

# Person Service Items Fit Based on the Novel KNN Algorithm in Household Service Industry

Hongying Zhu, Fengtao Liu

Glorious Sun School of Business and Management (GSSBM), DongHua University, Shanghai  
Email: 17724806683@163.com

Received: Dec. 21<sup>st</sup>, 2019; accepted: Jan. 3<sup>rd</sup>, 2020; published: Jan. 10<sup>th</sup>, 2020

---

## Abstract

In the Internet era, the domestic service industry generally has problems such as low customer satisfaction and low professionalism of housekeeper, which hinders the development of the industry. Under these circumstances, the reasons for this can be found that the mismatching between the domestic housekeeper and the service items is the most fundamental reason. Based on this background, this paper uses KNN, a distance-based algorithm in the big data environment, to match the housekeeping staff with the service project, and improve it by the sample distance weight based on the traditional KNN algorithm. An improved person service items fit model is obtained. Experiments show that the improved person service items fit model has an accuracy rate of 69.36% compared with the traditional person service items fit model based on the traditional KNN algorithm. The classification result is better and the error rate is lower. It can match the housekeeper and the service project well and promote the professional training of housekeeping staff, so that it will enhance customer satisfaction and promotes the long-term development of the domestic service industry.

## Keywords

Person Service Items Fit, KNN Algorithm, Housekeeping Service, Accuracy Comparison

---

# 基于改进的KNN算法的家政服务行业 人单匹配

朱虹影, 刘峰涛

东华大学旭日工商管理学院, 上海  
Email: 17724806683@163.com

## 摘要

互联网时代下家政服务行业普遍存在顾客满意度低, 服务人员专业性低等问题, 阻碍了行业发展的脚步。探究其中的原因可以发现家政服务员与服务项目不匹配是其最根本的原因。基于这一背景, 本文利用大数据环境下KNN这一距离类算法对家政服务员与服务项目进行人单匹配, 并在传统KNN算法的基础之上通过样本距离权重将其进行改进, 得到改进的人单匹配模型。实验表明, 改进的人单匹配模型相对于基于传统KNN算法的人单匹配模型, 准确率达到69.36%, 分类结果更好, 误差率更低, 可以很好地将服务员与服务项目相匹配, 促进家政服务员的专业化培训, 提升顾客满意度, 推动家政服务行业的长远发展。

## 关键词

人单匹配, KNN算法, 家政服务, 准确率比较

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

### 1.1. 研究背景

随着经济的快速发展和人民生活水平的不断提高, 家政服务的需求不断扩大, 推动了家政服务行业的蓬勃发展。但是在家政服务业蓬勃发展的同时, 员工流动性强, 供需不平衡, 顾客对家政服务平台缺乏黏性, 顾客满意度, 忠诚度低等问题也日渐凸显, 严重制约了家政服务行业的发展。对顾客来说, 家政服务人员缺乏专业性, 专业化程度低下, 不仅会影响顾客的满意度和服务体验, 甚至会因此发生摩擦; 而对于家政服务企业来说, 顾客满意度低, 忠诚度低, 直接带来的后果就是影响家政服务企业的效益, 而目前行业中大多数家政服务企业规模较小, 所以也间接限制了企业的后续发展。所以, 提高家政服务人员的专业性, 以满足顾客的需求, 提升企业的品牌声誉, 推动行业健康发展, 已经成为当下亟待解决的问题。

如何提高家政服务人员的专业性, 最直接的途径便是加强对服务人员的职业培训, 规范化每个服务项目的服务标准。但是目前家政行业中家政服务人员在实际服务过程中通常会服务于多个服务项目, 导致的结果就是多而不精。根据人职匹配理论, 每个个体存在差异, 人们要根据自己的特点找到合适自己的职业, 以此达到人职匹配的目的。因此不同的个体应该有适合自己的服务项目培训。所以在如何提升家政服务人员的专业性这一问题就可以转化成家政服务人员与服务项目的人单匹配这一科学问题。在已有数据基础之上, 利用科学合理的方法对家政服务人员按照服务项目进行分类, 进而根据分类进行特定的服务项目培训, 以此提高家政服务人员的专业性, 提升服务质量。

### 1.2. 文献回顾

本节将对服务人员专业性及人岗匹配相关的文献进行梳理, 整理出目前两个领域的研究现状, 最后给出本研究中“人单匹配”的定义。

经过检索, 现有文献中关于服务人员的专业性的研究主要集中在服务人员的专业性与顾客满意度, 品牌忠诚之间的关系。在物流行业, 陆子平等人将服务人员专业性纳入顾客满意度评价体系, 并通过调查发现, 企业对于目前第三方物流企业服务人员专业性的满意度较高[1]。杨浩雄等人在构建顾客满意度测评体系时, 利用层次分析法将顾客满意度的影响因素分解成为各个更加易于观测的指标, 通过显著性检验, 得出工作人员专业度作为构成感知质量的四级指标, 与企业形象之间关系显著[2]。解芳以快递行业为背景, 将快递人员专业性等维度作为自变量, 最后通过实证研究, 证明了快递人员专业性正向影响顾客再次购买意愿以及顾客对快递公司的信任的假设[3]。

通过以上对于以往文献的爬梳可以看出, 关于服务人员的专业性的研究都集中在服务人员专业性与顾客满意度等的关系探讨上, 而鲜少有关于如何提高服务人员专业性, 从而提升顾客满意度的研究。因此本文针对以人单匹配推进家政服务人员的培训从而提高家政服务人员的专业性的问题进行了后续研究。

关于人岗匹配问题, Weller 等人认为匹配是个人与组织内的角色, 工作, 情境和任务动态对齐的过程[4]。杨续昌等人在实现人员任务分配过程中, 确定人员的属性指标集包括性别, 年龄, 学历, 专业能力, 工作经验等[5]。朱丽娜等人在人岗匹配的实例分析过程中, 以职工的工作经验、学习创新能力等五项作为评价指标, 并通过模型求解得到最终的职工与岗位匹配的结果[6]。基于人岗匹配的文献整理, 在匹配过程中主要涉及到的人员属性包括工作经验等, 并且都有约束条件。因此本文结合实际家政服务业的情况将“人单匹配”定义为: 家政服务员与服务项目达到一个人员与任务匹配的程度, 人指家政服务员, 单指服务项目, 也就是顾客需求。

### 1.3. KNN 分类算法

KNN 算法是大数据环境下经典的分类算法, 理论相对比较成熟而且比较简单易懂。

Nguyen 等人以 KNN 算法为基础, 对越南奎霍普热带丘陵区降雨诱发的浅层滑坡进行空间预测, 将坡度、坡长、坡向、土壤类型等 12 个影响因素作为特征属性, 预测是否会发生滑坡, 并将其与支持向量机等算法的结果进行对比, 发现  $k$  近邻模型优于支持向量机等模型[7]。王波, 程福云根据时间序列模型构建了一种基于 KNN 算法的股票预测模型, 模型相较普通的时间序列模型更为简单, 且更为精确[8]。Srividya 等人提出应用 KNN 识别高中生、大学生和职场人士等不同群体的心理健康状态, 以此监测具有异常行为的个人[9]。

通过对 KNN 算法的文献梳理可以看出, KNN 算法的应用对象不仅有物, 其应用对象也慢慢地扩展到了人, 人的分类逐渐成为热点。

## 2. KNN 算法

### 2.1. KNN 算法分类流程

KNN 算法是一种距离类大数据分类算法, 主要是以训练样本与测试样本的距离为条件进行分类。主要分类过程如下:

- (1) 收集数据集, 准备数据, 对数据进行预处理, 将数据进行清洗及标准化, 使样本数据处于同一个量纲上。
- (2) 将数据集分为训练集和测试集, 即分离数据集。
- (3) 设定参数  $k$ , 选用距离公式。
- (4) 计算测试样本到这训练样本的距离, 并取前  $k$  个距离最近的样本作为测试样本的  $k$  个近邻。 $k$  个近邻中出现频率最高的样本的类别标签就是该测试样本的类别标签, 以此为依据, 得出各个测试样本的类别标签。

## 2.2. KNN 算法的改进

本研究将利用样本距离权重对 KNN 算法进行改进。

基于模糊理论的 KNN 算法, 主要是在样本不均衡的情况下, 针对样本距离的权重问题进行了研究及改进。基于模糊理论进行改进主要是通过样本密度和样本距离来计算样本的隶属度函数, 大大减少了样本的不均匀对分类结果产生的影响。

基于模糊理论对 KNN 算法进行改进的过程中一个关键的概念就是模糊隶属度。在模糊 KNN 算法中, 一个样本属于很多类别, 只是隶属程度不同。一般来说, 处于分类边界的样本隶属度值低, 越靠近类别中心点隶属程度越高, 隶属度值也就越高。在这里利用下面的表达式来计算隶属度值[10]:

$$\mu_{ic} = \begin{cases} 0.51 + 0.49 \left( \frac{n_i}{k} \right), & \text{如果训练样本本身的类别标签为 } l_c \\ 0.49 \left( \frac{n_i}{k} \right), & \text{其他情况} \end{cases} \quad (1)$$

上述表达式中,  $\mu_{ic}$  表示训练样本  $x_i$  属于类别  $l_c$  的可能性, 也就是隶属程度。 $k$  表示训练样本的  $k$  个近邻,  $n_i$  表示在  $k$  个近邻中属于类  $l_c$  的个数。上述表达式的条件就是如果训练样本  $x_i$  本身的类别与  $l_c$  相同, 那么,  $\mu_{ic} = 0.51 + 0.49 \left( \frac{n_i}{k} \right)$ , 否则的话,  $\mu_{ic} = 0.49 \left( \frac{n_i}{k} \right)$ 。接下来基于这样的条件来计算测试样本  $y_j$  属于各个类别的可能性, 取其中的最大值所对应的类别作为测试样本的类别标签。这个可能性即隶属程度我们用  $\mu_c(y_j)$  来表示, 表达式如下:

$$\mu_c(y_j) = \frac{\sum_1^k \mu_{ic} \frac{1}{\text{dist}(y_j, x_i)^{\frac{2}{b-1}}}}{\sum_1^k \frac{1}{\text{dist}(y_j, x_i)^{\frac{2}{b-1}}}} \quad (2)$$

上式中,  $\frac{1}{\text{dist}(y_j, x_i)^{\frac{2}{b-1}}}$  是将测试样本  $y_j$  与训练集样本  $x_i$  之间的欧氏距离用  $\frac{2}{b-1}$  进行加权并取其倒数

来作为对距离进行加权的因子,  $b$  用来控制加权的程度, 一般来说  $b > 1$ 。最终测试样本  $y_j$  的类别标签用最大隶属度原则来计算, 即  $l(y_j) = l_{c'}, c' = \max(\mu_c(y_j))$ 。

## 3. 基于改进的 KNN 算法的家政服务人单匹配研究

### 3.1. 数据预处理

公司的服务者数据中缺失值数据占总体数据样本的比例较小, 所以采取直接将其删除的方法进行数据清洗。另外, 本文的研究目的是将家政服务员与服务项目进行专业度匹配, 所以在数据样本处理时将顾客对每个服务项目对应的家政服务员的评分阈值设定为小于等于 5, 以此筛选出对于顾客给予高分的家政服务员样本, 作为研究的数据集。

数据集中关于家政服务员的特征属性主要包括家政服务员的性别, 年龄, 户籍, 从业年限, 工作时长, 学历。首先将数据集分为训练集和测试集, 其中 90% 为训练集, 10% 为测试集, 训练集中每个服务者相对应的服务项目已知。其中服务项目分别为: 1 代表深度保洁, 2 代表企业保洁, 3 代表日常保洁, 4 代表代表母婴护理。其中服务者的户籍主要分为农村户籍和城镇户籍, 以 0 代表农村户籍, 1 代表城镇

户籍。工作年限主要分为进入家政服务行业工作 1~2 年, 3~4 年, 5~6 年, 7~8 年及八年以上, 分别用 1~5 表示。每日平均工作时长是通过工作总时长除工作总天数所得。服务者的性别中 0 代表性别为女, 1 代表性别为男。服务者的学历分为小学、初中、高中、大专、大学及以上, 分别用 1~5 表示。

通过数据清洗, 去除噪声数据及缺失数据之后, 共采集到 3460 条家政服务员数据。数据集基本信息如下表, 其中家庭保洁样本数为 1622, 企业保洁样本数为 168, 日常保洁样本数为 1504, 母婴护理样本数为 166。如表 1 是家政服务员数据集的基本信息。

**Table 1.** Basic information of data set

**表 1.** 数据集基本信息

数据集	样本数	属性数	类别数
家政服务员数据	3460	6	4

根据每个特征属性的数据性质进行标准化, 其中二元型数据在进行标准化是直接用 0, 1 表示, 即  $X_1 = \begin{cases} 0, & \text{户籍为农村} \\ 1, & \text{户籍为城镇} \end{cases}$ ,  $X_5 = \begin{cases} 0, & \text{性别为女} \\ 1, & \text{性别为男} \end{cases}$ 。而数值型数据则用  $x = \frac{x_j - x_{\min}}{x_{\max} - x_{\min}}$  将数据归一化到 [0, 1] 之间。

截取部分归一化后的数据如表 2。

**Table 2.** Standardization of household service personnel data

**表 2.** 家政服务人员数据标准化

户籍	年龄	工作年限	平均时长	性别	学历	服务项目
$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$Y$
0	0.375	0	0.437	1	0.25	1
0	0.275	0	0.566	1	0.25	1
1	0.325	0.75	0.493	0	0.75	2
0	0.45	0.5	0.393	0	0.25	2
1	0.325	0.25	0.728	1	0.5	3
0	0.425	0.25	0.653	0	0.25	3
0	0.325	1	0.658	0	0.25	4
0	0.7	1	0.654	0	0.25	4

### 3.2. 模型分类准确度的计算

在进行数据归一化后, 基于原始的 KNN 算法的人单匹配模型利用欧氏距离来计算训练样本与测试样本的距离。欧氏距离的公式为  $\text{dist}(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$ 。

将 KNN 算法分类结果与原始服务项目标签对比, KNN 算法分类结果的准确率达到 61.56%。

表 3 是基于原始 KNN 算法的人单匹配模型和基于模糊算法改进的人单匹配模型的分类型准确率对比。

**Table 3.** Comparison of accuracy

**表 3.** 准确率对比

KNN	FKNN
61.56%	69.36%

从以上分类准确率的对比可以看出基于模糊算法改进的人单匹配模型分类准确率高于基于原始的 KNN 算法的人单匹配模型, 能够更好地将家政服务员与家政服务项目进行分类匹配。

#### 4. 研究总结

研究着眼于目前家政服务行业服务人员专业性薄弱, 流动率高, 顾客满意度低等问题, 以人单匹配作为切入点, 利用 KNN 算法将家政服务人员以服务项目为标签进行分类, 从服务人员专业性研究方面探讨推动专业性的路径问题, 推动服务人员专业性研究。另一方面本文基于服务项目利用改进的 KNN 模型对家政服务人员进行分类, 在家政服务行业普遍存在顾客服务体验差等问题的现状下, 实现更精确的人和服务项目的人单匹配, 以促进专项职业培训, 提升服务质量, 促进家政服务行业的整体发展。家政服务公司可以开展相关职业技能培训, 细化每个服务项目的服务标准, 使不同的家政服务员在适合自己的家政服务项目上不断提升专业技能, 深耕细作, 提升服务质量, 从而提高顾客满意度, 推动家政服务业的长远发展。

#### 参考文献

- [1] 陆子平, 聂鸿飞. 我国第三方物流企业顾客满意度评价体系的实证分析[J]. 物流技术, 2014, 33(1): 166-168+177.
- [2] 杨浩雄, 王雯. 第三方物流企业顾客满意度测评体系研究[J]. 管理评论, 2015, 27(1): 181-193.
- [3] 解芳. 快递服务质量与顾客再次购买意愿关系实证研究——基于顾客信任的中介作用[J]. 财经理论与实践, 2016, 37(3): 123-127.
- [4] Weller, I., Hymer, C.B., Nyberg, A.J. and Ebert, J. (2019) How Matching Creates Value: Cogs and Wheels for Human Capital Resources Research. *Academy of Management Annals*, **13**, 188-214. <https://doi.org/10.5465/annals.2016.0117>
- [5] 杨续昌, 陈友玲, 兰桂花, 等. 基于聚类分析和双边匹配的产品开发任务分配方法[J]. 计算机集成制造系统, 2017, 23(4): 717-725.
- [6] 朱丽娜. 基于熵权理论的人岗双边匹配决策[J]. 吉林金融研究, 2018(4): 61-67.
- [7] Bui, D.T., Nguyen, Q.P., Hoang, N.D., et al. (2016) A Novel Fuzzy K-Nearest Neighbor Inference Model with Differential Evolution for Spatial Prediction of Rainfall-Induced Shallow Landslides in a Tropical Hilly Area Using GIS. *Landslides*, **14**, 1-17. <https://doi.org/10.1007/s10346-016-0708-4>
- [8] 王波, 程福云. KNN 算法在股票预测中的应用[J]. 科技创业月刊, 2015, 28(16): 25-26.
- [9] Srividya, M., Mohanavalli, S. and Bhalaji, N. (2018) Behavioral Modeling for Mental Health Using Machine Learning Algorithms. *Journal of Medical Systems*, **42**, 88. <https://doi.org/10.1007/s10916-018-0934-5>
- [10] Keller, J.M., Gray, M.R. and Givens, J.A. (2012) A Fuzzy K-Nearest Neighbor Algorithm. *IEEE Transactions on Systems Man & Cybernetics*, **SMC-15**, 580-585. <https://doi.org/10.1109/TSMC.1985.6313426>