

短波话音数据监测分析方法研究

汪 晗¹, 王海京¹, 王 坤², 汤强晋¹, 郭凯瑞¹

¹94860部队, 江苏 南京

²陆军工程大学指挥控制工程学院, 江苏 南京

收稿日期: 2022年9月29日; 录用日期: 2022年12月22日; 发布日期: 2022年12月31日

摘 要

本文介绍了一种短波话音数据监测采集和分析挖掘方法。首先设计实现图形化人机交互界面对短波通联属性信息进行监测采集存储; 其次在存储数据库的基础上, 对监测数据进行分析挖掘。通过对各参数进行单项分析揭示了内部规律; 通过对多维参数间进行多项分析和关联分析, 揭示了短波话音通联质量与其他通联参数间的关联规律, 为短波收/发电台选取、优化短波通信频率提供了依据。

关键词

短波话音, 监测采集, 分析挖掘, 关联规则

Exploration on Short-Wave Data Monitoring and Analysis Method

Han Wang¹, Haijing Wang¹, Kun Wang², Qiangjin Tang¹, Kairui Guo¹

¹94860 PLA Troops, Nanjing Jiangsu

²College of Command and Control Engineering, PLA Army Engineering University, Nanjing Jiangsu

Received: Sep. 29th, 2022; accepted: Dec. 22nd, 2022; published: Dec. 31st, 2022

Abstract

This paper introduces a method of short-wave voice data monitoring, collecting, analyzing and mining. Firstly, a graphical human-computer interaction interface is designed to monitor, collect and store the short-wave communication attributes; Secondly, based on the stored database, the monitoring data is analyzed and mined. The internal relationship of each parameter is revealed through single-term analysis; The relationship among multi-dimensional parameters is analyzed through multiple term analyses and correlation analysis. The correlation rules between the quality of short-wave voice communication and other parameters are revealed, which provide bases for

short-wave receiver/transmitter selection and frequency optimization of short-wave voice communication.

Keywords

Short-Wave Voice, Monitoring and Collection, Analysis and Mining, Correlation Rules

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

1. 引言

短波通信频率在 3~30 MHz, 对应波长范围为 100~10 m。由于短波可以利用电离层反射进行天波传输, 可实现几百乃至上万千米通信, 具有通信距离远、开通迅速、机动灵活、抗毁性强、机动重建性好等优点[1], 是军事通信和应急通信的重要手段, 在军事、气象、航空、外交等领域具有广泛的应用[2]。在大数据时代[3], 有必要对短波语音通联过程中产生的大量数据进行实时动态监测, 并以极低的成本采集和存储, 形成短波语音监测大数据; 基于历史监测大数据进行分析挖掘, 从中提取出短波语音通联规律[4], 科学高效地实现短波收/发电台的选择(切换), 合理分配短波通信频率, 对提高短波语音通联效果具有重要意义。

短波语音通联效果的影响因素较多, 为提高短波通信可靠性, 早期主要依靠短波通信模型预测、人工分析、经验选频等方法[5] [6] [7], 这些方法效率较低且准确性不高。近年, 美国开展第 4 代短波自动链路建立工作, 实现短波全频段监测感知、信号捕获等实现快速建链[8], 但这些探测基于点对点进行, 对于短波组网而言, 需要为全网节点规划探测过程, 存在探测时间长、操作复杂、效率低的问题。[9]设计了一种基于数据驱动的短波监测系统, 将短波组网监测过程中产生的大量数据, 进行采集存储, 分析挖掘, 制定监测策略, 提高短波监测效率。[10]使用统计学方法对短波监测数据进行分析, 以增强短波监测数据的可用性。本文立足点与之不同, 本文目标是提高短波语音通联效果。为实现该目标, 首先设计实现了短波语音数据采集软件, 对影响短波语音通联质量的关键参数进行采集存储, 并使用数据分析挖掘方法揭示参数内部规律及参数间的关联规律, 为短波收/发电台选取、优化短波通信频率提供了依据。

2. 短波数据监测采集

为了更好地掌握短波电台使用、频率设置情况, 提高短波语音通联质量, 需要对短波语音通联数据进行监测采集存储, 积累形成短波语音监测大数据。日常短波语音监测工作采集的主要参数有: 短波发射电台、短波接收电台、时间、天气、发射频率、语音通联质量等。

为了便于采集存储, 设计实现了图形化短波语音数据采集人机交互界面, 如图 1 所示。通过操作界面, 可以将每次短波语音通联的属性信息有效保存, 并可以将短波终端机输出的语音数据以指定的采样频率和采样精度存储为.wav 文件, 便于后期分析挖掘。

短波语音通联的属性信息包含短波信号的频率、发射电台、接收电台、监测时间、当时天气、语音质量等数据, 每一项均称为数据项。本系统中使用的数据项的数据类型包含双精度浮点数字类型, 例如频率和语音质量; 字符串文本类型, 例如发射电台/接收电台名称、天气; 时间类型, 例如数据产生的标签时间等。监测采集的短波语音通联属性数据格式如表 1。

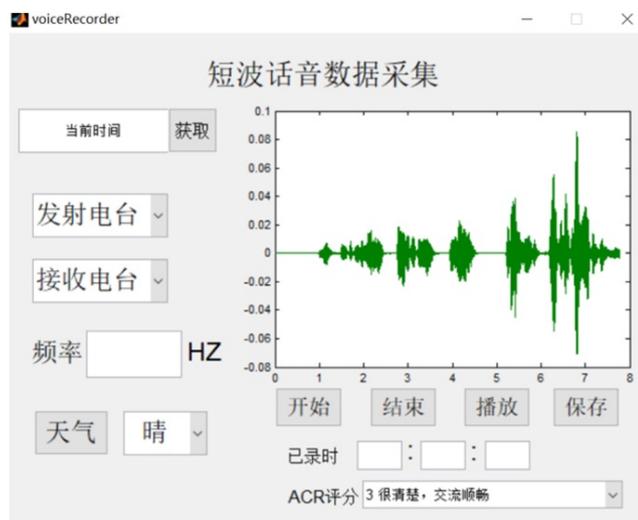


Figure 1. Interface of short-wave voice data acquisition
图 1. 短波话音数据采集界面

Table 1. Data format of short-wave voice communication attributes

表 1. 短波话音通联属性数据格式

监测时间	发射电台	接收电台	频率	话音质量	天气
20220402204834	发 1	收 1	6800	3	晴
20220403091523	发 1	收 2	12000	1	阴

将以上数据项在数据库中存储，可以从中提取各种数据表，其中“时间”项为时间格式，“发射电台”、“接收电台”和“天气”项为字符串格式，“频率”和“话音质量”为数字格式。

3. 短波监测数据分析

在短波话音监测信息存储数据库的基础上，设计实现短波数据管理模块，在人机交互分析界面上，人工选择数据源，按照每条属性逐一选择过滤条件，之后选择算法，选择统计的数据项，最后在界面上展示统计结果图或分析结果，操作流程如图 2。根据数据分析对象的项目数，可以分为单项分析和多项分析。



Figure 2. Flow chart of statistical analysis
图 2. 统计分析流程图

3.1. 单项分析

由于数据库中存在数字、字符串和时间三种不同的数据类型，因此对其进行统计分析挖掘需要使用不同的方法。

3.1.1. 数字项的分析方法

对单项为数字的数据可采用分布直方图来显示数据的分布情况，例如对短波监测数据中的使用频率

进行统计分析,通过频率分布直方图可直观体现频率的分布规律。示意图如下图 3,从图中可直观看出频率在历史监测数据中的实际使用情况。

短波话音质量分为“3、2、1”三类,分别对应三类不同的人耳听觉感受“好(话音清晰,交流顺畅)、中(勉强能听到)、差(有人声听不懂)”。短波话音监测数据分析挖掘的目的在于优化频率选择,提高短波通联质量,因此可以对短波话音质量为“3”的数据进行着重统计分析。例如,从数据库中提取“话音质量”值等于“3”的数据项,并将其对应的频率用直方图的形式展示,如图 3 所示,从中可直观看出使用不同频率时通联质量“好”的统计次数。

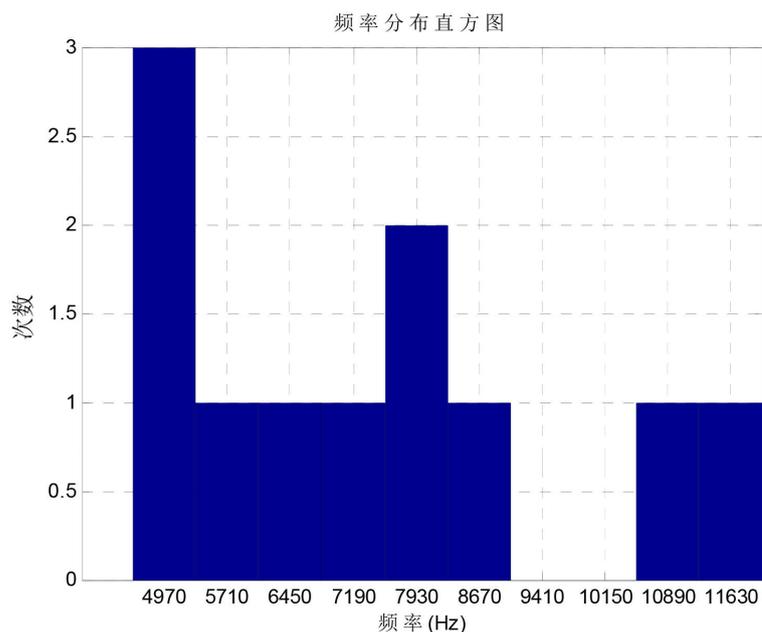


Figure 3. Histogram of monitoring frequency distribution
图 3. 监测频率分布直方图

3.1.2. 时间项的分析方法

由于短波通信依赖电离层反射进行天波传播,而电离层的活跃程度受时间(昼/夜)的影响较大,在实际短波通联时,为提高通联质量,需要在不同的时间段使用不同的频率。为了分析在历史通联数据中,不同时间段内的频率使用情况及相应的通联质量,可以对表 1 中的“监测时间”属性进行单项分析统计,以探究短波通联的时间分布规律。

在数据库中存储的“监测时间”是标准的“yyyy-MM-dd-HH-mm-ss”格式,在实际分析中,将其转换为字符串格式,并只取“yyyy-MM-dd-HH”部分进行统计,对于分钟 mm 和秒 ss 不作统计。

对以上简述的单项或多项数据统计按照时间分类,又可得到数据的时间变化情况,例如对频率单项分析后按照时间分类则可知某年某月的频率使用情况。

3.2. 多项分析

由于短波通联质量是诸多因素(发/收站、频率、时间、天气)综合作用的结果,仅依靠单项分析无法厘清因素间的相关关系,需要进行多项分析。但单项分析是基础,可以基于单项分析结果,进一步对不同因素进行组合分析,揭示多项因素间的规律分布。例如,在上述对时间单项分析的基础上,可以对 xxxx 年 xx 月的二维数据集集合进一步分析,从中分析当月频率的使用规律以及对应的语音质量分布,并用图、

表的形式可视化直观显示；也可实际运用于频率与收/发电台等要素之间的综合分析，例如从历史数据中提取特定的收/发电台组合后，对相应的频率进行单项数据分析，可以获得给定收/发电台组合下短波通信的常用频率分布。

在以话音质量为核心的分析中，需要在部分相关因素确定的前提下，例如给定发射/接收电台和时间，分析其他因素，例如频率与话音质量之间的关系；或者指定频率和时间，统计不同发射/接收电台与话音质量之间的关系；或者指定频率和发射/接收电台，统计分析时间与话音质量之间的关系。

3.3. 关联规则分析

综合考虑各种数据挖掘算法的优缺点[11]，采用 Apriori 算法对短波通信要素之间的关联规则进行挖掘。Apriori 算法本质由两部分组成：频繁项集和关联规则。频繁项集的计算主要是指频繁项集的支持度计算；关联规则的计算主要包括关联规则的置信度计算。

3.3.1. 频繁项集的支持度(Support)和阈值

一个项集的支持度被定义为数据集中包含该项集的记录所占的比例。阈值通常手动设置，比如设置阈值为 30%，这样支持度小于该阈值的时候，就认为该项集并不频繁，也就没有挖掘关联规则的意义，因此不会进行后续的计算。

3.3.2. 频繁项集的计算方法

频繁项集的计算方法主要是三步的不断循环：计算当前频繁项的所有并集，得到新的频繁项候选集；对得到的所有频繁项候选集进行支持度的计算；使用阈值过滤掉不符合要求的频繁项，最终剩下的成为新的频繁项。流程图如图 4 所示。

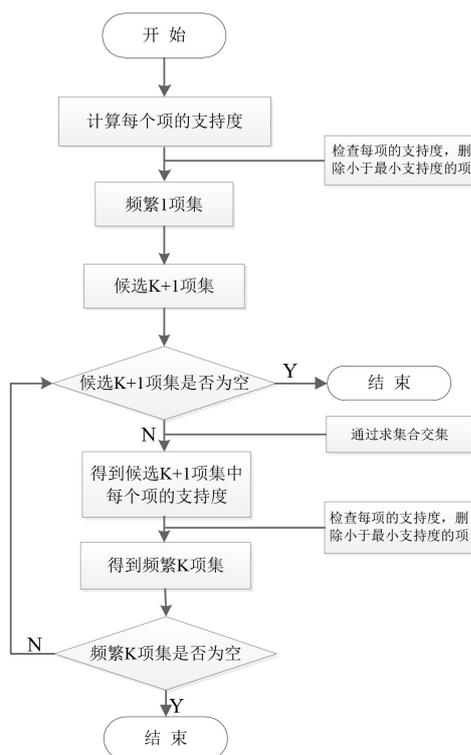


Figure 4. Flow chart of frequent item set calculation

图 4. 频繁项集计算流程图

3.3.3. 关联规则挖掘

关联规则的意思是通过某个数据项可以推导出另一个数据项，反映了一个数据项与其他数据项之间的相互依赖性，如果多个数据项之间存在某种关联关系，那么就能通过其他数据项对某一数据项进行预测[12]。根据频繁项集可以挖掘出关联规则，比如一个频繁项集{P1, P2, P3}，就可以推导出六个可能的关联规则：

{P1} → {P2, P3}
 {P2} → {P1, P3}
 {P3} → {P1, P2}
 {P1, P2} → {P3}
 {P2, P3} → {P1}
 {P1, P3} → {P2}

其中置信度最高的就是最有可能的关联规则[13]。通过在短波语音监测数据集上运行 Apriori 算法，可以挖掘出短波通联因素的频繁项集，从中提取关联规则。部分短波语音监测数据集如表 2 所示。

Table 2. Dataset of short-wave voice monitoring
表 2. 短波语音监测数据集

监测时间	发射电台	接收电台	频率	话音质量
20220402204834	发 1	收 1	7800	3
20220402112306	发 1	收 1	7800	3
20220403091523	发 1	收 2	12000	1
20220405111012	发 1	收 2	7800	3

在表 2 中，监测时间是标准的“yyyy-MM-dd-HH-mm-ss”格式，在前面对时间项的分析中，将其转换为字符串格式，并只取“yyyy-MM-dd-HH”部分进行统计；在关联规则挖掘时，考虑到短波信道传播和短波话音通信保障的实际情况，将“小时”字段“HH”以“07”和“17”为界分为“昼间 day”和“夜间 night”，对表 2 数据执行 Apriori 算法示意图如图 5。

Apriori 算法在表 2 的短波数据集上逐层搜索频繁项集，经过五次迭代运行连接和剪枝步骤，得出最终的频繁项集{day, 发 1, 收 1, 7800, 3}，算法到此结束，随后根据频繁项集找出强关联规则。

在{day, 发 1, 收 1, 7800, 3}中，对非空子集进行组合，部分关联规则如下：

{day, 发 1, 收 1, 7800} ≥ {3}，置信度 = $P(\{3\}|\{day, 发 1, 收 1, 7800\}) = 2/2 = 100\%$ ；

{day, 发 1, 7800, 3} ≥ {收 1}，置信度 = $P(\{收 1\}|\{day, 发 1, 7800, 3\}) = 2/3 = 67\%$ ；

{day, 收 1, 7800, 3} ≥ {发 1}，置信度 = $P(\{发 1\}|\{day, 收 1, 7800, 3\}) = 2/2 = 100\%$ ；

{day, 发 1, 收 1, 3} ≥ {7800}，置信度 = $P(\{7800\}|\{day, 发 1, 收 1, 3\}) = 2/2 = 100\%$ ；

{发 1, 收 1, 7800, 3} ≥ {day}，置信度 = $P(\{day\}|\{发 1, 收 1, 7800, 3\}) = 2/2 = 100\%$ ；

若最小置信阈值设置为 90%，则规则{day, 发 1, 收 1, 7800} ≥ {3}、{day, 收 1, 7800, 3} ≥ {发 1}、{day, 发 1, 收 1, 3} ≥ {7800}、{发 1, 收 1, 7800, 3} ≥ {day}可以输出，这就是强规则。

例如，{day, 收 1, 7800, 3} ≥ {发 1}，从该规则可以得出，给定工作时间(昼 day)、接收电台(收 1)、任务用频(7800 KHZ)，若要短波话音通联质量好(3 分)，可以使用的发射电台(发 1)；同理，根据规则{day, 发 1, 收 1, 3} ≥ {7800}可以在给定工作时间(昼 day)、发射电台(发 1)和接收电台(收 1)的情况下，推荐给话音通联质量好(3 分)的频率(7800 KHZ)。

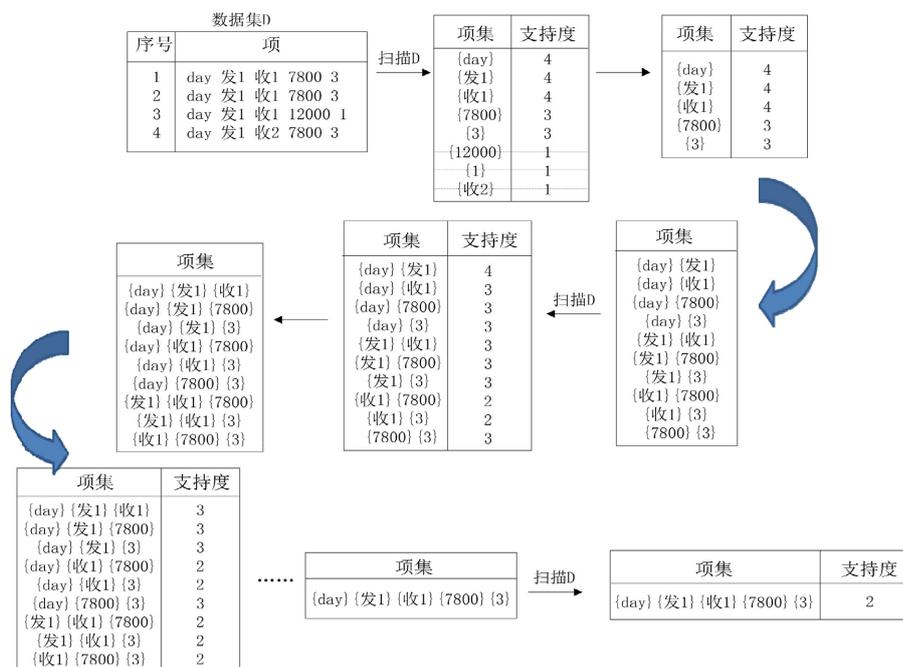


Figure 5. Schematic diagram of Apriori algorithm running on short-wave dataset
图 5. Apriori 算法在短波数据集上的运行示意图

4. 结束语

本文介绍了短波话音数据监测采集和分析挖掘方法。首先通过图形化人机交互界面对短波通联属性信息进行监测采集存储，其次在存储数据库的基础上，对监测数据进行分析挖掘。通过单项分析揭示各参数的内部关系，通过多项分析和关联分析对参数间的关系进行多维分析。从繁杂的短波话音监测数据表中提取有用信息，揭示短波话音通联质量与其他通联属性参数间的关联规律，最终形成大量有效的信息流，为短波收/发电台选取、优化短波通信频率设置提供依据。

参考文献

- [1] 周忠超, 刘曦元. 一种数据挖掘方法在短波监测中的应用[J]. 数字通信世界, 2019(3): 37-38.
- [2] 王金龙, 陈瑾, 徐煜华. 短波通信技术研究进展与发展需求[J]. 陆军工程大学学报, 2022, 1(1): 1-7.
- [3] 刘博. 通信行业大数据分析及应用研究[J]. 通讯世界, 2017(6): 139.
- [4] 文世敏. 大数据的分类挖掘优化技术创新发展[J]. 现代信息科技, 2018(7): 87-89.
- [5] International Telecommunication Union (2007) Method for the Prediction of the Performance of HF Circuit. 1-24.
- [6] 王斌, 胡中豫, 杨丽芬, 等. 一种短波远程通信选频估计方法[C]//中国通信学会. 2011年亚太青年通信学术会议. USA: Scientific Research Publishing, 2011: 57-60.
- [7] 张春艳, 徐开军, 王书旺, 等. 短波通信实时选频技术研究及其实现[J]. 电子世界, 2021(19): 18-19.
- [8] 周相成, 吉彬, 胡飞. 海洋短波组网通信自动选频[J]. 广东海洋大学学报, 2020, 40(2): 89-93.
- [9] 刘师范. 基于数据驱动方法的短波监测系统的设计[J]. 通讯世界, 2019(12): 14-15.
- [10] 刘曦元, 周忠超. 数据挖掘在短波监测数据分析中的应用设计[J]. 数字通信世界, 2019(2): 37-38.
- [11] 张启徽. 关联规则挖掘查找频繁项集的改进算法[J]. 统计与决策, 2015(4): 32-35.
- [12] 陈弘扬, 江建兴, 林于新. 基于 WEKA 的短波监测数据关联性分析方法浅析[J]. 中国无线电, 2021(6): 75-77.
- [13] 葛俊言. 数据挖掘技术的应用研究[J]. 中国新通信, 2018(15): 84-85.