

# 利用大数据技术分析 IT 行业薪酬的变化

王龙灿<sup>1</sup>

1 大连理工大学软件学院 辽宁省 大连市 116000

**摘要** IT 行业自出现以来便吸引着众多人士的关注，其中关于 IT 行业薪酬这个敏感性话题更是会引起人们热议。在收集到近些年来 IT 行业薪酬数据的情况下，利用大数据技术对 IT 行业薪酬数据进行挖掘，并对接下来几年 IT 行业薪酬的前景做出了一定的预测。大数据采集用于收集相关数据，将所收集到的数据进行预处理后存储，使用多种方式对数据分析挖掘。利用大数据技术对 IT 行业薪酬变化的分析结果可以看出，IT 行业薪酬在持续增长，但也显示出两极分化严重等一些比较奇妙的现象。结合其他因素和这些现象可以发现，IT 行业薪酬要高于多数其他行业，但同时 IT 行业门槛高等多重因素导致多数 IT 人士拿着一个不上不下的薪酬。预计随着 IT 行业的成熟，在未来该行业薪酬将大致稳定，和现在相比会大概高出 10%到 15%。

**关键词：** IT 行业薪酬；大数据分析；IT 行业发展趋向

## 1 引言

IT (Information Technology) 行业一般指信息技术产业，又称信息产业，是运用信息手段和技术，收集、整理、储存、传递信息情报，提供信息服务，并提供相应的信息手段、信息技术等服务的产业。IT 行业作为一个新兴行业，从出现以来便受到广泛关注。薪酬问题是一个敏感性话题，IT 行业的薪酬更能引起大众的关注。因此，本文利用大数据技术对 IT 行业薪酬进行了分析，并做出了未来几年内 IT 行业的发展趋势。

## 2 分析结果

从分析结果来看，IT 行业薪酬在这 20 年来一直保持着增长的态势，相比而言，前五年的增长速度不及之后十五年增长速度的 10%，结合时代背景可以发现，在国家决定大力发展高新技术产业的目标下达之后，IT 行业便步入了高速增长时期，薪酬随之飞速提升也是理所当然的事情了。另外分析结果还显示出了一些比较有趣的现象，IT 行业虽说整体来看横向对比其他行业拥有着高薪酬，但是单说 IT 行业内部薪酬分布，两极分化现象极其严重，业界精英人士月薪可以达到十万元甚至更高，但 IT 行业低收入人群月薪只有几千元，中层人士月薪多数都在一万元以上。结合其他相关数据不难看出，IT 行业虽然热门，但它也是一个高门槛行业，真正拥有技术的人员享有高薪酬和高待遇，技术水平低的人只能勉强度日。

通过预测行分析，我们可以看到 IT 行业的持续火爆虽会有所降温，但它依旧是热门行业，吸引着源源不断的人进入其中。在接下来的几年的，IT 行业前景仍旧不错，薪酬方面也符合人们对它的预期。另外，预测结果显示，IT 行业从业人员素质将会不断提高，未来的 IT 行业将会在真正意义上成为一个高薪酬行业，而不是现在的薪酬分布两极分化现象严重的局面。

## 3 分析方法

大数据分析最近被广泛应用于各个方面。大数据技术将被设计用于在成本可承受(economically)的条件下，通过非常快速(velocity)的采集、发现和分析，从大量化(volumes)、多类别(variety)的数据中提取价值(value)，将是 IT 领域新一代的技术与架构[1]。本文应用大数据技术，按照 Hadoop 数据处理流程（见图 1）对 IT 行业薪酬进行分析处理。

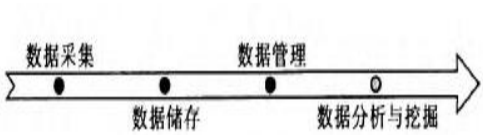


图 1 Hadoop 数据处理流程

### 3.1 大数据采集

首先是数据采集，我们在官方网张上获取了近 20 年 IT 行业薪酬的具体数据以及可能会影响到薪酬的其他数据，然后主要采用关系型数据库 MySQL

存储这些获取到的数据。

### 3.2 大数据预处理

在数据分析之前,我们对先前获取到的数据就进行了数据预处理,对原始数据进行包括但不限于清洗、填补、一致性检验的一系列操作,主要分为四个部分:数据清理、数据集成、数据转换、数据规约。

### 3.3 大数据存储

至于大数据存储这一存储器以数据库的形式存储采集到的数据的过程,我们主要采用基于Hadoop的技术扩展和封装这一方式。通过利用Hadoop开源优势以及善于处理非结构、半结构数据、复杂的ETL流程、复杂的数据挖掘和计算模型等特性,针对所收集到的数据进行处理。

### 3.4 大数据分析挖掘

在大数据分析挖掘这一过程中,我们从可视化分析、数据挖掘算法、预测性分析、数据质量管理等方面对现有的数据进行萃取、提炼和分析。

其中我们着重使用了可视化分析、数据挖掘算法和预测性分析。通过可视化分析我们借助图形化手段,清晰有效的展示出近20年来IT薪酬变化情况及相关因素之间的关系,简单明了、清晰直观。

通过数据挖掘算法我们意识到了IT薪酬变化的深层次原因,在这个阶段我们从不同侧重点创建了数据挖掘模型,利用模型对数据进行试探和计算,

## 5 参考文献

[1]刘蔚然, 刘莉娜. 大数据技术[J]. 冶金设备管理与维修, 2014, 32(004):33-36.

寻找到了分析结果的最佳模型参数。将这些参数用于整个数据集,我们很轻易的便得到了预期的数据之间存在的结果。

经过了数个步骤之后,预测性分析便相对简单了许多。我们发现IT行业在未来依旧有着极高的人气和与之对应的薪酬。

## 4 结论

从分析结果来看,近20年IT行业薪酬不断上涨,但公众切勿盲目跟风进入IT行业,究其根本,IT行业仍旧是一个比较看重个人素质的一个行业,在没有知识背景的情况下直接进入IT行业,可能会得到一个更坏的不如从前的生活。不过数据同样显示,IT行业薪酬在接下来的几年依旧有着不断上涨的潜力,若是想要谋求一个更好的职业发展,在学好行业基本功的情况下成为IT行业中的一员倒是不失为一种良策。IT行业门槛是高,但在时代背景下它有着极高的上限,可以说IT行业入门难、上升难,但一旦有了真才实学,这对想要获得高薪酬的人来说也是一个极好的机会。总之,切莫一时眼热上头忽略自身实际情况投身IT行业。

从此次大数据技术在IT薪酬分析方面的使用情况来看,大数据技术在分析数据方面有着极高的洞察力,相比于传统老套的数据分析方式有着速度快、覆盖广、信息利用率高等优势。