# 基于大数据的住宅自动评估系统的构建

侯纯涛 徐萍萍

**【**摘要**】**本文提出了基于大数据、数据分析技术支持下，以住宅小区为评估单位，批量实现住宅房屋自动评估的思路和数学模型，同时对模型中的常用评估参数的实现提出了见解，最后对自动评估系统的功能架构做了初步的探究。

**【**关键字**】**大数据；住宅；自动评估系统

## 一、引言

21世纪以来，随着互联网技术的高速发展，人类迎来一个信息爆炸的时代。在此背景下，大数据（Big Data）及数据挖掘（Data Mining）等相关概念也应运而生。大数据时代的到来在各个领域都开启了一场革命，已成为各行业转型的重要驱动力。在中国著名的互联网企业家马云看来，“未来最大的能源是数据”①， “未来30年大数据将重新定义变革和行业”②

房地产估价行业是一个与数据息息相关的行业。从评估技术的角度来看，无论是房地产市场的分析，还是估价中选用交易案例、资本化率的确定等，均离不开大数据的支持。目前，由于商品房交易市场的活跃，海量的房地产数据通过互联网交互共享，房地产交易中心及房地产中介机构中每天也产生大量的交易数据和挂牌数据，这些数据的产生为房地产估价的大数据分析和应用提供了较为成熟的条件。在众多的房地产数据中，又以住宅相关数据最多。故而笔者从住宅评估为切入点，探讨大数据在实现房地产自动评估系统中的应用。

房地产自动评估系统是在大数据的支持下，脱离人工干预的或者只有少量人工干预的，实现批量的、精确的、自动的估价系统。大数据的发展已经为房地产自动评估系统的实现提供了可能。自动评估系统在金融、税务等领域具有广泛的应用前景。

## 二、基本思路和数学模型

传统的估价程序，包括受理估价委托、搜集估价所需资料、选用估价方法进行测算、撰写估价报告等主要过程，这种方式需要估价师对估价过程的各细节进行完全的掌控，从估价案例的选取，到评估参数的确定，都需要估价师亲自确定，这种精细化、小作坊式的作业方式难以适用于自动评估系统的快速、自动、批量计算的要求。

目前，由于住宅交易市场的活跃，互联网上与交易中心存在大量的可利用的房屋价格交易数据。对这些交易数据有目的的搜集、分析，提取主要的评估参数，简化评估过程，辅助计算机编程技术，最终实现自动化评估。

在大数据的支持下，可以以住宅小区为基本单位，吸取传统评估方法中的比较和基准价格调整的思想，简化评估过程，以小区基准价格为基础，经过因素调整得到小区内各房屋的价格。

用数学公式表示为：

 

上述公式中：-待估房屋的价格

 -小区基准价格

 -区位因素修正

 -个别因素修正

 -时间修正系数

分别为区位因素和个别因素修正系数，假设有n个区位修正系数，修正值分别为，采用累积的方式，则。同理，假设有m个个别因素修正系数，修正值分别为，则。

如前文分析，住宅小区内房屋价格的评估简化为小区基准价格P、区位因素修正、个别因素修正、时间修正系数四个因素的确定过程，而这四个修正因素的确定，可以在大数据的支持下，采用回归分析等数学方法得到，下文将详细论述。

## 三、小区基准价格

### （一）、内涵

所谓“小区基准价格”，是指在住宅小区中，选取某一“条件居中”的房屋在价值时点的市场价格。它是其他房屋评估时的参考价格，一般选择住宅小区中居中位置的楼栋居中楼层的房屋。

小区基准价格，对于不同的建筑类别，选择的楼层不同。如多层住宅小区，一般选择2层-4层的房屋；对于高层建筑（共N层），选择N/2或N/2+1的房屋。对于有多种建筑类别的小区，如多层和高层楼房混搭的住宅小区，也可以定义两个基准价格，分别对应多层楼房和高层楼房。



图1 某典型住宅小区

例如所示的某典型住宅小区，该小区由编号A~I 9栋多层建筑组成，可选择位置居中的楼栋E的3层房屋的价格作为小区基准价格。

### （二）、小区基准价格的确定

小区基准价格的确定一般有如下两种方法。

（1）通过近期该小区大量的成交案例或挂牌案例统计分析得到。根据小区基准价格的定义，小区基准价格是小区内“条件居中”的房屋在价值时点的市场价格。所以可以近似的认为通过统计分析得到小区平均价格等于小区基准价格。需要特别说明的是，如果采用的案例是挂牌案例，其价格一般会偏高，需要估价师对最终的结果进行调整。

如图2所示，为某小区2016年1月至3月成交价格散点图，其趋势线是一条近似水平的直线，故可认为这一段时间小区的成交价格是平稳的，可以将这些案例的平均价格作为小区基准价格。

图2 某住宅小区3个月成交价格散点图

（2）通过估价师评估得到。对于一些成交较少或者市场价格波动较大的住宅小区，其基准价格需要评估师评估得到，数学方法可以作为辅助验证的手段。

## 四、区位修正因素和个别修正因素

由于本系统的评估模型是以小区为基本单位，故而可以最大限度的减少房屋价格的修正因素。需要特别考虑的一些区位和个别因素修正如下表：

表1 小区因素修正举例

|  | **修正因素** | **备注** |
| --- | --- | --- |
| 区位因素 | 临路 |  |
| 临景观 | 如河流、假山等 |
| 其他 | 如便利店、垃圾站、高压线等因素造成房产价值的增减 |
| 个别因素 | 建成年代 |  |
| 楼层 |  |
| 朝向 |  |
| 户型 |  |
| 建筑面积 |  |
| 其他 | 其他可能的加分项、减分项 |

确定修正因素后，需对因素指标进行量化，以构建修正体系。

区位修正因素量化。由于以小区为评估基本单位，基准房屋为同小区的房屋，房屋区位因素的差异在整体上体现为楼栋之间的差异。因此区位修正因素的量化有两种思路。一是按照表一的示例分项目编制修正体系，对各修正因素分别专家打分，确定最终的修正值；二是整体求取修正系数，表现为“楼栋比价关系”。楼栋比价关系的获得也有两种方法。一是数据分析法。这种方法是利用小区的历史成交数据，通过回归分析的方法，得到栋与栋之间的比价关系。二是Delphi打分法。即通过数个资深评估师打分得到比价关系。最终得到的成果如下图所示（以基准楼栋的系数为1）。



图3 区位修正因素（楼栋比价关系）

个别修正因素量化：楼层、朝向等个别修正因素的获得也有两种方法。例如楼层修正系数的量化。一是通过历史成交数据回归分析获得（如图4）。采用这种方法，一般选择交易时间较集中的一手房成交数据。二是参考当地相关部门公布的修正系数（如表2）的参考值，或者专家打分得到。

图4 某住宅小区楼层和成交价格的回归关系

表2 某部门公布的楼层修正系数（ 多层住宅）

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 总层数层次 | 一 | 二 | 三 | 四 | 五 | 六 | 七 |
| 1 | 0% | 0% | 0% | 0% | 0% | 0% | 0% |
| 2 |  | 2% | 2% | 2% | 2% | 2% | 2% |
| 3 |  |  | 1% | 4% | 4% | 4% | 4% |
| 4 |  |  |  | -1% | 4% | 4% | 4% |
| 5 |  |  |  |  | -2% | 2% | 2% |
| 6 |  |  |  |  |  | -3% | 0% |
| 7 |  |  |  |  |  |  | -4% |

## 五、时间修正系数

经过上述步骤，已建立房地产价值与各修正因素之间的函数关系。时间修正系数是从纵向调整房地产价格的手段。时间修正系数的量化，一方面可以直接参考当地房地产价格研究部门定期公布的房地产价格指数，例如上海二手房指数办公室（http://www.ehomeday.com/）定期公布的上海二手房指数报告。也可以以小区为单位，通过计算机分析各时点的房屋成交案例获得比价关系。

## 六、功能架构

从系统开发的角度来看，住宅自动评估系统的功能架构应分为四个层次（图5）。

首先是数据层。数据库建库是整个系统的核心部分。数据的搜集与整理也是系统运行的日常工作。房地产成交数据的来源有多个途径，可以使用爬虫软件从网络抓取大量的属性数据和挂牌信息，也可以来自房地产中介机构及房地产交易中心的共享。原始的数据经过必要的处理后，构成系统的底层核心数据库。

其次是计算层。计算层的主要任务是构建数学模型，对底层数据进行分析、测算，得到小区基准价格、修正因素和修正值以及必要的成果校验。

再次是基本应用层。这一层次是向客户展示系统功能的部分。是后台数据的前端体现。应包括基本的信息展示、自动估价以及房地产价格趋势分析的功能。

最后是功能拓展层。对于一些有特殊功能需求的用户，可定制开发符合其要求的功能，也可以与内部OA办公系统相衔接。



图5 住宅自动评估系统功能架构

## 七、结语

目前，市面上已经有不少公司对房地产自动估价系统的开发做了有益的尝试，也有较为成熟的产品出现，在笔者看来，房地产自动估价系统进一步发展仍存在诸多的难点：

（1）国内的房地产成交数据开放程度不够。由于国情的原因，国内大部分的真实成交数据在房地产交易中心及房地产中介手里，而这些部门的数据是不对外公开的，评估机构要想使用这些数据通常需要通过特别的途径。笔者建议由政府牵头，组建市级的房地产成交数据库，并有条件的共享给各大评估机构使用，以进一步促进房地产交易市场的公开化、透明化。

（2）目前市面上的房地产自动评估系统虽多，但对于系统的准确性，缺少一个统一的评价标准，评估结果没有公信力，同时也缺乏有效的监管，无法代替评估报告的作用，系统的应用受到很大的掣肘。

（3）系统的开发维护成本较大，单个中小评估机构难以承受。据笔者了解，有兴趣开发自动评估系统的机构并不在少数。但由于系统开发维护成本较高，且应用前景尚不明朗，导致很多中小评估机构望而却步。笔者认为可以由协会牵头，联合中小评估机构，采用抱团的方式分摊成本。

参考文献

（1）马云.在2015贵阳国际大数据产业博览会的演讲稿.[EB／OL] <http://news.xinhuanet.com/politics/2015-05/28/c_127851327.htm>

（2）马云. 在2017年中国国际大数据产业博览会的“机器智能”高峰对话会上的演讲. [EB／OL] <http://tech.qq.com/a/20170526/047688.htm>

（作者单位：上海上睿房地产估价有限公司）