城市之间逐日、逐小时人群流动数据来推断北京和上海的人口流动流量、流向及其新变化。逐日、逐小时人群流动数据字段说明见**表1**。进行跨省人口流动流量、流向推断时，单向线路数（singleNum）和标识常量（floatFlag）两个字段并不需要，但为方便后期处理，同样将其保存在同一数据表文件中。考虑以同样的口径来构建全国跨省人口流动流入/流出平衡表，然而Excel办公软件只能处理65 536行数据，因此只采用2015年2月7-16日共10天的数据。值得注意的是，虽然只有10天的数据，但其采

表1 2015年春运及春季全国城市间逐日、逐小时人群流动数据字段说明

PtopLineIn（各省份热门迁入线路）					
province（迁入省）	name（始发地）	num（迁入热门线路数）	singleNum（单向线路数）	per（在总线路中的占比）	floatFlag（常量，无意义）
例如吉林	辽宁	209	209	0.241 3	0
PtopLineOut（各省份热门迁出线路）					
province（迁出省）	name（终点站）	num（迁出热门线路数）	singleNum（单向线路数）	per（在总线路中的占比）	floatFlag（常量，无意义）
例如吉林	辽宁	218	218	0.292 2	0

集的全国样本总量（迁徙线路总数）无论是流入还是流出都已超过2.5亿条，即使在本文的研究中，北京和上海迁出线路数量分别为2 269.754 4万条和2 259.152 1万条，远远超出北京和上海外来人口数量④，从而保证了本文推断的有效性。

春节前跨省客流主要由“外来人口返乡流”（长期在外就业人口跨省返乡过年）、“商旅流”（商务、旅游等人口短期流动）和“留守人口逆向流”（跨省流动就业人员的子女或长辈等农村留守人口流向农民工工作地团聚过年的流动）等组成。其中，最大的客流是“外来人口返乡流”，而农民工是该客流的主要群体。据国家发展和改革委员会预测，2015年春运，在2.6亿农民工中，跨省流动的农民工将达到1.6亿人⑤。由于“商旅流”具有对称性，虽然春节期间商旅流动规模小于平时，但“留守人口逆向流”可作为一种填补，因此“留守人口逆向流”和“商旅流”二流合一，即可作为日常人口流动，从而为推断春节前人口流出地区的净流量提供了方便。另一方面，相对于超千万规模的外来人口而言，京、沪两市流向市外的户籍人口基本可以忽略不计，由于京、沪两市外来人口中邻近省份占了较大比例，因此在春节期间两市净流出人口总量大致与其外来人口总量相等，而春节前净流出人口总量则等于节前流出人口总量减去节前流入人口总量。

2015年春运从2015年2月4日至3月16日。来自北京市交通委员会的信息表明，2015年春运40天，北京铁路、民航、公路省际客运进出京客流总量达3 999.23万人次。其中，铁路北京地区上下车旅客2 668.70万人次；民航首都国际机场、南苑机场累计进出港旅客1 051.82万人次；公路省际客运累计进出京旅客278.71万人次。大数据分析表明，进出京客流总量中，30.225%的客流属于日常性流动，69.775%属于外来人口返乡（入城）流。据此推算，2015年春节北京铁路、民航、公路省际客运进出京人数为1 395.23万人，非自驾车返乡入城客流占全部客流的82%，节后15天返城占88%，由此推算，节前净流出人口为1 497.32万人。2014年末，北京户籍人口1 333.4万人，由于相当一部分为外来人口落户北京，2015年春运中有20%左右的京籍人口节前出城，据此计算，2014年末北京外来人口总量为1 230.64万人，2014年末总人口规模为2 564.04万人，这表明部分咨询机构或个人声称的北京人口规模超过3 000万的结论不实。依据同样的方法，2014年末上海外来人口总量为1 185.78万人。

利用居民生活用水量估计的北京年平均实有人口如**图2**所示。同理推算，2014年、2015年北京年均实有人口2 546.9万和2 713.5万。该估算方法缺陷在于，没有考虑降水变化和人均用水量的变化，但优点

④ 有咨询机构根据迁出热门线路数推断北京拥有超过3 000万人口的结果是错误的

⑤ 长期以来我国跨省流动人口数据一直被低估，国家统计局发布的《2014年全国农民工监测调查报告》表明，我国跨省流动农民工为7 867万人，仅占2014年全国农民工总量27 395万人的28.7%。而笔者2011年的研究表明，早在2004年，我国跨省农民工数量就达到1.4亿左右，现有的大数据分析证实了其早期估计结果的可信性

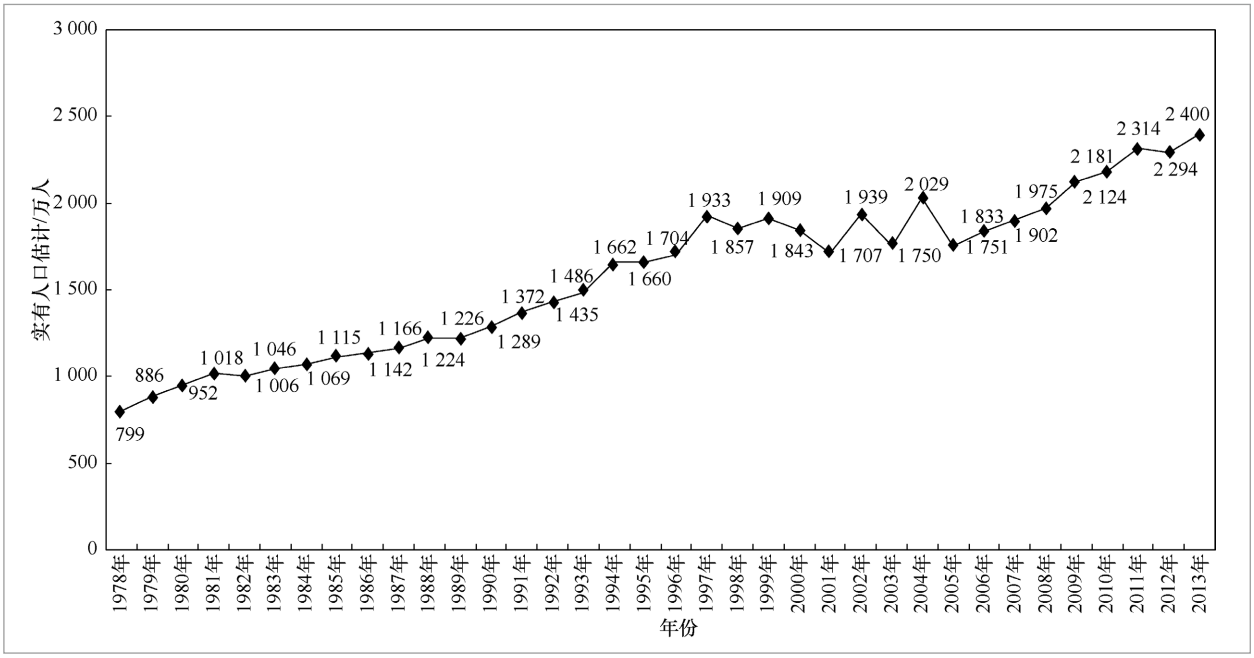

图 2 基于城市居民生活用水量的北京实有人口估计 / 万人

在于能够显示城市人口增长趋势，而非官方公布的线性均衡增长。**图2**中较好地解释了1982年、1989年、1997年东南亚金融危机、申办奥运会等外部冲击对城市人口的影响。相对于奥运会前后严格的人口控制，2008年奥运会之后人口爆发性增长实际上是北京人口在加入WTO之后的一种理性回归。2010年第六次全国人口普查时北京市常住人口为1 961.2万人，而利用居民生活用水量估计的实有人口是2 181万，相差220万人，根源在于第六次全国人口普查时外来人口的统计口径差别。2012年的人口下降源于2009年以来的宽松财政和货币政策收紧后的自然反应。

4 人口流动流量、流向变化比较分析

将北京迁入、迁出热门线路数汇总，见**表2**。第1列为节前北京迁出/迁入省级区域，该列为2015年2月7-16日共10天内的由这些省级区域迁入北京的热门线路总数，由于节前黑龙江、四川、湖北、安徽、山西等省迁入热门线路数排名没有进入前10位，采用最小均值法给予补足；第2列为迁出热门线路数，除了山西省和四川省需要补足缺省数据外，其余省份迁出热门线路数皆进入前10位；第3列为节前北京迁向各省级区域的热门线路数减去各省级区域迁入的热门线路数，得到净迁出热门线路数。计算北京实有外来人口流量时，先计算各省级区域净迁出热门线路数占全部净迁出热门线路数比例，再根据比例推算人口流动流量。在**表3**中，考虑到春节期间既有“留守人口逆向流”，也有北京本地户籍春节假期“探亲度假流”，还有部分常住外来人口留在北京过年，受条件限制，假定这种流入量和流出量相等，即假定常住外来人口在春节期间全部回到其原籍所在地过年。

与2010年第六次全国人口普查相比，北京外来人口构成并没有发生太大的变化，**见表3**。但是有个趋势非常明显，即邻近省份占比增加，如河北、山西、山东等邻

表 2 基于大数据推断的北京外来人口数量和构成（2014 年）

省份	PLineInNum（迁入热门线路数）	PLineOutNum（迁出热门线路数）	NetPLineOutNum（净迁出热门线路数）	外来人口构成	
				占比	数量/万人
河北	4 741 520	10 411 515	5 669 995	23.65%	291.0
河南	520 203	3 343 797	2 823 594	11.78%	145.0
山东	585 866	2 728 625	2 142 759	8.94%	110.0
山西	339 789	1 551 978	1 212 189	5.06%	62.3
四川	25 200	1 229 612	1 204 412	5.02%	61.8
安徽	17 777	1 185 096	1 167 319	4.87%	59.9
湖北	17 184	1 171 593	1 154 409	4.82%	59.3
黑龙江	28 821	1 149 553	1 120 733	4.67%	57.5
内蒙古	385 018	1 195 663	810 644	3.38%	41.6
江苏	308 552	959 489	650 937	2.72%	33.5
陕西	31 248	529 872	498 624	2.08%	25.6
辽宁	319 978	709 852	389 874	1.63%	20.1
天津	875 213	1 238 671	363 458	1.52%	18.7
其他	2 189 028	6 954 913	4 765 886	19.88%	244.7
合计	10 385 396	34 360 230	23 974 833	100%	1 230.6

表 3 北京外来人口数量和构成及其变化

省份	2010年全国第六次人口普查		2015年春运大数据推断	
	占比	数量/万人	占比	数量/万人
河北	22.1%	155.9	23.65%	291.0
河南	13.9%	98	11.78%	145.0
山东	8.5%	59.8	8.94%	110.0
山西	3.8%	26.9	5.06%	62.3
四川	4.6%	32.5	5.02%	61.8
安徽	6.1%	43	4.87%	59.9
湖北	4.7%	33.5	4.82%	59.3
黑龙江	5.7%	40.3	4.67%	57.5
内蒙古	3.4%	24.1	3.38%	41.6
小计	72.9%	514.0	72.2%	888.4
其他	27.1%	190.5	27.8%	342.2
合计	100%	704.5	100%	1 230.6

近省份流入北京的人口数量和占比都增加了⑥。湖北和四川的占比和数量的增加很可能仅仅只是一种虚假的表象，实际人口很可能并没有增加，且以农民工为代表的劳动力数量甚至可能减少。相对于耗费大量人力物力的人口普查而言，由于采样更接近随机，且样本量更接近总体，因此由大数据推断的跨省流动人口构成更为准确和可靠。各省实际来京人口数量与第六次全国人口普查差距极大，其中约55%的差距是由于第六次全国人口普查外来人口的统计口径比较小，另有45%的差距是由于第六次全国人口普查注重劳动力人口统计，非劳动力人口的统计存在较大的误差，而大数据推断则是所有的外来人口。

基于同样的方法，推算了上海外来人口数量和构成，见**表4**。结果表明，除了浙江、山东之外，上海跨省外来人口中大部分来自全国主要人口流出省份；江苏的情况比较特殊，由于苏北相对苏南而言发展较为滞后，流入上海的江苏人大部分来自苏北地区，而苏南地区历史上就与上海在经济上和人口流动上往来密切，日常性人口流动频繁；浙江的情况与江苏比较类似，不同的是浙江区域发展比较均衡，外出就业农民工占比远低于江苏。总体而言，上海外来人口来源广泛，大体上呈现沿海（往北辐射至山东，往南向浙江与福建辐射）、沿江（主要是沿长江流域辐射）、沿线（沿沪昆线辐射浙江、江西、湖南、贵州等省；沿京沪线辐射江苏、皖北、山东；沿沪深线向浙江、福建等南向辐射）由内向外依次辐射。随着中国高铁建设的深入，未来由上海至合肥—信阳—南阳—西安线路，则上海对河南的辐射力将大为增强；其次是由上海出发，经南通—盐城—连云港—威海高铁线路亦将增强上海的辐射力。

与2010年第六次全国人口普查及2000年第五次全国人口普查相比，上海外来人口主要来源地省份并没有变化，见**表5**，但是占比较第六次全国人口普查增加1个百分点；其次是四川所占比例下降最大，表明川渝经济区在全国的地位迅速增

⑥值得注意的是，由于第六次全国人口普查时很多自雇性外来人口并没有纳入统计范围，而在实有人口估计和大数据推断中这部分人口得以显现出来，因此并非是真正的外来人口增量。以河北省为例，第六次全国人口普查时在155.9万流入北京的常住外来人口中，并没有包括那些自雇性或其他服务业从业人口。虽然大数据推断表明，来自河北的外来人口比第六次全国人口普查多出104.2万，实际新增人口估计只有50万左右，山西、内蒙古的情况也是如此

表4 基于2015年百度春运大数据推断的上海外来人口数量和构成（2014年）

省份	PLineInNum（迁入热门线路数）	PLineOutNum（迁出热门线路数）	NetPLineOutNum（净迁出热门线路数）	外来人口构成	
				占比	数量/万人
安徽	468 341	5 645 703	5 177 362	24.90%	295.3
江苏	3 541 307	7 419 278	3 877 971	18.65%	221.1
河南	178 056	2 343 338	2 165 282	10.41%	123.4
江西	143 407	1 652 985	1 509 578	7.26%	86.1
湖北	108 325	1 418 151	1 309 826	6.30%	74.7
浙江	1 614 857	2 603 501	988 644	4.76%	56.4
山东	195 364	1 165 213	969 849	4.66%	55.3
湖南	14 803	732 783	717 980	3.45%	40.9
四川	95 729	791 018	695 288	3.34%	39.6
福建	139 130	818 717	679 587	3.27%	38.8
其他	936 833	3 636 381	2 699 547	12.98%	153.9
合计	7 436 153	28 227 068	20 790 915	100.00%	1 185.8

表 5 上海外来人口数量和构成及其变化

省份	2014年比重	排名	2010年比重	排名	2000年比重	排名
合计	100%		100%		100%	
安徽	24.9%	1	29%	1	32.2%	1
江苏	18.65%	2	16.8%	2	24%	2
河南	10.41%	3	8.7%	3	4.1%	6
四川	3.34%	9	7%	4	7.3%	4
江西	7.26%	4	5.4%	5	6%	5
浙江	4.76%	6	5%	6	9.9%	3
湖北	6.3%	5	4.5%	7	2.7%	8
山东	4.66%	7	4.2%	8	2.1%	9
福建	3.27%	10	2.9%	9	2.8%	7
湖南	3.45%	8	2.5%	10	1.4%	10
其他省份	12.98%		14%		7.5%	

强；受产业转移影响，安徽占比则持续下降，但仍旧占据第一位；沪昆线上的江西、湖南占比提高，实际反映了广东产业转移后的两省外出人口流向多元化；江浙地区占比下降也是区域经济均衡化发展和产业转移的必然结果。与北京类似，虽然外来人口总量增加，但劳动力人口减少，特别是外来农民工绝对值在减少。

5 结束语

在上述流动人口来源的推算中，并没有加入距离衰减系数，从而会导致四川、河南、安徽、湖南等离京沪较远的人口流出大省的流量和占比，相对邻近京沪的省份而言会有一定程度的低估，但这种低估在推算人口流出大省的时候会有一定程度的抵消。在编制跨省人口流动平衡表时，这种邻近省份的高估和相距较远省份的高估就会表现出来，但这恰恰为推算人口流入目的地省份在春节时返乡过年的比率及留在流入地过年的比率及其数量情况提供了方便。通过取均值等方法，过去只能依靠人口普查或全国性的抽样人口普查才能编制的跨省人口流动平衡表，在利用大数据之后变成了可能和现实，大数据的应用价值得到了极大的体现。另一方面，上述推算是以净流入为参照计算人口来源省份数量及占比，这实际上给出了人口流动流量的下限，而分别计算流入或流出人口数量及占比时，则可视为人口流动流量的上限，结合距离衰减系数，可以对跨省人口流动进行更精确的分析。

基于大数据的城市人口规模估计可以很好地研究不同年份人口变动及其背后变动因素，而官方提供的常住人口数据则是一条完美的近乎线性增长的曲线，更无法反映人口变动的原因。研究发现，基于城市人均生活用水量（日人均生活用水量取2010年以来的均值：114.67升/日）估计的2014年上海总人口规模比2013年减少了25万人。最新发布《2015年上海市国民经济和社会发展统计公报》表明，2015年末上海市外来常住人口981.65万人，同比下降1.5%（14.77万人），而基于城市人均生

活用水量估计的2015年上海人口规模比2014年减少16.67万人。虽然人口变动相差1.9万人，但相对于上海2 500万人口规模而言，无论是绝对值还是相对值都处于可接受的范围之内。2009-2013年的4年里，上海人口累计增加155万人，年均增加38.75万人，2013年以来上海人口规模进入下降阶段。上海人口规模的下降一方面可能来自于政府再一次严厉推行的人口控制政策，另一方面可能是高涨的生活成本及投资率下降所导致的农民工群体再一次大规模的逃离。笔者估计，2014年全国农民工较2013年减少555万人，而上海农民工约占全国跨省农民工的4.71%，同比例推算，上海同期农民工减少人数应为26.1万人，与基于城市人均生活用水量所计算的人口减少相当接近⑦。表明在全国农民工供应减少的大趋势下，以上海为代表的特大城市并没有劳动力吸引优势，另一方面也说明随着老龄化时代的到来，特大城市无法像以往那样依靠吸引外来年轻人口来优化人口结构，由此表明，特大城市的人口控制将可能对其造成长期的负面影响。

特别需要指出的是，人口大数据是其他大数据的基础和核心。严格说来，纲要提出的政府数据充其量只是大数据化的数据，而非真正意义上的大数据，仅仅依靠政府数据进行跨部门共享校核，所得到的国家人口基础信息库只能是大数据化的数据。距离真正意义上的大数据生成动态化、实时化、大容量化，还有相当大的差距。政府公布的人口数据在获取的过程中往往需要付出巨大的成本。以人口普查为例，每5年一次的人口普查需要支付纳税人500亿元的直接成本，平均每年的直接成本就在100亿元以上。这种基于人力进行的普查会无可避免地存在各种各样的误差，而基于移动通信、交通和社交网络的人口流动大数据分析，将有助于提出更为精细和准确的人口基础信息，且这些基础信息可以进行动态化、实时化的更新和可视化呈现。无论是从培养大数据人才还是推进国家治理现代化角度，都需要移动通信、交通和社交网络等运营商向高校和科研机构提供尽可能过滤了私人信息的可用于大数据研究的数据接口。国家人口基础信息库的信息生成依旧离不开企业数据，特别离不开对移动通信、交通和社交网络等大数据的分析。无论是商业类国有企业还是公益类国有企业，抑或是以BAT为代表的私营企业，向公共智库、大数据研究人员和社会公众提供实时的活动用户数量、QQ活动用户登录量、微信用户登录数量等统计信息，并不构成用户私人信息泄密，而这些信息对于大数据分析却极为重要，对于完善国家人口基础信息库的工作显得尤其重要和关键。国家应以立法的形式要求这些企业提供数据，并给予这些企业相当形式的补偿或税收减免。高校或科研机构也应明确数据需求。国家可设立大数据基础研究数据基金，基础研究数据向高校和科研机构开放。

⑦ 实际上是完全吻合，其1.1万差值可解释为1 000万常住外来人口的子女或老人来沪定居的增量

参考文献:

[1] 段成荣. 中国流动人口研究[M]. 北京：中国人口出版社，2011.
DUAN C R. Study on the floating population in China [M]. Beijing: China Population Press House, 2011.

[2] 周晓津. 劳动力流动视野下的中国区域经济增长研究[M]. 北京：经济科学出版社，2011.
ZHOU X J. Study on China regional economic growth from the perspective of labor flow [M]. Beijing: Economic Science Press, 2011.

[3] 胡巧玲，茹金平. 基于大数据分析的人口迁移量预测模型仿真[J]. 计算机仿真，2014，31(10)：246-249.

HU Q L, RU J P. Population migration quantity simulation and forecasting based on the big data analysis[J].Computer Simulation, 2014, 31(10): 246-249.

[4] 王峰, 唐美华. 基于移动通信大数据的城市人口管理解决方案[J]. 移动通信, 2014, 13(13): 38-41.
WANG F, TANG M H. Management solution of urban population based on mobile communication big-data analysis [J]. Mobile Communications, 2014, 13(13): 38-41.

[5] 赵时亮, 高扬. 基于移动通信的人口流动信息大数据分析方法与应用[J]. 人口与社会, 2014, 30(3): 20-26.
ZHAO S L, GAO Y. Big data migration analysis method and application based on mobile communication[J]. Population and Society, 2014, 30(3): 20-26.

[6] 李红娟. 大数据时代下的人口信息管理及应用探析[J]. 现代管理科学, 2014(10): 111-113.
LI H J. Population information analysis on the management and application in the era of big data[J]. Modern Management Science, 2014(10): 111-113.

[7] 刘瑜, 康朝贵, 王法辉. 大数据驱动的人类移动模式和模型研究[J]. 武汉大学学报: 信息科学版, 2014, 39(6): 660-666.
LIU Y, KANG C G, WANG F H. Towards big data-driven human mobility patterns and model[J]. Journal of Wuhan University: Information Science Edition, 2014, 39(6): 660-666.

[8] 张强, 周晓津. 我国大城市人口规模估算与调控路径选择[J]. 西部论坛, 2014(2): 1-16.
ZHANG Q, ZHOU X J. Population size estimation and control path selection of China's large cities[J]. Western Forum, 2014(2): 1-16.

[9] 王广州. 大数据时代中国人口科学研究与创新[J]. 人口研究, 2015(5): 15-26.
WANG G Z. Research and innovation in the population science of China in the era of big data[J]. Population Research, 2015(5): 15-26.

作者简介

周晓津(1971-), 男, 博士, 广州市社会科学院研究员, 主要研究方向为人口与城市经济学、高铁经济学和大数据应用。

姚阳(1979-), 女, 广州市社会科学院经济学副研究员, 主要研究方向为区域发展与地方治理、城市经济。

收稿日期: 2016-03-14
基金项目: 2015年国家社会科学基金一般项目"基于大数据的人口流动流量、流向新变化研究"(No.15BRK037)
Foundation Item: General Project of National Social Science Fund 2015 "Research on Population Migration, Population Flow and New Change of Diretions Based on Big Data" (No.15BRK037)