

基于 KMP 算法的电网远动数据校核方法研究

为解决调度自动化系统主站系统与站端远动系统的远动数据无自动校核手段的问题，基于远动数据的特点，提出了基于 KMP 算法的数据校核方法，并引入了匹配度控制参数 MLP (Matching Length Parameter) 来实现校核精度的控制，以满足校核精度和效率的要求。实际应用情况表明，该方法具有很强的科学性和实用性。

郝蛟 江伟 邓彬/深圳供电局有限公司



郝蛟/助理工程师

关键词/Keywords

远动 ·
校核 ·
KMP ·
控制参数 ·

0 引言

调度自动化系统的远动功能是指传送遥测、遥信、遥控以及遥调等变电站运行及控制信息的功能，由变电站的远动系统终端与主站系统共同实现，是调度自动化系统实现监控电网、变电站无人值班最基本的功能^[1]。遥测和遥信由变电站远动系统采集并远程上送至主站系统，遥控和遥调则由主站系统远程下发遥控和遥调命令至变电站远动系统。

远动功能的安全性和有效性有赖于变电站远动系统和主站系统对远动信息配置的正确性和一致性。如出现变电站与主站配置不一致的情况，将造成电网的遥测、遥信等采集数据不正确，而如果遥控数据有误的话，还会使电网的遥控操作存在很大的安全风险。因此确保变电站远动系统和主站系统的远动信息配置的一致，对调度自动化系统的安全有效运行，防止调度人员误操作，具有重要意义。

远动信息表是主站系统与站端系统远动信息的配置文件，定义了远动信息名称以及上传顺序编号等信息。

对于调度自动化主站系统 (EMS)，远动信息存储在相应的数据库表中，并一般由数据维护人员人工输入和维护，存在因人员经验、工作状态等因素造成数据维护错误的风险。目前的通用做法是数据维护完成后由数据维护人员人工检查，同时定期开展专项工作对远动数据进行核查，效率低下，同时无法规避人工核查出错的风险。目前尚无针对远动信息校核的高效的数据校核方法，也无针对调度自动化系统远动信息自动校核的自动化软件。

本文针对调度自动化系统远动信息的特点和管理方法，引入了高效的字符串匹配算法，提出了基于 KMP 算法的电网远动数据校核方法，并在此基础上引入了匹配度控制参数 MLP，通过对参数的合理设置，能够有效地分离出远动信息的关键性错误和名称不规范问题，便于数据维护人员迅速定位关键错误，降低电网监控风险。该方法在某地电网 EMS 系统的应用情况，说明了其有效性和科学性。

1 KMP 算法简介

KMP 算法是一种高效的字符串匹配算法，由 D.E.Knuth、V.R.Pratt 和 J.H.Morris 同时发现，因此称它为 Knuth-Morris-Pratt 操作 (简称 KMP 算法)。其对于任何模式和目标序列，都可以在线性时间内完成匹配查找^[2]。KMP 算法完成的任务是：给定两个字符串 O 和 f ，长度分别为 n 和 m ，判断 f 是否在 O 中出现，如果出现则返回出现的位置。

朴素匹配算法是遍历 O 的每一个位置，然后从该位置开始和 f 进行匹配，但是这种方法的复杂度是 $O(nm)$ 。而 KMP 算法的核心在于通过引入一个跳转表 $next[]$ ，在匹配失配时，根据之前匹配记录的信息，可以跳转多位到跳转表定义的位置继续匹配，避免了不需要的匹配任务，使匹配的复杂度降为 $O(n+m)$ ，从而大大提高了匹配的效率^[3]。

2 基于 KMP 算法的远动数据校核

2.1 远动信息的基准数据

每个变电站都维护有一张远动信息表，表中记录了该站的所有遥测、遥信、遥控信息等远动信息的名称、顺序序号、信息体地址以及端子号等配置信息^[4]。其根据变电站一次设备监控需求以及实际测控装置的情况设计，并由变电站远动系统运维人员根据现场情况和需求的变更而不断维护和修改。

变电站远动系统运维人员根据远动信息表完成站端远动信息的配置，包括二次回路接线以及远动机远动信息的下装，使站端实际配置与远动信息表一致。主站系统数据维护人员根据站端提供的远动信息表完成主站系统数据库相关数据包括远动名称与顺序编号的录入，使主站系统远动信息与远动信息表一致。

由此可见，远动信息表是站端和主站系统远动信息配置的基准，也是站端和主站信息保持一致的桥梁。对于调度自动化主站系统，保持其远动信息的正确性即是保证主站系统数据库中的四遥信息数据与远动信息表一致。

2.2 远动数据的特点

远动信息表包含了遥测、单位位置遥信、双位置遥信、遥控的信息。主站系统则分别对应了前置遥测定义表、前置遥信定义表和遥控关系表。主站系统将表中的远动名称与顺序编号维护录入到相应的数据库表中^[5]。

遥测：设备的各种电气量，如有功、无功、电流等。远动信息表中按照调度编号+电气量命名，如 1246 P，主站系统数据库里按照中文名+调度编号+电气量，如蕤科 II 线 1246 有功值。

单位位置遥信：保护信号信息。远动信息表中按照电压等级+中文名+调度编号+信号名称命

名，如 110 kV 蕤科 II 线 1246 控制回路断线，主站系统数据库里类似，如 110 kV 蕤科 II 线 1246 控制回路断线。

双位置遥信：开关设备的分合位置信息。远动信息表中按照调度编号+下划线+合位（或分位）命名，如 1246_合位，主站系统数据库里按照中文名+调度编号+设备类型+遥信值，如蕤科 II 线 1246 开关遥信值。

遥控：对开关设备的远程控制信息。远动信息表中按照调度编号+合（分）命名，如 1246 合，主站系统数据库里按照中文名+调度编号+设备类型遥信值命名，如蕤科 II 线 1246 开关遥信值。

可以看到远动信息表和主站系统数据库对远动名称的记录方式有所不同，远动信息表由于维护的便利性要求，其名称描述比较简练，是主站系统数据库该远动信息名称的子集。因此名称匹配的问题就等效于上节描述的在主站系统数据库远动信息名称字符串 O 中匹配远动信息表远动信息名称 f 的问题，可以使用 KMP 算法解决。

2.3 基于 KMP 算法的校核方法

远动信息名称往往与实际一次设备直接相关，而调度编号则是设备的唯一标识，因此一次设备的调度编号是远动名称的关键信息。通过上小节的分析可知，双位置遥信和遥控名称只需调度编号匹配上，则该信号信息就是匹配的。遥测和单位位置遥信则需比对除调度编号外的额外信息，遥测还需比对电气量类型，单位位置遥信则更为复杂，还需要比对信号名称。

由于远动信息数据量庞大，对效率要求较高，朴素匹配算法已无法满足效率要求，而 KMP 算法能够在朴素匹配算法的基础上大大简化匹配复杂度，提高校核效率。

(1) KMP 算法匹配

为了提高 KMP 算法的效率，在匹配之前需要对从远动信息表读取的名称进行预处理：将双位置遥信和遥控名称中的“合（或分）”、下划线等过滤，只剩调度编号；将遥测名称中的英文字符电气量转换成中文字符；单位位置遥信则不需做预处理。

预处理后，如远动信息匹配那么远动信息表

的名称 f 应该是主站系统名称 O 的子字符串,直接使用 KMP 算法匹配,无法匹配的报错。方法能够很好地完成遥测、双位置遥信以及遥控运动信息的校核,单位置遥信的校核则因其复杂度较高需另作讨论。

(2) 改进 KMP 算法匹配

实际应用中发现单位位置遥信名称有两类问题:1) 描述不规范:单位位置遥信名称较长,字符种类多,手工输入时会出现非关键信息的错字、漏字或插入空格字符等情况,会无法匹配上并报错,但这些不规范问题并不影响信号的辨识和理解;2) 关键性错误:单位位置遥信名称中弄错间隔或信号类型等重大错误,直接影响调度监视的关键错误,这些错误必须及时提示并整改。描述不规范问题的大量报错会淹没掉关键信息错误,因此单位位置遥信的校核中有必要过滤这些不规范问题,将关键错误凸显出来。

在使用 KMP 算法进行两个字符串匹配时,匹配上的最大连续字符的长度表征了其匹配度,长度越大说明字符串的匹配程度越高,而当最大连续匹配字符的长度等于远动信息表的远动信息名称 f 时,为完全匹配。

本文基于 KMP 算法提取出一个匹配度控制参数 MLP (Match Length Parameter),当最大连续匹配长度大于 MLP 时,则认为通过匹配度检验,如小于 MLP,则认为未通过检验进而发出报警。MLP 参数可设置,通过配置其大小设置匹配算法的精度。实际使用中,首先设置 MLP 较小值,进行粗校核,然后根据实际情况提高 MLP 的值,进行细校核。

粗校核的目的是过滤掉不规范问题,将关键性的重大错误凸显出来并予以首先解决。实际应用经验表明粗校核中 MLP 能包括调度编号就能筛选出 95%以上的错误,因此进行粗校核时 MLP 值一般选取 4~5。

细校核的目的在于实现名称的全匹配,将人工维护中的不规范问题全部展示出来。MLP 数值越大,则精度越高,而极端情况是普通的 KMP 算法匹配。

实际应用中采用粗匹配和精匹配相结合的方法对单位位置遥信名称进行校核,通过 MLP 参数的设置将名称中的关键性错误和不规范问题

分离出来,首先用粗匹配校核关键性错误,再辅以精匹配校核不规范问题。

3 运动数据校核流程

远动信息数据校核的具体流程如下图所示。

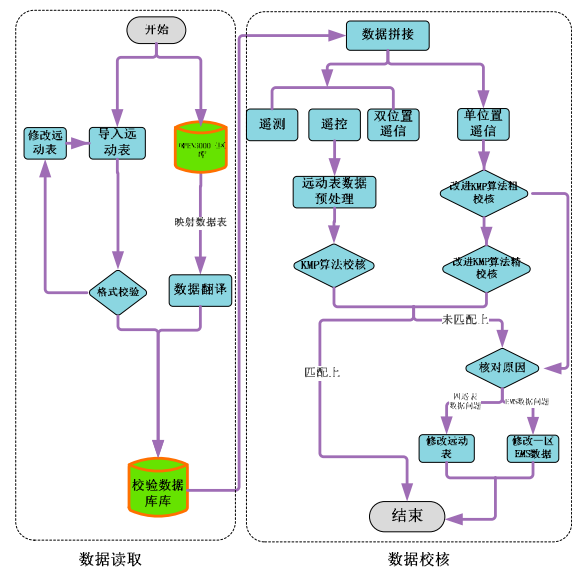


图 远动数据校核流程图

远动信息表格式校验 :在将远动信息表数据导入校验数据库时按照远动信息表格式规范对表格进行格式校验,不满足要求的信息表必须修改规范后完成导入。

数据读取 :将远动信息表数据导入到校验数据库 ;将主站系统三区数据库的远动数据映射到校核数据库。

数据拼接 :将两方数据以变电站为单位,以顺序编号为基准,将相同顺序编号的远动名称拼接起来,作为相互校核的对象。

数据校核 :对遥测、遥控及双位置遥信名称进行预处理,采用基本的 KMP 算法校核,将无法匹配的远动信息进行报警发布 ;对单位位置遥信名称则采用改进 KMP 算法校核,粗校核和细校核相结合,优先解决关键性错误。

数据整改 :根据数据校核结果,查明错误原因,相应地修改远动信息表或主站系统的数据库。

4 应用情况

本文以远动数据校核流程为基础开发了远

动信息数据校核工具,工具基于采用 C/S 的软件架构,采用 Microsoft Visual Studio 2008 开发,数据库使用 Oracle 11g,工具界面采用开源的多标签窗体控件 DockPane。

4.1 算法测试

在应用中分别采用朴素匹配算法和本文提出的改进 KMP 算法,对某电网的远动信息进行校核,从算法效率与关键错误筛选能力上比较,结果如下表。

表 匹配算法比较情况

	效率 /s	筛选能力 (关键性错误个数/报错个数)	
		单位置信号	其他
朴素匹配算法	20 左右	41.2% (2 095/5 088)	99.4% (1 072/1 078)
改进 KMP 算法 (MLP 取 5)	5 左右	99.2% (2 119/2 137)	99.4% (1 072/1 078)

对某中调电网的远动信息共计 251 593 条数据进行测试,通过比较可以看到,改进 KMP 算法完成全网远动信息的校核仅需要 5 s,效率相较朴素匹配算法提高了约 4 倍;而筛选关键错误方面,因双位置遥信、遥控以及遥测能够完全匹配,因此准确度均能接近 100%,而单位位置遥信的校核中,朴素匹配算法没有过滤能力,算法输出的错误结果中包括了大量不影响信号监视的不规范问题,难以辨识。而改进 KMP 算法通过合理设置匹配度控制参数 MLP,可以较为准确地筛选出关键错误,关键错误比例由 41.2%提高到了 99.2%,能够迅速定位远动信息的错误,及时揭示风险。

4.2 应用效果

某中调调度自动化系统目前包含遥测点 37 000 个,遥信点 190 800 个,遥控点 23 793 个,每年遥测、遥信、遥控等维护量在 10 万点以上。远动数据维护工作量大,人工成本很高,潜在风险也很高。

将校核工具在该中调应用后,完成了对远动信息表格式的校验以及电网远动信息的全数据校核,共计 251 593 条数据。程序自动校核时间

在 5 s 左右,极大地提高了效率和人力成本。而通过校核后的人工统计,该方法校核的准确率达到 99%,满足精度要求。

数据维护人员只需对工具校核输出的结果进行审核和确认,完成整改工作。通过校核工具的应用,远动信息数据的校核和维护工作极大地提高了效率和防控风险能力。

5 结束语

本文针对调度自动化系统远动信息的特点和管理方法,引入了高效的字符串匹配算法,提出了基于 KMP 算法的电网远动数据校核方法,并在此基础上引入了匹配度控制参数 MLP,通过对参数的合理设置,能够有效地分离出远动信息的关键性错误和名称不规范问题,便于数据维护人员迅速定位远动信息关键错误,降低电网监控的风险。该方法在某地电网 EMS 系统的应用情况,说明了该方法的有效性和科学性。

参考文献

[1] 赵巧云,张海江.无人值班变电站远方监控“远动”信息的配置与规范[J].中国电力:技术版,2012,1.

[2] 算法导论[M].高等教育出版社,2002.

[3] Knuth D E, Morris, Jr J H, Pratt V R. Fast pattern matching in strings[J]. SIAM journal on computing, 1977, 6 (2): 323-350.

[4] 胡玲霞,邱军旗.“远动”信息管理系统的设计[J].科技资讯,2006,35:008.

[5] 王向东,宋新航,王文海,等.基于 SCADA 的全息数据应用研究[J].价值工程,2012,31 (3): 140-140.

作者简介

郝蛟(1990-),男,助理工程师,主要从事调度自动化系统运行维护管理工作。

(收稿日期:2015-12-10) EA
(编辑:朱金凤)