

表 3 中每一行代表一个聚类簇中心坐标, H1 至 H9 分别代表供电单位、地市、区县、客户地址经度、客户地址维度、工单类型、业务类型一级、业务

类型二级、业务类型三级对应的坐标。簇中心 1 至簇中心 23 分别代表的用电问题如表 4 所示。

表 4 聚类簇中心

簇中心编号	用电问题	簇中心编号	用电问题
1	E 市客户申请服务侧用电需求配合	13	C 市客户申请核实抄表数据异常
2	E 市客户反映停电时间长	14	A 市客户申请核实电能表异常
3	I 市客户申请服务侧用电需求配合	15	C 市客户申请服务侧用电需求配合
4	H 市客户催办业务进程	16	E 市客户申请核实电能表异常
5	A 市客户申请欠费复电登记	17	G 市客户咨询停电信息发布渠道
6	E 市客户投诉抢修服务超时限	18	A 市客户投诉抢修服务超时限
7	L 市客户咨询停电信息发布渠道	19	I 市客户申请核实电器损坏
8	G 市客户申请用电信息变更定量定比调整	20	L 市客户催办业务进程
9	C 市客户催办业务进程	21	H 市客户反映电压质量长时间异常
10	L 市客户申请用电信息变更客户联系方式调整	22	I 市客户反映停电时间长
11	A 市客户咨询抄表时间	23	C 市客户咨询抄表收费催收电费问题
12	L 市客户投诉抢修服务超时限		

通过将各个簇中心业务类型、业务分级的数值与事先约定的编码规则对照, 可以得到具体的用电问题, 从而降低服务风险. 以簇中心 6 为例, 该位置有较多的客户进行投诉, 反应抢修服务超出时限. 该模型可以实时导入新的数据, 实现对簇中心的实时调整以应对新的服务风险的出现.

3.5 结果分析

K-Means 算法作为一种常用的聚类算法, 对球状分布的数据具有很好的效果, 但是算法对初始聚类中心敏感, 容易受到孤立点的影响. 文中在聚类之前排除了孤立点的影响, 提出了一种新的选取初始聚类中心的方法. 针对客服工单数据中存在较多孤立点, 对聚类分析结果产生巨大影响的情况下, 文章对传统 K-Means 算法进行改进, 使改进后的 K-Means 算法更加适用于客户工单数据.

分别利用原 K-Means 算法和改进后的 K-Means 算法进行聚类分析对比, 聚类结果如表 5 所示.

实验结果表明, 改进算法更接近实际数据分布. 虽然需要查找少量孤立点, 会增加时间消耗, 但是改进算法准确度较高, 聚类效果较好.

为了更加直观的表现改进算法的优越性, 根据经典 K-Means 算法和改进后的 K-Means 算法的聚类结果分析对比画出曲线对比图, 如图 7 所示.

通过两者的聚类结果分析对比曲线图可以很明显

的看出改进后的 K-Means 算法 Calinski-Harabasz 得分更高, 聚类效果更好, 更加准确挖掘出潜在具有相同问题的电力客户.

表 5 原算法和改进后算法的 Calinski-Harabasz 分值对比

K 值	原算法分数值s	改进算法分数值s
20	379 409.683 233 6517	399 273.138 103 4749
21	393 956.542 504 964	402 513.044 880 1194
22	369 111.210 733 772 14	407 287.921 328 6048
23	407 448.930 150 306 14	421 681.062 466 495 44
24	397 800.846 911 3769	419 919.294 205 112 27
25	394 101.155 369 231 36	419 269.938 281 428 36
26	395 118.782 970 824 16	410 063.233 460 827 73
27	399 619.081 821 8347	408 707.420 899 271 56
28	398 331.317 749 9216	407 725.120 096 935 54
29	394 660.427 922 181 26	398 859.358 177 6948

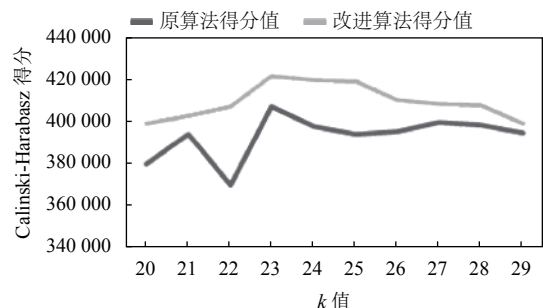


图 7 聚类结果分析对比曲线图

比较每一条聚类结果是否和真是的结果一致, 计算聚类结果的准确率 (Accuracy), 如式 (7) 所示.

$$acc = \frac{N_{cor}}{N} \quad (7)$$

其中, N 表示工单总数, N_{cor} 表示正确聚类的工单数. 改进后的 K-Means 聚类模型准确率高达 91.2%. 而采用传统的聚类算法模型, 准确率只有 85.7%. 通过验证认为, 改进后的 K-Means 模型能从工单数据出发, 较为精准地将具有相同问题的电力客户聚类.

4 结论

针对客户服务数据的特征, 本文给出了一种改进的 K-Means 聚类算法. 应用该算法可以从大量工单数据中找到若干个簇中心, 以挖掘出客户服务中的用电问题, 不仅为改进客服服务质量提供数据支撑, 还能为潜在服务风险的预测奠定数据基础, 从而让相关企业为客户提供更优质的服务.

以电力工单数据分析为例, 根据用电问题产生情况的不同, 可将类比分析分为被动服务(事件驱动)的类比分析和主动服务(服务驱动)的类比分析两种情况.

(1) 对于因大批量初装用户、举行促销活动、中介恶意查询、系统故障(缴费未到账)、出现极端天气等因素导致某类用电问题的用户达到一定数量或占一定比例的业务场景, 则可施行被动服务类比分析, 将发生该类问题的客户服务工单数据, 作为改进 K-Means 算法的输入, 进行聚类分析, 从而得到代表着该类用电问题的簇中心. 当再次接入新的客户时, 可以通过计算新客户与该类问题簇中心的欧式距离来判定潜在的风险: 若新客户在簇类内, 则客户是该类问题的潜在风险客户, 若客户在簇类外, 则客户发生该类问题风险较小. 因此, 改进的算法可以预先判断客户是否具有发生该类问题的风险, 从而提前实施相应的措施.

(2) 对于定期(每月、每周或每天)发起的如电费账单出账、线路系统升级改造、安全隐患定期排查等业务场景, 则可以采用主动服务类比分析, 将存在多种用电问题的客户服务工单数据, 作为改进的聚类算法的输入, 从而得到代表前 N 个最频繁出现的用电问题的簇中心. 以此数据为支撑, 再结合业务处置的历史经验, 可做出相应的日常风险预判. 比如, 该方法还可以通过往年同期数据的聚类, 挖掘出高概率发生的具体用电问题的信息包括时间和地点等, 通知相关部门做好预防措施; 再如, 通过对实时工单数据的聚类, 可以挖

掘出突发问题, 从而能及时通知相关部门前往验证并解决突发问题, 与此同时通知出现电力问题区域的客户, 让客户知晓当前的情况, 以减少投诉, 减轻客服压力.

值得一提的是, 该方法还可以应用到其他相关行业的客服系统, 以提升客户服务质量.

参考文献

- 1 谭火超. 关于完善电力客户服务、提升供电服务质量的探讨. 机电信息, 2018, (33): 174-175. [doi: 10.3969/j.issn.1671-0797.2018.33.100]
- 2 Kruppa J, Ziegler A, König IR. Risk estimation and risk prediction using machine-learning methods. Human Genetics, 2012, 131(10): 1639-1654. [doi: 10.1007/s00439-012-1194-y]
- 3 邹云峰, 何维民, 赵洪莹, 等. 文本挖掘技术在电力工单数据分析中的应用. 现代电子技术, 2016, 39(17): 149-152.
- 4 陈璐. 浅谈电力公司的电力营销信息系统的建设. 电子世界, 2014, (22): 208. [doi: 10.3969/j.issn.1003-0522.2014.22.200]
- 5 Hartigan JA, Wong MA. Algorithm AS 136: A K-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics), 1979, 28(1): 100-108.
- 6 Michie MG. Use of the Bray-Curtis similarity measure in cluster analysis of foraminiferal data. Journal of the International Association for Mathematical Geology, 1982, 14(6): 661-667. [doi: 10.1007/BF01033886]
- 7 Rahman G, Islam Z. A decision tree-based missing value imputation technique for data pre-processing. Proceedings of the Ninth Australasian Data Mining Conference. Ballarat, Australia. 2010. 41-50.
- 8 Bechtel W, Shagrir O. The non-redundant contributions of Marr's three levels of analysis for explaining information-processing mechanisms. Topics in Cognitive Science, 2015, 7(2): 312-322. [doi: 10.1111/tops.12141]
- 9 Rushton G, Armstrong MP, Gittler J, et al. Geocoding in cancer research: A review. American Journal of Preventive Medicine, 2006, 30(2S): S16-S24.
- 10 Guo XJ, Song ZX. To study on coding for identity card of citizen and its application. Value Engineering, 2007, 26(10): 114-116.
- 11 Zhang ZQ, Yang QY, Dou A. An improved K-means algorithm for reciprocating compressor fault diagnosis. Proceedings of 2018 Chinese Control and Decision Conference. Shenyang, China. 2018. 276-281.

- 12 Jain AK, Murty MN, Flynn PJ. Data clustering: A review. *ACM Computing Surveys*, 1999, 31(3): 264–323. [doi: [10.1145/331499.331504](https://doi.org/10.1145/331499.331504)]
- 13 Appelboam A, Reuben AD, Bengler JR, *et al.* Elbow extension test to rule out elbow fracture: Multicentre, prospective validation and observational study of diagnostic accuracy in adults and children. *BMJ*, 2008, 337: a2428. [doi: [10.1136/bmj.a2428](https://doi.org/10.1136/bmj.a2428)]
- 14 Crozier SN, Falconer DD, Mahmoud SA. Least sum of squared errors (LSSE) channel estimation. *IEE Proceedings F- Radar and Signal Processing*, 1991, 138(4): 371–378. [doi: [10.1049/ip-f-2.1991.0048](https://doi.org/10.1049/ip-f-2.1991.0048)]
- 15 Mahmud S, Rahman M, Akhtar N. Improvement of K-means clustering algorithm with better initial centroids based on weighted average. *Proceedings of International Conference on Electrical and Computer Engineering*. Dhaka, Bangladesh. 2012. 647–650.
- 16 Tsai CY, Chiu CC. Developing a feature weight self-adjustment mechanism for a K-means clustering algorithm. *Computational Statistics & Data Analysis*, 2008, 52(10): 4658–4672.
- 17 Aranganayagi S, Thangavel K. Clustering categorical data using silhouette coefficient as a relocating measure. *Proceedings of the International Conference on Computational Intelligence and Multimedia Applications*. Sivakasi, Tamil Nadu, India. 2007. 13–17.
- 18 Łukasik S, Kowalski PA, Charytanowicz M, *et al.* Clustering using flower pollination algorithm and Calinski-Harabasz index. *Proceedings of 2016 IEEE Congress on Evolutionary Computation*. Vancouver, BC, Canada. 2016. 2724–2728.