说 明 书 摘 要

本发明涉及一种电信网络告警预测方法，包括：按时间顺序分批读入电信网络告警数据并对所述告警数据预处理；对预处理后的告警数据按照Unix时间戳进行DBSCAN聚类；对DBSCAN聚类后的告警数据挖掘频繁项集和告警关联规则；合并并保存所有批次告警数据挖掘出的告警关联规则；利用合并并保存的上述告警关联规则预测未发生的告警。本发明能够很好地处理具有突发性特点的告警数据，使用时间聚类提高了关联规则的有效性，采用分批处理的思路提高了处理大批量数据的能力。

摘 要 附 图



权 利 要 求 书

1.一种电信网络告警预测方法，其特征在于，该方法包括如下步骤：

a.按时间顺序分批读入电信网络告警数据并对所述告警数据预处理；

b.对预处理后的告警数据按照Unix时间戳进行DBSCAN聚类；

c.对DBSCAN聚类后的告警数据挖掘频繁项集和告警关联规则；

d.合并并保存所有批次告警数据挖掘出的告警关联规则；

e.利用合并并保存的上述告警关联规则预测未发生的告警。

2.如权利要求1所述的方法，其特征在于，所述的步骤a具体包括：

步骤a1，读取告警记录，对每一条告警记录提取相应特征信息；

步骤a2，检查该条告警故障原因是否存在故障表中，如是，则进入步骤a3；如否，则丢弃并返回步骤a1；

步骤a3，检查数据库中该告警的告警类型字段是否属于“告警清除”类型，如是，丢弃并返回步骤a1；如否，进入步骤a4；

步骤a4，检查内存中是否存在网元、故障原因、故障类型均相同，且故障发生时间绝对值之差小于给定阈值的告警记录，如是，则丢弃并返回步骤a1；如否，进入步骤a5；

步骤a5，将该告警添加进内存；

重复步骤a1到步骤a5，直至该批告警数据全部被处理完。

3.如权利要求2所述的方法，其特征在于，所述的步骤a1具体包括：

从数据库中读取告警记录，对每一条告警记录提取相应特征信息，组成以下数据格式：{网元、故障原因、故障类型、故障发生时间}。

4.如权利要求3所述的方法，其特征在于，所述的步骤b具体包括如下步骤：

步骤b1，将告警数据格式中的发生时间字符串转换成Unix时间戳；

步骤b2，使用DBSCAN算法处理时间列数据；

步骤b3，使用轮廓系数评价聚类效果；

步骤b4，将聚类得出的簇标签结果添加到告警数据的最后一列：{网元、故障原因、故障类型、故障发生时间，簇类别}；

步骤b5，对事务中的告警按照时间顺序排序，并且去掉重复告警。

5.如权利要求4所述的方法，其特征在于，所述重复告警指：网元、故障原因、故障类型都相同的告警。

6.如权利要求5所述的方法，其特征在于，所述的步骤c具体包括如下步骤：

步骤c1，设定最小支持度阈值min\_sup，扫描所有聚类后得到的告警事务，对每个项进行计数，剔除出现小于min\_sup的项集，得到频繁1-项集，记为L1；

步骤c2，对L1进行迭代，生成候选2-项集，对候选2-项集进行剪枝处理得到C2；

步骤c3，对C2进行支持度计数，将小于min\_sup的项集剔除，得到L2；

步骤c4，重复步骤c2和步骤c3，生成L3，L4…Lk，直至Lk+1为空；

步骤c5，设定最小置信度阈值min\_conf，输出提升度大于1的强关联规则；

步骤c6，使用步骤c1到步骤c5的方法，分别单独挖掘告警中网元、故障原因的关联规则。

7.如权利要求6所述的方法，其特征在于，所述的步骤d具体包括如下步骤：

步骤d1，将第一批告警数据挖掘出的规则读进内存，格式为：{前件，后件，conf，samples}；

步骤d2，读进下一批告警数据挖掘出的规则，逐条对比规则是否存在内存当中，若否则添加；若已存在，则更新内存中该规则的conf及samples：，；

步骤d3，重复步骤d2直至所有规则合并完成，将规则保存为可读写文件。

8.如权利要求7所述的方法，其特征在于，所述的步骤e具体包括如下步骤：

步骤e1，读取存放规则的文件，将其转换为键值对；

步骤e2，按照时间顺序读入当前发生的告警，判断该告警与上一条读进的告警发生时间之差是否小于规定的阈值，如是，则进入步骤e3；如否，执行步骤e5；

步骤e3，将该告警添加到告警集合中，遍历上述键值对规则表，检查规则表中是否存在一行数据属于当前告警集合的子集，若是，执行步骤e4；若不是，继续执行步骤e2，扩大告警集合；

步骤e4，输出该规则的后件和概率并清空告警集合；

步骤e5，查找单独的属性关联规则表，判断是否存在当前网元和故障各自对应的后件，若有，进入步骤e6；若无，进入步骤e7；

步骤e6，记其置信度为p1和p2，将网元和故障所对应的后件组合成预测告警，并输出概率p=p1\*p2；

步骤e7，预测无故障发生；

清空告警集合，重复步骤e2到步骤e7，继续预测新的告警。

9.如权利要求8所述的系统，其特征在于，所述的键值对表示为，键为前件，值为后件及置信度：{key：前件，value：[后件，conf]}。

说 明 书

**电信网络告警预测方法及系统**

**技术领域**

本发明涉及一种电信网络告警预测方法及系统。

**背景技术**

随着电信网络规模的不断扩大，每天由电信设备故障产生的告警也越来越多，而且一个设备的故障经常会导致另一个设备的故障，从而给维护人员带来很大的麻烦。70年代有人提出利用专家知识建立告警关联规则库来处理电信网络产生的告警，但时至今日，这种依赖专家知识的告警关联工具显然不能很好地应付越来越复杂的网络结构。

由于近年来数据挖掘技术的发展，很多行业都开始使用这项技术来处理各自的业务，并取得了令人瞩目的效果，通信行业也不例外。在告警关联方面，运营商开始尝试使用数据挖掘技术对以往积累的大量告警历史信息进行分析，但复杂的业务逻辑和巨大的数据量给他们带来了很大的挑战。

目前电信网络中的告警关联规则一般是通过专家基于积累的相关经验进行总结提炼，然后讨论决定的。这种人工提取规则的方法，存在效率低、不完整、依赖性强等特点，无法很好适应当前复杂的电信网络结构。

已经有一些使用数据挖掘技术进行告警关联的案例，比较经典的算法是WinEPI。WinEPI通过设置固定的时间窗口宽度来提取告警序列，并发现告警在时间上的偏序关系。但由于告警序列通常是一组不均匀的数据，往往在一个短的时间段内密集产生，之后一段时间就恢复平静。且不同时间段内产生的告警事件频率、持续时间也各不相同，如果使用固定的时间窗口宽度来提取告警事务则可能存在很多无效的数据，导致最终提取的关联规则无效。

**发明内容**

有鉴于此，有必要提供一种电信网络告警预测方法及系统。

本发明提供一种电信网络告警预测方法，该方法包括如下步骤：a.按时间顺序分批读入电信网络告警数据并对所述告警数据预处理；b.对预处理后的告警数据按照Unix时间戳进行DBSCAN聚类；c.对DBSCAN聚类后的告警数据挖掘频繁项集和告警关联规则；d.合并并保存所有批次告警数据挖掘出的告警关联规则；e.利用合并并保存的上述告警关联规则预测未发生的告警。

其中，所述的步骤a具体包括：

步骤a1，读取告警记录，对每一条告警记录提取相应特征信息；

步骤a2，检查该条告警故障原因是否存在故障表中，如是，则进入步骤a3；如否，则丢弃并返回步骤a1；

步骤a3，检查数据库中该告警的告警类型字段是否属于“告警清除”类型，如是，丢弃并返回步骤a1；如否，进入步骤a4；

步骤a4，检查内存中是否存在网元、故障原因、故障类型均相同，且故障发生时间绝对值之差小于给定阈值的告警记录，如是，则丢弃并返回步骤a1；如否，进入步骤a5；

步骤a5，将该告警添加进内存；

重复步骤a1到步骤a5，直至该批告警数据全部被处理完。

所述的步骤a1具体包括：

从数据库中读取告警记录，对每一条告警记录提取相应特征信息，组成以下数据格式：{网元、故障原因、故障类型、故障发生时间}。

所述的步骤b具体包括如下步骤：

步骤b1，将告警数据格式中的发生时间字符串转换成Unix时间戳；

步骤b2，使用DBSCAN算法处理时间列数据；

步骤b3，使用轮廓系数评价聚类效果；

步骤b4，将聚类得出的簇标签结果添加到告警数据的最后一列：{网元、故障原因、故障类型、故障发生时间，簇类别}；

步骤b5，对事务中的告警按照时间顺序排序，并且去掉重复告警。

所述重复告警指：网元、故障原因、故障类型都相同的告警。

所述的步骤c具体包括如下步骤：

步骤c1，设定最小支持度阈值min\_sup，扫描所有聚类后得到的告警事务，对每个项进行计数，剔除出现小于min\_sup的项集，得到频繁1-项集，记为L1；

步骤c2，对L1进行迭代，生成候选2-项集，对候选2-项集进行剪枝处理得到C2；

步骤c3，对C2进行支持度计数，将小于min\_sup的项集剔除，得到L2；

步骤c4，重复步骤c2和步骤c3，生成L3，L4…Lk，直至Lk+1为空；

步骤c5，设定最小置信度阈值min\_conf，输出提升度大于1的强关联规则；

步骤c6，使用步骤c1到步骤c5的方法，分别单独挖掘告警中网元、故障原因的关联规则。

所述的步骤d具体包括如下步骤：

步骤d1，将第一批告警数据挖掘出的规则读进内存，格式为：{前件，后件，conf，samples}；

步骤d2，读进下一批告警数据挖掘出的规则，逐条对比规则是否存在内存当中，若否则添加；若已存在，则更新内存中该规则的conf及samples：，；

步骤d3，重复步骤d2直至所有规则合并完成，将规则保存为可读写文件。

所述的步骤e具体包括如下步骤：

步骤e1，读取存放规则的文件，将其转换为键值对；

步骤e2，按照时间顺序读入当前发生的告警，判断该告警与上一条读进的告警发生时间之差是否小于规定的阈值，如是，则进入步骤e3；如否，执行步骤e5；

步骤e3，将该告警添加到告警集合中，遍历上述键值对规则表，检查规则表中是否存在一行数据属于当前告警集合的子集，若是，执行步骤e4；若不是，继续执行步骤e2，扩大告警集合；

步骤e4，输出该规则的后件和概率并清空告警集合；

步骤e5，查找单独的属性关联规则表，判断是否存在当前网元和故障各自对应的后件，若有，进入步骤e6；若无，进入步骤e7；

步骤e6，记其置信度为p1和p2，将网元和故障所对应的后件组合成预测告警，并输出概率p=p1\*p2；

步骤e7，预测无故障发生；

清空告警集合，重复步骤e2到步骤e7，继续预测新的告警。

所述的键值对表示为，键为前件，值为后件及置信度：{key：前件，value：[后件，conf]}。

本发明提出了一种电信网络告警预测方案，通过这种方案能实现动态的时间窗口宽度，从而能很好地处理具有突发性特点的告警数据；对于提取出的告警事务数据库，采用Apriori算法进行关联规则挖掘；最后通过规则匹配和多维属性概率相乘的方法预测告警。本发明使用时间聚类提高了关联规则的有效性，采用分批处理的思路提高了处理大批量数据的能力，并给出了一种预测告警的方案。

**附图说明**

图1为本发明电信网络告警预测方法的流程图；

图2为本发明实施例步骤S1对所述告警数据预处理的流程图；

图3为本发明实施例步骤S2对预处理后的告警数据按照Unix时间戳进行DBSCAN聚类的流程图；

图4为本发明实施例步骤S3对DBSCAN聚类后的告警数据挖掘频繁项集和告警关联规则的流程图；

图5为本发明实施例步骤S4合并并保存所有批次告警数据挖掘出的告警关联规则的流程图；

图6为本发明实施例步骤S5利用合并并保存的上述告警关联规则预测未发生的告警的流程图。

**具体实施方式**

下面结合附图及具体实施例对本发明作进一步详细的说明。

参阅图1所示，是本发明电信网络告警预测方法较佳实施例的作业流程图。

步骤S1，按时间顺序分批读入电信网络告警数据并对所述告警数据预处理。具体而言，请一并参阅图2：

步骤S11，从数据库中读取若干条告警记录，对每一条告警记录提取相应特征信息，组成以下数据格式：{网元、故障原因、故障类型、故障发生时间}。

步骤S12，检查该条告警故障原因是否存在故障表中，如是进入步骤S13；如否说明该告警可能是记录出错的脏数据，丢弃并回到步骤S11。

步骤S13，检查数据库中该告警的告警类型字段是否属于“告警清除”类型，如是，丢弃并回到步骤S11；如否，进入步骤S14。

步骤S14，检查内存中是否存在网元、故障原因、故障类型均相同，且故障发生时间绝对值之差小于给定阈值的告警记录，如是，说明该告警是重复告警，丢弃并回到步骤S11；如否，进入步骤S15。

步骤S15，将该告警添加进内存。

重复步骤S11到步骤S15，直至该批告警数据全部被处理完。

步骤S2，对预处理后的告警数据按照Unix时间戳进行DBSCAN聚类。具体而言，请一并参阅图3：

步骤S21，将告警数据格式中的发生时间字符串转换成Unix时间戳，并将时间列数据单独提取出来。

步骤S22，使用DBSCAN算法处理时间列数据，算法的参数eps需要调节，min\_samples设为2，评价标准使用曼哈顿距离（Manhattan Distance）：，其中分别代表告警发生的时间。

步骤S23，使用轮廓系数（Silhouette Coefficient）来评价聚类效果的好坏，对于样本，轮廓系数的定义如下：，为样本到同簇其他样本的平均距离，为样本到其他簇所有样本的平均距离。越接近1，说明聚类越合理。

步骤S24，将聚类得出的簇标签结果添加到告警数据的最后一列：{网元、故障原因、故障类型、故障发生时间，簇类别}，根据所在类别划分告警事务，同一个类别的告警属于同一事务。

步骤S25，对事务中的告警按照时间顺序排序，并且去掉重复告警。其中所述重复告警指：网元、故障原因、故障类型都相同的告警。

步骤S3，对DBSCAN聚类后的告警数据挖掘频繁项集和告警关联规则。具体而言，请一并参阅图4：

步骤S31，设定最小支持度阈值min\_sup，扫描所有聚类后得到的告警事务，对每个项进行计数，剔除出现小于min\_sup的项集，得到频繁1-项集，记为L1。

步骤S32，对L1进行迭代，生成候选2-项集。根据Apriori的前提假设：如果一个项集是非频繁集，那么它的所有超集也是非频繁的，因此需要对候选2-项集进行剪枝处理，即剪掉候选2-项集中包含的不频繁1-项集，得到C2。

步骤S33，对C2进行支持度计数，将小于min\_sup的项集剔除，得到L2。

步骤S34，重复步骤S32和步骤S33，生成L3，L4…Lk，直至Lk+1为空。

步骤S35，设定最小置信度阈值min\_conf，输出提升度大于1的强关联规则。提升度的计算方式如下：=

步骤S36，使用步骤S31到步骤S35的方法，分别单独挖掘告警中网元、故障原因的关联规则。

步骤S4，合并并保存所有批次告警数据挖掘出的告警关联规则。具体而言，请一并参阅图5：

步骤S41，将第一批告警数据挖掘出的规则读进内存，格式为：{前件，后件，conf，samples}。

步骤S42，读进下一批告警数据挖掘出的规则，逐条对比规则是否存在内存当中，若否则添加；若已存在，则更新内存中该规则的conf及samples：，。

步骤S43，重复步骤S42直至所有规则合并完成，将规则保存为可读写文件。

步骤S5：利用合并并保存的上述告警关联规则预测未发生的告警。具体包括，请一并参阅图6：

步骤S51，读取存放规则的文件，将其转换为键值对，键为前件，值为后件及置信度：{key：前件，value：[后件，conf]}。

步骤S52，按照时间顺序读入当前发生的告警，判断该告警与上一条读进的告警发生时间之差是否小于规定的阈值min\_time，如是，则进入步骤S53；如否，执行步骤S55。

步骤S53，将该告警添加到告警集合中，遍历上述键值对规则表，检查规则表中是否存在一行数据属于当前告警集合的子集，若是，执行步骤S54；若不是，继续执行步骤S52，扩大告警集合。

步骤S54，输出该规则的后件和概率并清空告警集合。

步骤S55，查找单独的属性关联规则表（即网元🡪网元，故障🡪故障），判断是否存在当前网元和故障各自对应的后件，若有，进入步骤S56；若无，进入步骤S57。

步骤S56，记其置信度为p1和p2，将网元和故障所对应的后件组合成预测告警，并输出概率p=p1\*p2。

步骤S57，预测无故障发生。

清空告警集合，重复步骤S52到步骤S57，继续预测新的告警。

虽然本发明参照当前的较佳实施方式进行了描述，但本领域的技术人员应能理解，上述较佳实施方式仅用来说明本发明，并非用来限定本发明的保护范围，任何在本发明的精神和原则范围之内，所做的任何修饰、等效替换、改进等，均应包含在本发明的权利保护范围之内。

说 明 书 附 图



图1



图2



图3



图4



图5



图6