

doi: 10.7690/bgzd.2025.04.016

基于电网对账系统的 TF-IDF 优化算法

王 岩, 郭 威, 隋海滨, 符贵谦

(海南电网有限责任公司信息通信分公司数据资产运营管理部, 海口 570203)

摘要: 为提高电网集团对账系统的工作效率, 优化电网集团的服务效果, 设计一种 TF-IDF 词频-逆向文件频率 (term frequency-inverse document frequency, TF-IDF) 优化算法。论述电网集团对账系统的基本设计, 讨论服务器和浏览器 (browser/server, B/S) 架构下的 TF-IDF 算法优化设计方法, 对 B/S 架构下使用 TF-IDF 算法优化设计在电网系统中的综合应用效果进行分析。结果表明: 该算法的对账效果提升明显, 为优化电网对账系统提供了技术基础, 为提升电网集团服务质量做出了贡献。

关键词: 电网集团; B/S 架构; TF-IDF 算法; 电网对账系统; 对账效果

中图分类号: TM727 **文献标志码:** A

TF-IDF Optimization Algorithm Based on Power Grid Reconciliation System

Wang Yan, Guo Wei, Sui Haibin, Fu Guiqian

(Data Asset Operation Management Department, Information and Communication Branch of Hainan Power Grid, Haikou 570203, China)

Abstract: In order to improve the work efficiency of the power grid group reconciliation system and optimize the service effect of the power grid group, a term frequency-inverse document frequency (TF-IDF) optimization algorithm is designed. This paper discusses the basic design of the power grid group reconciliation system, discusses the TF-IDF algorithm optimization design method under the B/S (browser/server) architecture, and analyzes the comprehensive application effect of the TF-IDF algorithm optimization design in the power grid system under the B/S architecture. The results show that the reconciliation effect of the algorithm is significantly improved, which provides a technical basis for optimizing the power grid reconciliation system and contributes to improving the service quality of the power grid group.

Keywords: power grid group; B/S architecture; TF-IDF algorithm; power grid reconciliation system; reconciliation effect

0 引言

随着社会的发展, 电力已经成为了当前社会的主流能源, 社会对其依赖性也越来越高^[1]。随着电力应用量的增加, 近些年较为严重的电力对账问题已经成为电力集团迫切想要解决的首要问题^[2]。电力对账系统面临的问题主要包括数据量太大, 需要的人工较多; 人工技术有限, 数据处理时间太长; 以及人工处理数据处理误差较大等问题^[3]。需要采用新方法为电网对账系统提供技术支撑, 优化此类问题。

当前, 在科技飞速发展的情况下, 电网对账系统研究也在不断朝着智能技术方向进发, 自动对账系统已经成为了电网对账系统的主要目标。笔者设计 B/S 架构下的 TF-IDF 算法, 对电网对账系统进行综合优化。

1 研究理论与方法

1.1 研究问题

电网企业一般都是大型企业, 承担着保障国民正常用电的重要任务, 同时也为国家经济的发展提供了重要支撑^[4]。随着电网集团业务的不断增加, 电网集团的财务对账系统工作任务也越来越繁重; 因此, 通过科学技术手段提高电网集团对账系统的工作效率是当前电网集团的主要任务^[5]。

随着“信息化、智能化、精益化”管理要求和财务及时、规范化传递, 一体化运作、标准化管理模式的不断践行, 营销财务对账中数据统计、数据推送财务环节、银行接口采集数据等流程环节各类数据维护类问题层出不穷^[6]。财务管理系统作为和其他系统协同最多的系统, 协同类问题往往涉及到多个系统, 在一个应用场景中, 相互间调用频繁,

收稿日期: 2024-08-11; 修回日期: 2024-09-20

第一作者: 王 岩 (1995—), 女, 陕西人。

导致问题难于定位。同时,在银电联网服务上,缺乏全面完善的监控、报警机制,服务错误信息都是记录在日志中,无法及时提供预警信息而带来的问题也日益突出^[7];因此,亟需电网集团实现智能营财对账及银电联网服务监控优化,以业务改进为动力,完善电费对账以及相关系统应用,进一步提高业务信息化水平,从效率及管理角度提高电网企业的核心服务能力,提高客户满意度。

1.2 电网集团对账系统

由 1.1 节可知,报表核对过度依赖人工干预,对人员素质要求高,且人工的工作量较大^[8]。常见的问题如下:

1) 日常对账问题。

日常对账中数据统计、数据推送财务环节、银行接口采集数据、数据核对等流程环节存在数据汇总不透明,第 3 方收费接口不稳定,收费抹账数据写入重复,汇总规则错乱,自动对账模式单一,跨区交费无法对账问题,常导致财务到账信息和营销实收不在同一单位,跨区交费受限,数据重复无法重推,数据缺失对账池数据核查困难,数据量大自动对账卡死,对账池数据核查困难,数据采集不稳定漏导情况严重,很大程度影响了电费财务数据的一一对应关系,造成对账模式混乱的局面^[9]。

2) 月对账问题。

在营销财务月对账中,由于数据汇总不透明、无预收数据,导致统计汇总的数据无法穿透,数据差异难以定位排查,财务无法核对预收情况。同时,数据推送至财务管理系统后,若发现错误,采用作废单据方式,营销管理系统无法重新推送数据,常需后台人员手工重推,耗时耗力且难以保证数据准确性^[10]。

3) 应收、实收报表问题。

海南电网应收、实收报表由各供电局分别核查、处理,处理完毕后数据上报省公司。应收、实收报表核对工作主要依靠人工,报表数据异常问题未在系统进行管理^[11]。

当前,很多国内外研究已经为提升电网集团做出了重要贡献。刘思源等^[12]指出目前电网集团针对营销信息化系统正式提出了一个“十五年”规划,提出要以全面推行集团化运作以及精细化管理作为主要内容,重点实现各地市供电公司的自动化管理,完成对各个管理环节的统一控制,从而提升应对复杂市场环境的能力,降低运行成本,实现供电公司

的可持续发展。Logeswaran 等^[13]指出电网系统作为电力集团的重要支柱,对其对账系统进行综合研究,提高其性能,对电网集团的发展具有重要作用。综上所述,设计在浏览器和服务器 B/S 架构下使用词频-逆向文件频率 TF-IDF 算法构建电网集团自动对账系统,提高电网集团的对账效率,对推动电网集团的发展具有重要作用。

2 B/S 架构下的 TF-IDF 算法

2.1 B/S 架构下的 TF-IDF 算法优化设计

B/S 架构是对 (clientserver, CS) 架构进行改进后的产物。在使用过程中,只需在客户端安装浏览器,并在综合服务器中安装数据库,然后便可使用网页服务器与数据库进行数据交换^[14]。

关键字在文档中出现的频率 (term frequency, TF), 在总文档中包含关键字的局部文档的分布状况 (inverse document frequency, IDF), 关键字抽取 TF-IDF 算法,其基本理念就是根据计算权重对需要检测的文档中的关键词进行筛选,进而根据关键词筛选进行自动文档检测^[15]。TF-IDF 的计算公式为:

$$TF-IDF = TF * IDF = tf * \log(N/n_i + 0.01)。 \quad (1)$$

式中: tf 为关键词的词频; N 为文档的总数量; n_i 为出现关键词的文档数量。对上述公式进行改进,可得到更加全面的计算公式:

$$TF-IDF = tf * \log n_i * \log(N/n_i + 0.01) * \left[1 - \sqrt{\sum_{j=1}^n (tf - \hat{tf})^2 / (k-1)} / \hat{tf} \right]。 \quad (2)$$

式中: $\hat{tf}$ 为关键词在每篇文档中出现的平均次数; k 为其他类别文档的总数量。在 B/S 架构下的数据库中使用 TF-IDF 可以根据关键词对数据库中的文档进行自动筛选,然后根据筛选结果实现自动对账。笔者设计的 B/S 架构下 TF-IDF 算法的电网对账系统原理如图 1 所示。

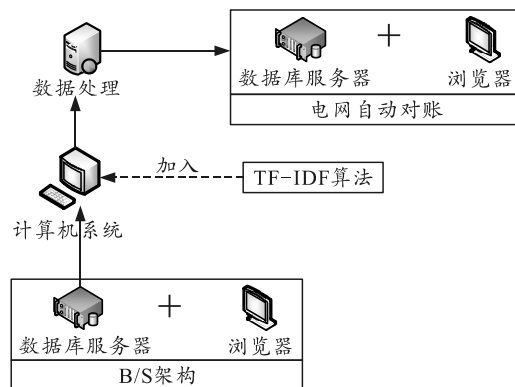


图 1 B/S 架构下 TF-IDF 算法的电网对账系统原理

图 1 中，笔者通过 TF-IDF 算法实现对账系统的优化，同时还可对电网对账系统原有的前置机服务、后台接入服务、后台调用服务等关键程序进行改造，在相应位置埋入监控点，及时将错误信息及正常业务统计信息以 socket 通信或者消息队列的方式发送到监控服务，由监控服务进行集中存储、处理、告警等。实现对供电发起的单笔扣款、银行发起的单笔扣款、供电发起的批扣、银行发起的批扣等银联网交易的实时监控预警，增加功能能够查看监控信息，弹出窗口进行提示，并通知相关人员。该优化设计的主要工作流程如图 2 所示。

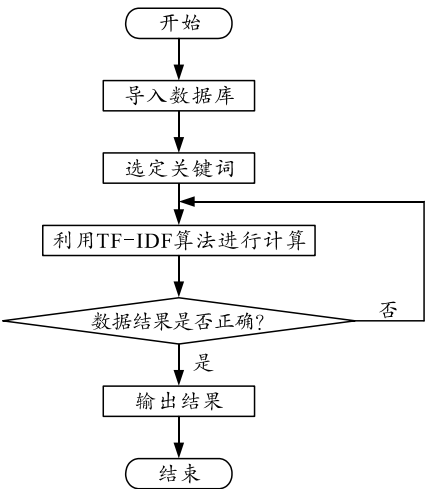


图 2 B/S 架构下 TF-IDF 优化算法的主要工作流程

图 2 中，在 B/S 架构下设计优化后的 TF-IDF 算法用与电网对账系统，通过关键词筛选不仅能够提高电网对账系统的计算效率，而且能实现电网对账系统的自动对账目标。

2.2 研究数据来源

本文中数据来源为电网集团市场营销部门，将电网对账系统的数据作为研究的样本，然后通过数据训练评估设计的 B/S 架构下的 TF-IDF 算法在电网对账系统的工作状况，并且以一定量的数据作为样本，对传统的电网对账系统和传统的 TF-IDF 算法，与设计优化后的 TF-IDF 算法对账效果进行对比，从而综合分析设计的 B/S 架构下的 TF-IDF 算法的具体性能。主要通过数据测试召回率 (recall rate) 和准确率 (accuracy) 作为主要的评价依据，并且通过综合计算以 F 值作为综合性评价依据， F 值的计算公式为：

$$F_1 = (2 * R * P) / (R + P)。$$
 (3)

并且采用的主要训练关键词来源为电网集团的日常工作原则：

1) 按供电单位、收费员、收费部门、交易方式、用户编号、结算户号、实收时间起止等条件查询需核查的收费数据。按供电单位、收费部门、交易方式查询时，按照勾选的分组方式(供电单位、收费部门、收费员、收费日期)进行汇总显示并支持点击明细可以弹窗查询到具体的实收记录；按用户编号、结算户号查询时，显示实收记录；

2) 按供电单位查询时支持勾选多条汇总记录调用核查服务，按收费员、用户编号、结算户号查询时支持勾选多条用户收费记录调用核查服务，对有异常的收费数据生成异常记录清单并自动下发督办工单，前台收费需按收费员/收费日期生成督办工单，非前台收费按收款银行/收费日期/收费方式生成督办工单。

依据以上原则选择个别关键词作为模型训练的关键词，并通过一定量的数据作为样本对模型进行训练，并记录结果。

基于此，笔者对设计的 B/S 架构下的 TF-IDF 算法对账性能进行评估，评估过程中在 Matlab 虚拟平台中进行实验，笔者使用的实验参数信息如表 1 所示。

表 1 虚拟实验平台信息

序号	实验工具	版本信息
1	Matlab	9.2.0.538062 (R2017a)
2	操作系统	Microsoft Windows 10 专业版 Version 10.0(Build 15063)
3	Java	Java 1.7.0_60-b19 with Oracle Corporation Java HotSpot(TM) 64-Bit Server VM mixed mode

表 1 中，笔者基于以上设置进行了虚拟实验，在过程中共使用 6 千万条数据量对模型进行了性能评估，并使用 6 千万条数据量对模型的实际对账效果进行了对比评估，从而突出了本文中研究的具体价值。其中数据项包括客户信息以及电量管理信息。

3 结果与分析

3.1 B/S 架构下的 TF-IDF 算法对账性能评估

传统的电网对账系统采用的是人工对账方法，在电力普及的状况下该方法需要大量的人工进行对账，因此对账效果非常差。通过优化客户端和服务 器实现 TF-IDF 算法进行自动化对账具有重要意义。传统的 TF-IDF 算法也存在一定的缺点，经过优化后的 TF-IDF 算法不仅能够实现突破性的自动化对账理念，还能很好地解决传统 TF-IDF 算法所面临的问题。B/S 架构下 TF-IDF 算法的对账性能评估结果如图 3 所示。

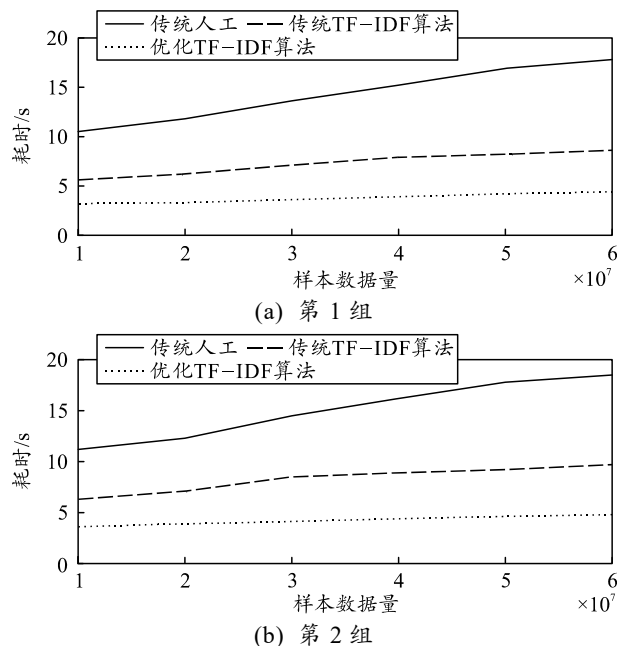


图 3 B/S 架构下 TF-IDF 算法的对账性能评估

图 3 中, 2 组采用不同的关键词原则。通过对账操作对比, 不同方法之间的数据处理平均耗时差异很大。其中, 传统人工对账方法平均耗时最低在 10 s 左右, 随着数据量的增大, 传统人工对账方法的数据处理平均耗时也在不断增加, 最高在 18 s 左右, 传统 TF-IDF 算法的数据处理平均耗时最低在 5 s 左右; 并且随着数据量的增大, 传统 TF-IDF 算法的数据处理平均耗时呈现缓慢增加状态, 最高在 10 s 左右。而设计的 B/S 架构下的 TF-IDF 算法数据处理平均耗时最低在 3 s 左右, 并且随着数据量的不断增加, B/S 架构下的 TF-IDF 算法数据处理平均耗时浮动并不大, 最高在 5 s 左右。可见设计的 B/S 架构下的 TF-IDF 算法在电网对账系统中的工作性能较传统的对账方法有了较大提升。

3.2 B/S 架构下的 TF-IDF 算法对账效果评估

对账性能提升是算法优化的技术基础, 而对账效果是算法应用的最终成果。设计的 B/S 架构下 TF-IDF 算法的对账效果评估结果如图 4—5 所示。

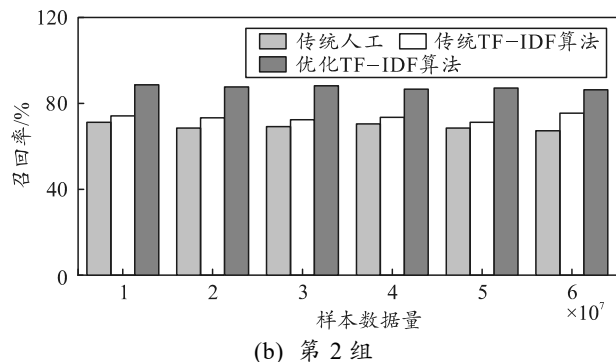
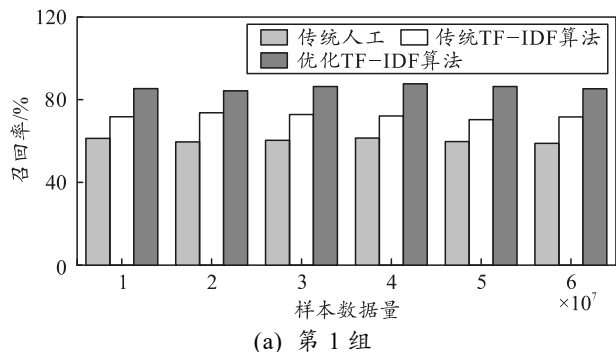


图 4 B/S 架构下 TF-IDF 算法的对账召回率评估

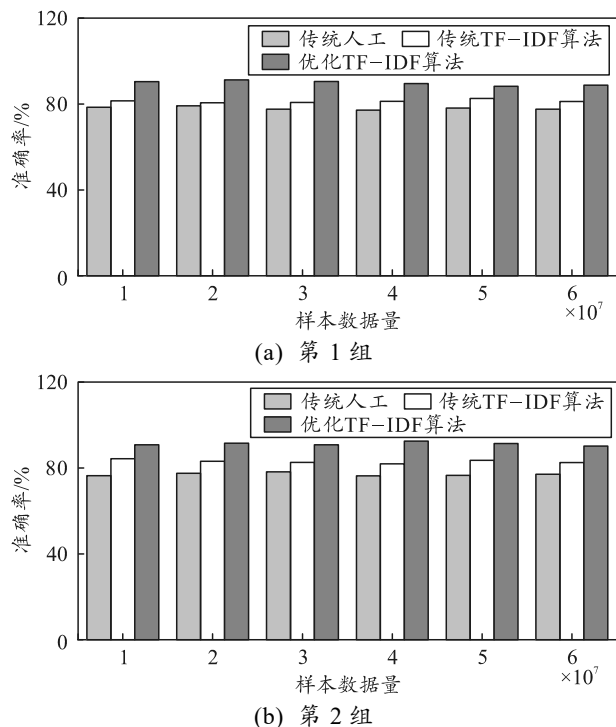


图 5 B/S 架构下 TF-IDF 算法的对账准确率评估

图 4—5 中, B/S 架构下 TF-IDF 算法的对账效果评估中, 设计的优化 TF-IDF 算法平均召回率在 89% 左右, 准确率在 90% 左右, 较其他 2 种方法有较大的优势。TF-IDF 算法对账效果 F 值的评估结果如图 6 所示。

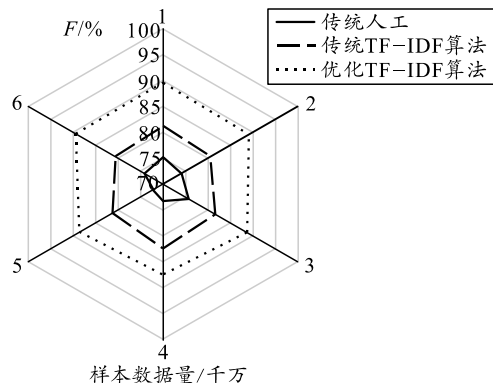
图 6 B/S 架构下 TF-IDF 算法对账效果 F 值评估结果

图6中,在3种对账方法综合效果评估中,传统人工的 F 在75%左右,传统TF-IDF算法 F 值在80%,设计优化的TF-IDF算法 F 值在90%左右。可见设计的B/S架构下TF-IDF算法的对账效果具有突破性的提升。

4 结论

研究表明:在电网对账性能评估中,传统人工的对账方法平均耗时最低在10s左右,传统人工对账方法的数据处理平均耗时也在不断增加,最高在18s左右;传统TF-IDF算法的数据处理平均耗时最低在5s左右,随着数据量的增大,传统TF-IDF算法的数据处理平均耗时呈现缓慢增加状态,最高在10s左右。笔者设计的B/S架构下的TF-IDF算法数据处理平均耗时最低在3s左右,并且随着数据量的不断增加,B/S架构下的TF-IDF算法数据处理平均耗时浮动并不大,最高在5s左右。在综合效果评估中,设计的优化TF-IDF算法平均召回率在89%左右,准确率在90%左右,较其他2种方法有较大的优势。并且传统人工的 F 在75%左右,传统TF-IDF算法 F 值在80%,设计优化的TF-IDF算法 F 值在90%左右。

参考文献:

- [1] 张颖. 探究提高营销系统电费自动对账率的对策[J]. 技术与市场, 2019, 26(11): 189-190.
- [2] AMULYA G, KUMARS V D A. Dealing With Reactive Power in Islanded Micro Grid Corresponding Power Distribution in Hierarchical Droop Control using Photovoltaic System[J]. Journal of Trend in Scientific Research and Development, 2019, 3(5): 12-14.
- [3] 杨春燕, 宾冬梅, 黎新. 基于PBFT区块链技术的电网企业仓储系统[J]. 电信科学, 2021, 37(2): 144-153.
- [4] 张君怡, 任鑫博. 非独立组网架构下5G网络在智能电网中的应用[J]. 农村电气化, 2020(6): 5-7.
- [5] 李芬, 李瑾, 仵松顺, 等. 基于智能合约的多微电网市场化交易模型及算法[J]. 网络与信息安全学报, 2020, 6(4): 56-66.
- [6] SHARMA D K, RAPA K G K, PASUPULLA A P, et al. A review on smart grid telecommunication system[J]. Materials Today: Proceedings, 2022, 51: 470-474.
- [7] BETANCOURTPAULINO P, CHAMORRO H R, SOLEIMANI M, et al. On the perspective of grid architecture model with high TSO - DSO interaction[J]. IET Energy Systems Integration, 2021, 3(1): 1-12.
- [8] 裴璐遥. 电网企业网管系统的设计和场景验证[J]. 南方能源建设, 2020, 7(4): 81-86.
- [9] SECK G S, KRAKOWSKI V, ASSOUMOU E, et al. Embedding power system's reliability within a long-term Energy System Optimization Model: Linking high renewable energy integration and future grid stability for France by 2050[J]. Applied Energy, 2020, 257(2): 114037.
- [10] YANG D, LI W, YAGLI G M, et al. Operational solar forecasting for grid integration: Standards, challenges, and outlook[J]. Solar Energy, 2021, 224(4): 930-937.
- [11] NESPOLI L, MEDICI V, LOPATICHKI K, et al. Hierarchical demand forecasting benchmark for the distribution grid[J]. Electric Power Systems Research, 2020, 189(7): 106755.
- [12] 刘思远, 李傲雪, 王同旭. IWOS 系统在配网故障处理中的应用[J]. 农村电气化, 2020(2): 52-56.
- [13] LOGESWARAN T, RAJA M S, HAMEED J B S, et al. Power flow management of hybrid system in smart grid requirements using ITSA-MOAT approach[J]. Applied Energy, 2022, 319(9): 119228.
- [14] GREEN H, DURRANT J D. Deepfrag: An open-source browser app for deep-learning lead optimization[J]. Journal of chemical information and modeling, 2021, 61(6): 2523-2529.
- [15] 路健, 范增民, 刘彩娜. 基于TF-IDF算法的供应链信息定向挖掘模型[J]. 计算机仿真, 2021, 38(7): 153-156.