

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

H04N 7/16 (2006.01)

H04N 5/00 (2006.01)

G06Q 30/00 (2006.01)



# [12] 发明专利申请公布说明书

[21] 申请号 200810222750.8

[43] 公开日 2009 年 1 月 28 日

[11] 公开号 CN 101355686A

[22] 申请日 2008.9.24

[21] 申请号 200810222750.8

[71] 申请人 中辉世纪传媒发展有限公司

地址 100029 北京市朝阳区裕民路 12 号中国  
国际科技会展中心 A 座 601

[72] 发明人 柴剑平 韩 光

[74] 专利代理机构 北京三友知识产权代理有限公司

代理人 樊一瑾

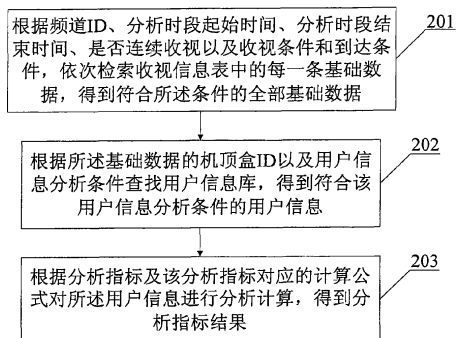
权利要求书 5 页 说明书 20 页 附图 8 页

## [54] 发明名称

收视率统计方法和系统

## [57] 摘要

本发明提供一种收视率统计方法和系统，所述方法包括：根据预先设置的分析条件中的频道 ID、分析时段起始时间 A、分析时段结束时间 B、是否连续收视以及收视条件和到达条件，依次检索收视信息表中的每一条基础数据，得到符合所述条件的全部基础数据；根据所述基础数据的机顶盒 ID 以及预先设置的分析条件中的用户信息分析条件查找用户信息库，得到符合所述用户信息分析条件的用户信息；根据预先设置的分析条件中的分析指标及该分析指标对应的计算公式，对所述用户信息进行分析计算，得到分析指标结果。 本发明可以提供准确及时的收视情况分析，数据准确率高，分析周期短。



1、一种收视率统计方法，其特征在于，所述方法包括下列步骤：

根据预先设置的分析条件中的频道 ID、分析时段起始时间 A、分析时段结束时间 B、最小收视时长 T 以及是否连续收视确定收视条件  $A < b \& B > a$  和到达条件  $t \geq T$ ，根据所述收视条件和到达条件依次检索收视信息表中的每一条基础数据，得到符合所述条件的全部基础数据；

根据所述基础数据的机顶盒 ID 以及预先设置的分析条件中的用户信息分析条件查找用户信息库，得到符合所述用户信息分析条件的用户信息；

根据预先设置的分析条件中的分析指标及该分析指标对应的计算公式，对所述用户信息进行计算，得到分析指标结果；

其中，所述每一条基础数据包括：机顶盒 ID、收视时段起始时间 a、收视时段结束时间 b 以及频道 ID；t 为分析时段内有效收视时长或分析时段内有效收视时长之和。

2、根据权利要求 1 所述的方法，其特征在于：

当所述 t 为分析时段内有效收视时长时：

$$t = \frac{(|A - b| + |a - B| - |a - A| - |b - B|)}{2};$$

当所述 t 为分析时段内有效收视时长之和时：

$$t = \frac{\sum_{i=1}^n (|A - b_i| + |a_i - B| - |a_i - A| - |b_i - B|)}{2 \times n}, \text{ 其中, } n \text{ 为一用户在分析时段内}$$

收看一频道节目的次数。

3、根据权利要求 1 所述的方法，其特征在于：

所述 T 或者为最小收视时长百分比，则所述 t 为分析时段内有效收视时长或分析时段内有效收视时长之和与分析时段时长的百分比。

4、根据权利要求 3 所述的方法，其特征在于：

当所述  $t$  为分析时段内有效收视时长与最小收视时长百分比时:

$$t = \frac{(|A-b| + |a-B| - |a-A| - |b-B|)}{2 \times (B-A)} \times 100\%;$$

当所述  $t$  为分析时段内有效收视时长之和与最小收视时长百分比时:

$$t = \frac{\sum_{i=1}^n (|A-b_i| + |a_i-B| - |a_i-A| - |b_i-B|)}{2 \times n \times (B-A)} \times 100\%, \text{ 其中, } n \text{ 为一用户在分}$$

析时段内收看一频道节目的次数。

5、根据权利要求 1 或 3 所述的方法, 其特征在于, 所述检索收视信息表中的基础数据的步骤包括:

判断所述基础数据中的频道 ID 与所述分析条件中的频道 ID 是否相等, 同时判断所述基础数据中的收视时段起始时间  $a$  和收视时段结束时间  $b$  是否满足所述收视条件, 并判断  $t$  是否满足所述到达条件, 如果是, 则提取所述基础数据的机顶盒 ID, 并保存该机顶盒 ID。

6、根据权利要求 1 或 3 所述的方法, 其特征在于, 根据所述机顶盒 ID 以及预先设置的分析条件的用户信息分析条件查找用户信息库, 得到符合所述条件的用户信息的步骤包括:

根据机顶盒 ID 以及所述用户信息分析条件查找用户信息库, 找到符合该机顶盒 ID 条件以及用户信息分析条件的用户信息, 并记录这条数据;

其中, 所述用户信息库包括: 人员信息表、地区表、收视信息表; 所述用户信息分析条件包括: 目标人群、地区。

7、根据权利要求 6 所述的方法, 其特征在于, 所述根据预先设置的分析条件中的分析指标及该分析指标对应的计算公式, 对所述用户信息进行计算, 得到分析指标结果的步骤包括:

根据所述用户信息查找权值表, 得到该用户信息对应的各项权值;

根据所述用户信息、权值以及分析指标对应的计算公式计算分析指标。

8、根据权利要求 7 所述的方法, 其特征在于:

所述分析指标为收视率，所述分析指标对应的计算公式为收视率计算公式：

$$\text{收视率}(\%) = \frac{\sum_{i=1}^n \text{收视时长(分钟)} i \times \text{权重} i}{\text{分析时段总时长(分钟)} \times \text{总体推及人口}} \times 100\%;$$

其中， $\sum_{i=1}^n \text{收视时长(分钟)} i$ 是从所述用户信息中获取，权重  $i$  是从权值表中获取，分析时段总时长、总体推及人口是预先设定的。

9、根据权利要求 1 或 3 所述的方法，其特征在于，所述方法还包括基础数据回传的步骤，该步骤包括：

产生收视行为记录；

判断所述收视行为记录是否满足了回传条件，如果是，则发送所述收视行为记录，继续下一步骤；如果否，则存储所述收视行为记录，返回；

根据所述收视行为记录生成基础数据并储存，该基础数据包括：频道 ID、机顶盒 ID、开始时间、结束时间。

10、根据权利要求 9 所述的方法，其特征在于，所述回传条件为收视行为记录回传要求的最大存储数据量，或者收视行为记录回传要求的最长回传时间间隔。

11、根据权利要求 9 所述的方法，其特征在于，根据收视行为记录生成基础数据的步骤包括：

对所述收视行为记录进行校验及格式转换；

将转换后的收视行为记录按照频道 ID、机顶盒 ID、开始时间以及结束时间的分类转换为基础数据。

12、一种收视率统计系统，该收视率统计系统包括统计服务器，其特征在于，所述统计服务器包括：

检索单元，用于根据预先设置的分析条件的频道 ID、分析时段起始时间 A、分析时段结束时间 B、最小收视时长或最小收视时长百分比 T、是否

连续收视确定收视条件  $A < b \& B > a$  和到达条件  $t \geq T$ , 根据所述收视条件和到达条件依次检索收视信息表中的每一条基础数据, 得到符合所述条件的全部基础数据;

查找单元, 用于根据所述基础数据的机顶盒 ID 以及预先设置的分析条件中的用户信息分析条件查找用户信息库, 得到符合所述条件的用户信息;

指标计算单元, 用于根据所述用户信息以及所述分析指标对应的计算公式计算得到分析指标结果;

其中, 所述每一条基础数据都包括机顶盒 ID、收视时段起始时间  $a$ 、收视时段结束时间  $b$  以及频道 ID; 当  $T$  为最小收视时长时,  $t$  为分析时段内有效收视时长或分析时段内有效收视时长之和; 当  $T$  为最小收视时长百分比时  $t$  为分析时段内有效收视时长或分析时段内有效收视时长之和与分析时段时长的百分比。

13、根据权利要求 12 所述的系统, 其特征在于, 所述统计服务器还包括分析条件设置单元, 用于根据统计需求设置分析条件。

14、根据权利要求 12 所述的系统, 其特征在于, 所述统计服务器还包括:

基础数据导入单元, 用于根据统计要求将基础数据导入到数据库中的收视信息表;

数据库, 用于通过表格的形式存储播出数据、家庭信息、成员信息、机顶盒信息、权值信息、广告价格以及所述基础数据; 以及

操作维护单元, 用于对所述数据库中的数据进行增加、删除、修改的操作。

15、根据权利要求 12 所述的系统, 其特征在于, 所述系统还包括:

机顶盒, 用于在满足回传条件时向数据中心服务器发送收视行为记录;

数据中心服务器, 用于接收所述收视行为记录, 生成基础数据。

16、根据权利要求 15 所述的系统, 其特征在于, 其特征在于, 所述机

顶盒包括：

收视行为采集单元，用于采集用户的收视行为，生成收视行为记录，该收视行为记录包括：正常收视记录、空记录、业务模式记录、待机记录；

存储单元，用于存储回传条件以及依所述记录产生的时间先后顺序依次存储各条收视行为记录；

判断单元，用于判断所述收视行为记录是否到达了所述回传条件；

回传单元，用于根据所述判断单元的判断结果发送收视行为记录到所述数据中心服务器。

17、根据权利要求 15 所述的系统，其特征在于，所述数据中心服务器包括：

收视行为记录接收单元，用于接收机顶盒自动回传的收视行为记录；

数据格式转换及校验单元，用于对所述收视行为记录进行校验及格式转换，生成基础数据；

数据中心，用于存储所述基础数据；

机顶盒参数修改单元，用于对机顶盒的参数进行修改，该参数包括回传条件。

## 收视率统计方法和系统

### 技术领域

本发明涉及数字电视领域的收视率情况调查，尤其涉及一种数字电视的收视率统计方法和系统。

### 背景技术

随着我国数字电视技术的快速发展，人们的收视习惯正在悄悄地发生变化，针对收视情况的收视率统计显得越来越重要，然而原始的收视率统计技术由于是针对模拟电视，因此不能适用于数字电视的收视率统计。

为解决这一问题，目前许多厂家和个人都在积极开发针对数字电视的收视率统计系统及方法，其中较为突出的是申请号为 200610034523.3 的专利申请中提出的一种数字电视系统的收视率数据自动采集、统计系统及方法，该专利申请提供的系统及方法是从终端用户的机顶盒中自动采集收视数据，并实时地通过双向优先电视网络传送到系统服务器，以供其对收视数据进行统计，得出数字电视节目的收视率数据，然而，现有技术中的这种解决方案仍至少存在如下三个缺点：

- 1、由于利用了 IP 网络的双向特性传输收视数据，因此在传输时会存在一定的延时，无法保证收视数据的准确实时传输，导致对于收视率的统计不准确；
- 2、由于收视数据被实时传输到系统服务器进行收视率统计，因此导致系统服务器的工作量过大，为收视统计工作带来不便；
- 3、系统服务器所采用的统计方法工作量大，造成了系统资源的极大浪费。

其中，传统的统计方法为了能从机顶盒实时回传的收视信息表中检索到符合统计要求的数据，并据以计算收视时长，往往需要不断地重复检索收视

信息表，造成了系统资源的极大浪费。

请参照图 1A—图 1D，由于机顶盒回传的基础数据包含了每条收视记录的机顶盒 ID、起始时间  $a$ 、结束时间  $b$  以及频道 ID，因此符合统计要求的分析时间段  $A\sim B$  的基础数据，存在四种不同的时间分布情况，即：1、收视起始时间  $a$  和收视结束时间  $b$  都在统计要求的分析时段  $A\sim B$  以外，也就是用户收视时间包括了分析时段，如图 1A 所示；2、收视起始时间  $a$  和收视结束时间  $b$  都在统计要求的分析时段  $A\sim B$  以内，也就是用户收视时间包含于分析时段，如图 1B 所示；3、收视起始时间  $a$  在分析时段之内，收视结束时间  $b$  在分析时段之外，如图 1C 所示；4、收视起始时间  $a$  在分析时段之外，收视结束时间  $b$  在分析时段之内，如图 1D 所示。此时，为了找到符合所述分析时段的用户，传统情况下，要使用四个判断条件依次搜索扫描四次基础表才能完成，而为了计算收视时长，不同的判断条件要使用不同的计算公式，即，条件 1： $a < A \& b > B$ ，收视时长  $t = B - A$ ；条件 2： $A < a \& B > b$ ，收视时长  $t = b - a$ ；条件 3： $a > A \& a < B \& B < b$ ，收视时长  $t = B - a$ ；条件 4： $a < A \& b < B \& b > A$ ，收视时长  $t = b - A$ 。因此，传统的统计方法不仅效率低下，而且浪费系统资源。

## 发明内容

为了解决上述现有技术中所指出的问题，本发明提供一种数字电视的收视率统计方法和系统。

本发明公开了一种收视率统计方法，所述方法包括下列步骤：根据预先设置的分析条件中的频道 ID、分析时段起始时间  $A$ 、分析时段结束时间  $B$ 、最小收视时长  $T$  以及是否连续收视确定收视条件  $A < b \& B > a$  和到达条件  $t \geq T$ ，根据所述收视条件和到达条件依次检索收视信息表中的每一条基础数据，得到符合所述条件的全部基础数据；根据所述基础数据的机顶盒 ID 以及预先设置的分析条件中的用户信息分析条件查找用户信息库，得到符合所述用户信息分析条件的用户信息；根据预先设置的分析条件中的分析指标及



该分析指标对应的计算公式,对所述用户信息进行计算,得到分析指标结果;其中,所述每一条基础数据包括:机顶盒 ID、收视时段起始时间 a、收视时段结束时间 b 以及频道 ID; t 为分析时段内有效收视时长或分析时段内有效收视时长之和。

本发明还公开了一种收视率统计系统,该收视率统计系统包括统计服务器,所述统计服务器包括:检索单元,用于根据预先设置的分析条件的频道 ID、分析时段起始时间 A、分析时段结束时间 B、最小收视时长或最小收视时长百分比 T、是否连续收视确定收视条件  $A < b \& B > a$  和到达条件  $t \geq T$ ,根据所述收视条件和到达条件依次检索收视信息表中的每一条基础数据,得到符合所述条件的全部基础数据;查找单元,用于根据所述基础数据的机顶盒 ID 以及预先设置的分析条件中的用户信息分析条件查找用户信息库,得到符合所述条件的用户信息;指标计算单元,用于根据所述用户信息以及所述分析指标对应的计算公式计算得到分析指标结果;其中,所述每一条基础数据都包括机顶盒 ID、收视时段起始时间 a、收视时段结束时间 b 以及频道 ID;当 T 为最小收视时长时, t 为分析时段内有效收视时长或分析时段内有效收视时长之和;当 T 为最小收视时长百分比时, t 为分析时段内有效收视时长或分析时段内有效收视时长之和与分析时段时长的百分比。

本发明的有益效果在于:本发明的系统和方法可以提供准确而及时的收视情况分析,数据准确率更高、分析周期短,且可以根据需要生成多种图形结果便于比较。

### 附图说明

此处所说明的附图用来提供对本发明的进一步理解,构成本申请的一部分,并不构成对本发明的限定。在附图中:

图 1A-图 1D 为用户收视起止时段与统计要求的分析起止时段的关系示意图;

图 2 为本发明统计方法流程图;

图3为基础数据生成和回传的方法流程图；

图4为检索基础数据表的方法流程图；

图5为本发明实施例统计系统的组成框图；

图6为本发明实施例的统计系统的系统应用实现示意图；

图7为本发明实施例机顶盒的组成框图；

图8为本发明实施例机顶盒回传收视行为记录的方法流程图；

图9为本发明实施例的数据中心服务器的组成框图；

图10为本发明实施例机顶盒与数据中心服务器的指令交互流程图。

### 具体实施方式

为使本发明实施例的目的、技术方案和优点更加清楚明白，下面结合实施例和附图，对本发明实施例做进一步详细说明。在此，本发明的示意性实施例及其说明用于解释本发明，但并不作为对本发明的限定。

实施例一：

请参照图2，本发明实施例的收视率统计方法主要包括下列步骤：

201：根据预先设置的分析条件的频道ID、分析时段起始时间A、分析时段结束时间B、最小收视时长或最小收视时长百分比T以及是否连续收视，确定收视条件 $A < b \& B > a$ 和到达条件 $t \geq T$ ，根据所述收视条件和到达条件依次检索收视信息表中的每一条基础数据，得到符合所述条件的全部基础数据；

202：根据所述检索得到的基础数据的机顶盒ID以及预先设置的分析条件的用户信息分析条件查找用户信息库，得到符合所述条件的用户信息；

203：根据预先设置的分析条件中的分析指标及所述分析指标对应的计算公式对所述符合条件的用户信息进行分析计算，得到分析指标结果。

其中，步骤202中的根据预先设置的分析条件的用户信息分析条件查找用户信息库，得到符合所述条件的用户信息的步骤可以在步骤201之前，本发明并不限定其先后顺序。

根据本实施例， $t$  为连续收视条件下分析时段内有效收视时长或不连续收视条件下分析时段内有效收视时长之和。连续收视条件是指用户在指定分析时段内连续收看节目达到了最小收视时长（用户收看节目时长不可累加作为该最小收视时长）；不连续收视条件是指用户在指定分析时段内收看节目超过一定时间但少于指定的收视时长，但在指定分析时段内，观众多次收看节目，而多次收看节目的时长之和满足用户指定的收视时长。

根据其他实施例，所述  $T$  可以为最小收视时长百分比，当  $T$  为最小收视时长百分比时，所述  $t$  为分析时段内有效收视时长或分析时段内有效收视时长之和与分析时段时长的百分比。

当  $T$  为预先设置的分析条件中的最小收视时长时， $t$  为连续收视条件下的用户的有效收视时长，则  $t = \frac{(|A-b|+|a-B|-|a-A|-|b-B|)}{2}$ ，或  $t$  为不连续

收视条件下的用户有效收视时长之和，则

$$t = \frac{\sum_{i=1}^n (|A-b_i|+|a_i-B|-|a_i-A|-|b_i-B|)}{2 \times n}$$

，其中， $n$  为某一用户在分析时段内收看某一频道节目的次数。

当  $T$  为预先设置的分析条件的最小收视时长百分比时， $t$  为用户的有效收视时长或用户有效收视时长之和与分析时段时长的百分比，且：

当所述  $t$  为分析时段内有效收视时长与最小收视时长百分比时：

$$t = \frac{(|A-b|+|a-B|-|a-A|-|b-B|)}{2 \times (B-A)} \times 100\% ;$$

当所述  $t$  为分析时段内有效收视时长之和与最小收视时长百分比时：

$$t = \frac{\sum_{i=1}^n (|A-b_i|+|a_i-B|-|a_i-A|-|b_i-B|)}{2 \times n \times (B-A)} \times 100\% , \text{ 其中, } n \text{ 为某一用户在}$$

分析时段内收看某一频道节目的次数。

根据本实施例，基础数据可以通过任何一种可实施的方式获取，例如可

以从机顶盒手动获得，也可以从机顶盒通过双向网络自动实时回传，还可以从机顶盒根据回传条件自动回传。

图3所示为机顶盒根据回传条件自动回传收视行为记录的方法流程图，该方法包括下列步骤：

301：采集收视行为，生成收视行为记录；

302：判断所述收视行为记录是否满足了回传条件，如果是，则发送所述收视行为记录（步骤303），继续下一步骤；如果否，则存储所述收视行为记录（步骤304），返回；

305：根据所述收视行为记录生成基础数据，该基础数据包括：频道ID、机顶盒ID、收视时段起始时间a、分析时段结束时间b；

306：将所述基础数据保存到收视信息表。

其中，回传条件为收视行为记录回传要求的最大存储数据量，或者收视行为记录回传要求的最长回传时间间隔。

其中，根据收视行为记录生成基础数据的步骤包括：

对所述收视行为记录进行校验及格式转换；

将转换后的收视行为记录按照频道ID、机顶盒ID、收视时段开始时间a以及收视时段结束时间b的分类转换为基础数据。

根据本实施例，还需要预先设置分析条件，以便于根据该分析条件过滤基础数据和用户信息，将符合要求的用户提炼出来，进行收视数据的计算和收视率的统计。

其中，分析条件包括了频道ID、分析时段起始时间A、分析时段结束时间B、最小收视时长或最小收视时长百分比T、是否连续收视，这些分析条件与收视条件和到达条件一起用于过滤基础数据。

其中的T可以根据统计要求而设置，例如可以设置为1分钟、3分钟或5分钟。

图4所示为本发明实施例根据预先设置的分析条件和收视条件以及到达

条件来检索收视信息表中的基础数据的方法流程图,在该实施例中,预先设置的 T 为最小收视时长,如图 4 所示,本发明实施例检索收视信息表中的基础数据的方法包括下列步骤:

401: 根据预先设置的分析条件查找第一条基础数据,该分析条件至少包括频道 ID、分析时段起始时间 A、分析时段结束时间 B、最小收视时长 T 以及是否连续收视;

402: 判断所述基础数据中的频道 ID 与所述分析条件中的频道 ID 是否相等,如果是,进行下一步骤,否则查找下一条基础数据(步骤 403);

404: 判断所述基础数据中的收视时段起始时间 a 和收视时段结束时间 b 是否满足收视条件  $A < b \& B > a$ , 如果是,则进行下一步骤,否则查找下一条基础数据(步骤 403);

405: 根据是否连续收视计算有效收视时长 t;

406: 判断所述有效收视时长是否满足到达条件  $t \geq T$ , 如果是,则进行下一步骤,否则查找下一条基础数据(步骤 403);

407: 提取所述基础数据的机顶盒 ID,并保存该机顶盒 ID;

408: 判断是否到达最后一条基础数据,如果是,则结束,否则查找下一条基础数据(步骤 403)。

根据本实施例,并限定上述 402、404 以及 406 的判断步骤的先后顺序,只要根据预先设置的分析条件,将满足该分析条件以及到达条件和收视条件的基础数据从收视信息表中检索出来,都包含于本发明的保护范围。

根据上述检索收视信息表的方法,由于只使用了一个收视条件  $A < b \& B > a$  就包含了如图 1A—图 1D 所示的四种时间分布情况,因此,只需要扫描一次收视信息表即可将满足该收视条件的基础数据筛选出来,以便于进一步筛选以及进行数据的统计工作,大大提高了统计效率和节约了系统资源。又由于采用了公式  $t = \frac{(|A-b|+|a-B|)-|a-A|-|b-B|}{2}$  计算有效收视时

长，也在节约系统资源的同时提高了统计效率。

在图4所示的实施例中，预先设置了最小收视时长 $T$ ，故在判断是否满足到达条件时，先根据是否连续收视选择有效收视时长 $t$ 的计算公式，根据选择出的公式计算有效收视时长或有效收视时长之和，以便比较判断。而根据其他实施例，也可以预先设置最小收视时长百分比 $T$ ，此时，在判断是否满足到达条件时，也要先根据是否连续收视选择计算公式，再根据选择出的公式计算有效收视时长或有效收视时长之和与分析时段的时长的百分比，以便比较判断。具体的计算公式已经在前述进行说明，在此不再赘述。

另外，分析条件还包括目标人群、地区等用户信息分析条件，这些分析条件用于对前述过滤出来的基础数据对应的用户信息进行进一步过滤筛选，以便根据该过滤筛选后的用户信息进行各项分析指标的计算和统计。该方法包括下列步骤：

根据前述方法得到符合条件的基础数据后，根据该基础数据的机顶盒ID以及用户信息分析条件从数据库中的人员信息表、地区表、收视信息表（将其统称为用户信息库）中查找到这些机顶盒ID以及用户信息分析条件对应的用户信息，包括人员信息、目标人群、地区等，然后记录该条符合条件的数据。

根据本实施例，根据预先设置的分析条件查找符合条件的基础数据和根据预先设置的分析条件中的用户信息分析条件查找符合条件的用户信息，可以同时进行，或者不分前后顺序进行，本发明并不以此作为限制。

如此，即可根据预先设置的分析条件中的目标人群、地区等用户信息分析条件，对所述用户信息进行进一步过滤筛选，得到符合条件的用户信息。

另外，为了进一步提高统计效率，本发明该实施例还可以预先按照数据库中的人物属性表的属性计算属性权值，并将该人物属性权值保存到数据库中的权值表，以便于进行指标计算时使用。

此时，在根据用户信息以及分析指标对应的计算公式计算分析指标时可

以添加人物属性权值，该方法包括：

根据上述步骤提取出来的用户信息查找权值表，得到人物属性权值；

根据所述用户信息、人物属性权值以及分析指标对应的计算公式计算分析指标。

其中，分析指标也是预先设置的分析条件之一，由统计人员根据统计要求预先设定好的，以便选择该分析指标对应的计算公式进行指标的计算和统计。

这里需要说明的是，加权是收视数据处理过程中的核心环节。在收视数据处理中，加权的目的在于对样本结构与总体结构的偏差进行校正，以取得能够准确反映总体收视情况的数据。

举例说明，假定某市总人口为 10 万，其中男性占 48%，女性占 52%；所抽样本数为 1000 人，在样本中性别结构与总体性别结构发生了偏离，分别为男性 420 人、女性 580 人，导致男性权值（样本中的一个男性在总体中所代表的人数）与女性权值不同，即，男性权值=总体男性人数/样本中男性人数=48000/420=114.3；女性权值=总体女性人口/样本中女性人口=52000/580=89.6。假定样本中收看某节目人数为 500 人，其中男性 180，女性 320，则：总体收看该节目人数为=样本男性观看人数×权值+样本中女性观看人数×女性权值=180×114.3+320×89.6=49246 人，则该节目收视率=49246/100000=49.3%，而不是 500/1000=50%。

上例中加权变量只有一个（性别），实际中还可能以年龄、城乡等作为权值。

根据上述方法和步骤，将符合预先设置的分析条件的用户信息提取出来以后，即可根据分析指标和分析指标对应的计算公式，计算计算各项指标。

在数字电视领域，分析指标是根据客户需要，或者根据运营商需要来设置和计算的，本发明并不以此作为限制，仅列举两个实施例加以说明本发明的统计方法的应用，而任何根据本发明的上述统计方法实施的分析指标的计

算都应包含于本发明的保护范围之内。

范例一、分析指标为到达率：

$$\text{到达率 (\%)} = \frac{\text{满足连续或不连续到达条件的收视总人口}}{\text{样本总人数}} \times 100\%。$$

其中，连续满足到达条件的收视人数总和可以从数据库中根据检索的结果取得，满足在指定分析时间段内连续收看节目的人数总和，分析条件为用户所选择的目标人群、地区、日期、分析时间段、频道。

其中，不连续满足到达条件的收视人数总和也可以从数据库中根据检索的结果取得，其满足在指定分析时间段内收看节目，且收看节目时间之和满足到达条件，分析条件为用户所选择的目标人群、地区、日期、时间段、频道。

其中，总体推及人口数量也就是样本总体数量。

其中，根据用户所选择的指标计算种类可以从权重表中取出相应权重值。

范例二、分析指标为收视率：

$$\text{收视率 (\%)} = \frac{\sum_{i=1}^n \text{收视时长 (分钟)} i \times \text{权重} i}{\text{分析时段总时长 (分钟)} \times \text{总体推及人口}} \times 100\%。$$

也就是说：

$$\text{收视率 (\%)} = \frac{\text{特定时段内观众收视总时长} \times \text{权重}}{\text{该时段总时长} \times \text{总体推及人口}} \times 100\%。$$

其中，特定时段内观众收视总时长满足条件为：用户所选择的目标人群、地区、日期、时间段、频道。

其中，权重是根据用户所选择的指标计算种类，从权重表中取出相应权重值。

其中，该时段总时长是指总体推及人口的特定时段总时长。

根据本发明的实施例，上述统计方法本发明实施例的统计系统来实现，



具体操作时，可以由统计人员根据下述步骤操作：

设置分析条件并保存；

打开分析条件并开始分析；

根据分析条件中输入的信息，从分析条件中取出设定的目标人群、地区、日期、时间段、频道、到达条件、有效收视条件、分析指标等信息并保存到临时变量中；

是否需要计算时间不连续的数据，如果是，则删除所述临时变量中的到达条件，进行下一步骤，否则直接进行下一步骤；

连接数据库，从地区表、频道信息表、人员信息表、收视信息表、权值表中搜索满足所述临时变量的分析条件的信息；

将搜索到的信息存放于中间变量中；

根据选取的分析指标，使用相应的计算公式和搜索到的中间变量，开始进行分析；

输出分析结果。

根据本发明实施例的上述方法，可以对机顶盒回传的收视记录进行各项指标的计算和统计，不但提高了统计效率，而且节约了系统资源，避免了系统资源的浪费和拥堵。

实施例二：

图5为实现本发明上述实施例的统计方法的数字电视统计系统的组成框图，如图5所示，本发明的数字电视统计系统包括统计服务器，该统计服务器包括：

基础数据导入单元51，用于根据统计要求将基础数据导入到数据库中的收视信息表，这里的基础数据也是指用户的收视数据，一般由机顶盒采集，并可以通过各种方式反馈到本发明的数字电视统计系统，如上述实施例一所述的方式，根据本发明实施例的反馈方式，将在下面对机顶盒说明时进行详细说明。上述统计要求可以是统计时间段的要求，举例说明，要统计6月份

的收视率，则基础数据导入单元 51 将数据中心服务器中的 6 月 1 日至 6 月 30 日的收视行为记录的基础数据导入。另对，由于各种可能的原因导致可能基础数据没有结束时间，对于没有结束时间的基础数据，可以按照预定的程序进行处理，如超过一默认时间，则视为无效处理等。

数据库 52，用于通过表格的形式存储播出数据、家庭信息、成员信息、机顶盒信息、权值信息、广告价格以及所述基础数据信息等，其中，基础数据信息储存于数据库的收视信息表中，权值信息是根据人物属性值预先计算的，储存于数据库的权值表中，其中，播出数据、家庭信息、成员信息、机顶盒信息、广告价格等可以通过操作维护单元 57 录入。

分析条件设置单元 53，用于根据统计需求设置分析条件。该分析条件包括频道 ID、分析时段起始时间 A、分析时段结束时间 B、最小收视时长或最小收视时长百分比 T、是否连续收视、用户信息分析条件、分析指标等。

检索单元 54，用于根据预先设置的分析条件的频道 ID、分析时段起始时间 A、分析时段结束时间 B、最小收视时长或最小收视时长百分比 T、是否连续收视以及收视条件  $A < b \& B > a$  以及或者到达条件  $t \geq T$ ，依次检索收视信息表中的每一条基础数据，得到符合所述条件的全部基础数据中的机顶盒 ID；

查找单元 55，用于根据所述机顶盒 ID 以及预先设置的分析条件中的用户信息分析条件查找用户信息库，得到符合所述条件的用户信息；

分析指标计算单元 56，用于根据所述用户信息以及所述分析指标对应的计算公式计算得到分析指标结果；

其中，所述每一条基础数据都包括机顶盒 ID、收视时段起始时间 a、收视时段结束时间 b 以及频道 ID；当 T 为最小收视时长时，t 为连续收视条件下的有效收视时长或不连续收视条件下的有效收视时长之和；当 T 为最小收视时长百分比时，t 为连续收视条件下的有效收视时长或不连续收视条件下的有效收视时长之和与分析时段时长的百分比。

其中，对于  $t$  的计算公式已于前述说明，在此不再赘述。

操作维护单元 57，用于对所述数据库 52 中的数据进行增加、删除、修改的操作。

根据本实施例，该统计系统还包括机顶盒和数据中心服务器。

下面结合图 6 说明本发明实施例的数字电视的收视率统计系统的组成，如图 6 所示，本发明的数字电视收视率统计系统主要包括机顶盒 61、数据中心服务器 62 以及统计服务器 63，其中：

该统计服务器 63 用于将数据中心服务器 62 储存的基础数据按照要求导入到数据库中的收视信息表，按照预先设置的分析条件、到达条件、有效收视条件等过滤该基础数据，并按照预先设置的分析条件过滤用户信息，得到符合预先设置的分析条件的用户信息，再根据分析指标和该分析指标对应的计算公式对所述筛选出来的用户信息进行各项指标的计算和统计，并生成所需形式的收视率数据输出。这部分已在前述中进行说明，在此不再赘述。

机顶盒 61 用于记录、存储经过筛选后的有效的原始收视数据，并在达到回传条件时，通过数字电视双向网络自动向数据中心服务器 62 发起连接请求，以回传原始收视数据到数据中心服务器 62。

根据本发明的一实施例，机顶盒 61 包括收视行为采集单元 71、存储单元 72、判断单元 73 以及收视行为记录回传单元 74，如图 7 所示，其中：

收视行为采集单元 71 用于采集用户的收视行为数据，生成收视行为记录存储于所述存储单元 72，该收视行为记录包括：正常收视记录、空记录、业务模式记录、待机记录。

上述记录可以采用不同的 ONID、TSID、SERVICEID 组合进行标识，包含了机顶盒在各个时刻的状态变化信息，反映了用户的各种收视行为。机顶盒状态的变化包括当前播出节目的切入/换出，开机/关机，进入/退出数据广播，进入/退出股票信息，进入/退出 NVOD，进入/退出游戏，进入/退出电子邮件等。统计服务器保留上述三个字段组合所对应的节目列表和业务模式

编号，如果接收到的数据记录中三个字段的组合在表中没有定义，即视为无效数据，不作收视指标的计算。

正常收视记录是指用户在正常收看电视节目的情况下，对同一频道节目连续收看达到一定时间长度，构成一次收视行为，产生正常收视记录。该时间长度作为阈值间隔参数可以通过前端，如数据中心服务器设置，并存储于机顶盒中。

空记录是指用户连续换台中每次换台停留时间没有超过预设的阈值间隔，这段时间记作空记录。空记录的 ONID 为 0x0000、TSID 为 0x0000、SERVICEID 为 0x0000。

业务模式记录是指用户在进入机顶盒的各种功能业务时产生的反映收视行为的记录，功能业务包括 EPG 节目指南、VOD/NVOD 收视、数据广播、游戏、股票、电子邮件等等。

上述只是举例说明，由于各地数字电视网络运营商、机顶盒生产厂商的差异性非常大，业务模式记录未提及或不详尽之处要由各方协商解决。同时，业务模式记录可能随着国内数字电视增值业务的快速发展及调查需要的变化进行相应的修正。而本发明的保护范围不限于此，任何根据本发明的精神所作的均等变化，都应包含于本发明的保护范围。

待机记录是由用户遥控关闭机顶盒操作产生，或由系统每隔一段时间记录开机状态的信息推算上次大致的关机时间。

根据上述实施例，本发明对收视事件按类型进行细分，除切入、换出频道的换台信息外，还包括开、关机记录、业务模式记录以及满意度信息、人员信息等，不同类型的事件通过不同的原始网络 ID、传输流 ID 和服务 ID 的组合表示。其中开关机、换台、业务模式等反映了客观的机顶盒工作状态，即用户的收视情况；而满意度、人员等信息可以反映谁在看、评价如何。从而除了得出客观的收视率指标外，更能接受用户反馈，得到用户的主观评价以提供个性化服务等扩展应用。

存储单元 72 则用于存储包含回传条件的预设参数、状态参数，以及依所述记录产生的时间先后顺序依次存储各条收视行为记录。该存储单元 72 可以采用非易失性存储器实现，预留有足够的存储空间，新收视事件，也即收视行为记录产生时则先将其记录在存储器中。

机顶盒 61 的存储单元 72 中存储的预设参数包括：阈值间隔，回传条件，更新关机时间记录的间隔，数据中心服务器的 IP 地址、端口号，连接失败后重新连接的延时等，该预设参数可以根据需求进行扩充和修改。其中，阈值间隔，回传条件，更新关机时间记录的间隔，连接失败后重新连接的延时等参数可以通过数据中心服务器修改，也可以通过机顶盒软件升级更新；而数据中心服务器 IP 地址、端口号由于考虑到安全因素，最好通过机顶盒软件升级更新。

根据本实施例，上述阈值间隔是用于对收视事件进行有效的判断，例如，用户在收看电视的过程中经常会浏览各个频道，在浏览过程中每个频道停留很短时间，此时不应该判断用户对这些频道进行了收看，实时传回这样的数据是无用信息，浪费了网络资源又加重了服务器的负担。本发明中的机顶盒预设了阈值间隔的参数，可以灵活的定义用户在一个频道停留多长时间记为一次有效的收视记录。在机顶盒端做这样的处理可以大大减少无用数据，减少存储空间和节省网络带宽。该参数可以在机顶盒软件升级时更新，也可以在回传的过程中根据通信协议由数据中心服务器对其更新。

根据本实施例，上述回传条件包括达到回传要求的最大存储数据量、达到回传要求的最长回传时间间隔。因为收视率的统计没有必要实时计算，根据业务需要，在一定的时间段内得到收视信息即可，所以根据本发明实施例的机顶盒，采用了先存储收视行为记录，再根据回传条件回传的方式。该回传条件可以通过软件中设置参数来灵活控制，可以在记录数据量累积到一定程度时触发回传机制，也可以在距上次回传时间达到一定间隔长度时触发回传机制。参数可以在机顶盒软件升级时更新，也可以在回传的过程中根据通

信协议由数据中心服务器对其更新。极端情况下，设为每秒或每条记录回传即可达到同背景技术中的专利申请相同的实时回传效果。

相对于实时回传，本发明采用先存储再回传的好处包括：（1）每次产生的收视信息仅为 16 字节，每次回传产生的握手信息以及 IP 包的封装等造成的开销远大于此，大量客户端的情况下服务器不断的处在建立和断开与各个机顶盒的连接过程中，造成网络资源的浪费。在高收视率节目开始或结束等情况下大量用户短时间内产生的换台信息甚至可能造成服务器过载瘫痪；存储再回传的情况下，不同用户达到回传条件的时间分布更为平均，且带宽利用率较高。（2）机顶盒虽然提供了 IP 网络接口，但仍要受到用户家 IP 网络接入方式的限制，实时回传需要随时在线，根据当地网络部署的情况，可能会产生 IP 接入网的费用；非实时的方式可以在用户上网时以较小的带宽代价完成回传，不必随时在线。（3）实时回传在用户快速浏览各个频道时会产生大量的对计算收视率无用的换台信息，本发明的先存储再回传的方案，在存储时即区分了有效数据和无效数据，而仅回传有效数据，则可以改善这一缺陷。

机顶盒的状态参数包括：机顶盒编号、智能卡号、机顶盒 IP 地址、未传输的数据量等。

机顶盒 61 的存储单元 72 还要依记录产生的时间先后顺序依次存储各条记录，如果存储区已满则进行循环覆盖最早的数据。系统记录当前最新写入的位置指针和上次传输结束的位置指针，两者之间的数据为本次回传的待传数据。如果待传数据占满整个存储空间，即当前写入指针与上次传输结束指针重合，新写入的数据仍覆盖最早的数据，这时最早的数据记录将被丢弃，只保留更新的记录。上次传输结束指针随当前写入指针后移，同时存储区满标记置位，下次回传将传整个存储区的数据。

根据本发明的其他实施例，机顶盒 61 还可以包括一个时钟同步单元（图未示），用于与数字电视系统中的时钟同步，以数字电视系统的时间记录收

视事件时间。根据 DVB 规范，数字电视传输流中含有时间表 TDT，机顶盒在正常工作时与系统时钟同步，按 DVB 标准从时间表中提取 5 字节的时间信息，该时间为准确的用户收看事件的时间点，避免了因采用统计服务器接收到收视数据的时间作为收视时间，可能引起的网络延迟、拥塞、丢包、服务器忙等情况产生错误信息甚至丢失信息。

机顶盒启动后与系统时钟同步，记录为开机事件；用户进行换台等操作时记下该频道切入时间，然后判断用户在阈值间隔未换出则判为换台事件，记下切入时间和原始网络 ID、传输流 ID 和服务 ID，否则不记录或记为空记录；用户使用 NVOD，数据广播，股票等增值业务时产生业务模式记录，通过上述三个 ID 的组合表示不同的业务应用。用户在收看时可以通过操作机顶盒菜单产生人员和满意度的评价记录。机顶盒每隔一小段时间更新关机时间值为当前时间，这样机顶盒在断电等突然关机的情况下下次开机提取该时间存为上次的关机记录。在达到一定的预设条件时机顶盒会通过网络回传所存的全部收视记录。传输完毕清空存储区。传输过程中机顶盒向服务器发起连接，报告自身 ID 及存储状态和自身各项参数，服务器应答要求机顶盒回传数据或是更新其参数。完成后断开连接。

判断单元 73 用于判断收视行为记录是否到达了回传条件，关于收视行为记录以及回传条件前述已作说明，在此不再赘述。

如果判断的结果为是，则通过收视行为记录回传单元向数据中心服务器 62 发起建立连接请求，将收视行为记录自动回传到数据中心服务器 62。

如果判断的结果为否，则由机顶盒 61 继续采集收视行为数据，生成收视行为记录保存到存储单元 72。

回传单元 74 用于根据所述判断单元 73 的判断结果发送收视行为记录到所述数据中心服务器 62。对于机顶盒 61 与数据中心服务器 62 的交互流程将在下面再作详细说明。

再请参照图 8，机顶盒根据回传条件自动回传收视行为记录的步骤包括：

801: 采集收视行为数据, 生成收视行为记录并保存;

802: 判断是否达到了回传要求的最长时间间隔, 如果是, 则进行步骤 803, 否则进行步骤 804;

803: 回传所述收视行为记录;

804: 判断是否达到了回传要求的最大存储数据量, 如果是, 则进行步骤 803, 否则进行步骤 801。

数据中心服务器 62 用于接收机顶盒 61 上传的原始收视数据, 也即收视行为记录, 进行校验和格式转换后, 生成基础数据, 准备入库。

根据本实施例, 该数据中心服务器 62 包括收视行为记录接收单元 91、数据格式转换及校验单元 92、数据中心 93 以及机顶盒参数修正单元 94, 如图 9 所示, 其中:

收视行为记录接收单元 91 用于接收机顶盒 61 的回传单元 74 回传的收视行为记录。

数据格式转换及校验单元 92 用于对所述收视行为记录进行校验及格式转换, 生成基础数据存储到数据中心, 具体包括下列步骤:

(1) 校验机顶盒 61 回传的包含收视行为记录的报文信息, 去除掉非正常信息 (包括乱码、错误等信息);

(2) 按照用户预定的需求将所述报文信息按照地区、频道等关键字转换为基础数据;

(3) 转换后的基础数据以文件形式存放到数据中心进行保存, 等待入库。

数据中心 93 用于存储所述基础数据, 该基础数据包括了频道 ID、机顶盒 ID、收视时段起始时间 a、收视时段结束时间 b。

机顶盒参数修改单元 94 用于对机顶盒的参数进行修改, 可修改的参数包括阈值间隔, 回传条件, 更新关机时间记录的间隔, 连接失败后重新连接的延时等。



再请参照图 10, 数据中心服务器 62 与机顶盒 61 的交互流程包括下列步骤:

机顶盒 61 在存储数据量达到最大存储数据量或距上次回传时间间隔达到预设的最长回传时间间隔时, 自动向数据中心服务器 62 发起连接。机顶盒 61 与数据中心服务器 62 连接时, 发送请求连接指令 (步骤 1001), 该请求连接指令包含了自身标识、IP 地址、运行参数、未传输数据量等信息, 其中, 自身标识是指能唯一标识此用户的号码, 可以是智能卡号、MAC 地址或出厂号。

数据中心服务器 62 对机顶盒 61 的身份进行确认, 根据要进行的动作, 发送相应的指令进行应答。

机顶盒 61 接收数据中心服务器 62 发送的指令 (步骤 1002), 判断指令类型 (步骤 1003)。

大多数情况下, 数据中心服务器 62 发送取数据指令 (2), 机顶盒 61 收到后开始发送数据指令 (步骤 1004) 回传, 数据中心服务器 62 接收后, 则发送接收数据确认指令, 机顶盒 61 接收该确认指令 (步骤 1005), 根据上述确认指令的内容判断是否发送数据成功 (步骤 1006), 如果是, 则机顶盒 61 将取数据指针后移所传的字节数 (步骤 1007), 否则继续向数据中心服务器 62 发送数据。

当需要对机顶盒 61 的参数做修改时, 数据中心服务器 62 在连接建立后发送设置参数指令 (4), 机顶盒 61 收到该指令, 修改自身参数 (步骤 1008), 并向数据中心服务器 62 发送设置确认指令 (步骤 1009)。如果成功设置, 则发送指令内容为成功的参数设置确认指令, 数据中心服务器 62 收到成功确认, 即可发取数据指令; 如果设置失败, 返回指令内容为失败的参数设置确认指令, 数据中心服务器 62 收到失败确认, 则重发设置参数指令。

为了满足大量数据的储存要求, 本发明实施例的数据中心服务器可以采取服务器集群负载均衡技术。

本发明定义了一整套利用数字电视机顶盒进行收视率调查以及一些扩

展应用的协议和规范。目前市场上已经有很多具备 IP 网络接入功能的机顶盒。采用本发明的方案完全可以在现有的硬件平台上通过修改机顶盒嵌入软件实现，不需要硬件的改动，一般的机顶盒产品其存储容量都留有一定富余，如果机顶盒内置的存储空间不够可能需要扩展存储芯片。

工作流程简述为：机顶盒启动后与系统时钟同步，记录为开机事件；用户进行换台等操作时记下该频道切入时间，然后判断用户在阈值间隔未换出则判为换台事件，记下切入时间和原始网络 ID、传输流 ID 和服务 ID，否则不记录或记为空记录；用户使用 NVOD，数据广播，股票等增值业务时产生业务模式记录，通过上述三个 ID 的组合表示不同的业务应用。用户在收看时可以通过操作机顶盒菜单产生人员和满意度的评价记录。机顶盒每隔一小段时间更新关机时间值为当前时间，这样机顶盒在断电等突然关机的情况下下次开机提取该时间存为上次的关机记录。在达到一定的预设条件时机顶盒会通过网络回传所存的全部收视记录。传输完毕清空存储区。

传输过程中机顶盒向数据中心服务器发起连接，报告自身 ID 及存储状态和自身各项参数，数据中心服务器应答要求机顶盒回传数据或是更新其参数。完成后断开连接。数据中心服务器对所接收到的数据进行校验和格式转换，准备入库。统计服务器则按照统计要求导入数据中心服务器经过校验和格式转换后的基础数据，根据数据库中存储的播出数据、家庭信息、成员信息、机顶盒信息、广告价格信息等，将所述基础数据转换为收视数据保存。统计服务器的分析平台再根据分析条件，包括分析指标、特定样本、播出时间、地区和频道等对所述收视数据进行指标分析，输出收视率图表。如此即完成了收视率的调查、统计。

以上所述的具体实施例，对本发明的目的、技术方案和有益效果进行了进一步详细说明，所应理解的是，以上所述仅为本发明的具体实施例而已，并不用于限定本发明的保护范围，凡在本发明的精神和原则之内，所做的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。



图 1A



图 1B

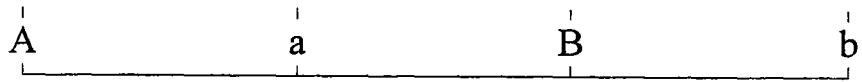


图 1C



图 1D

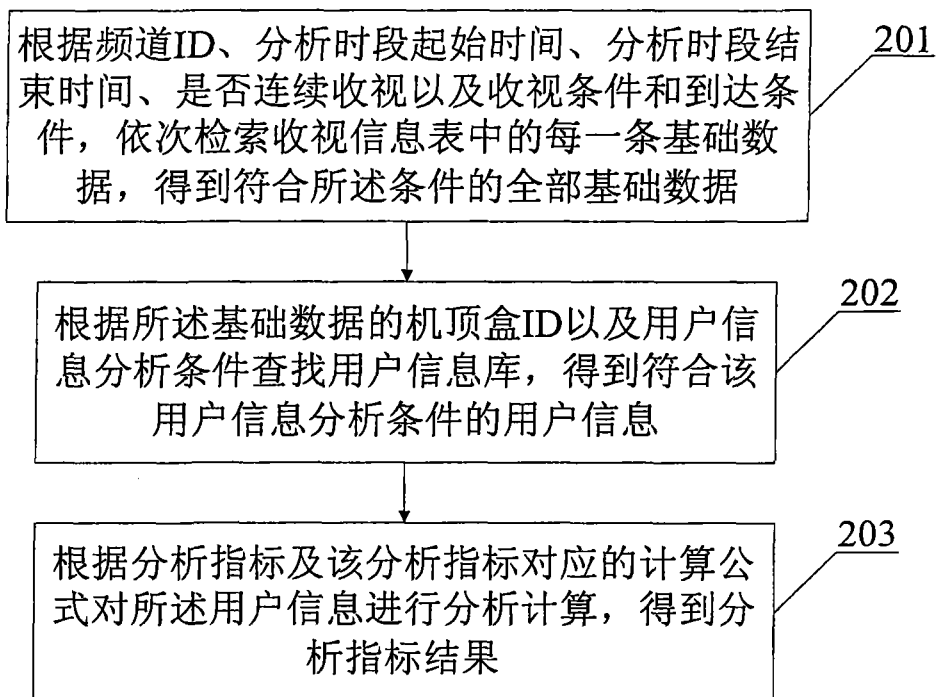


图 2

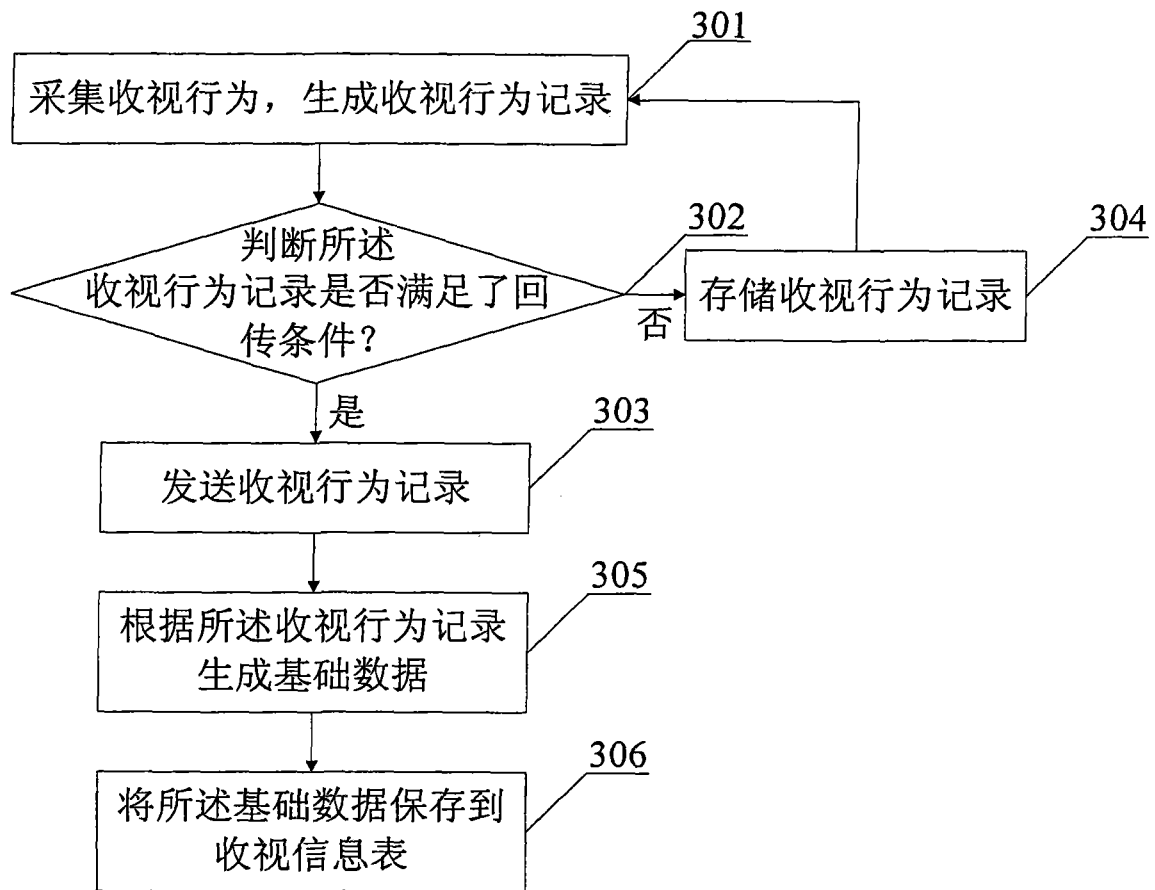


图 3

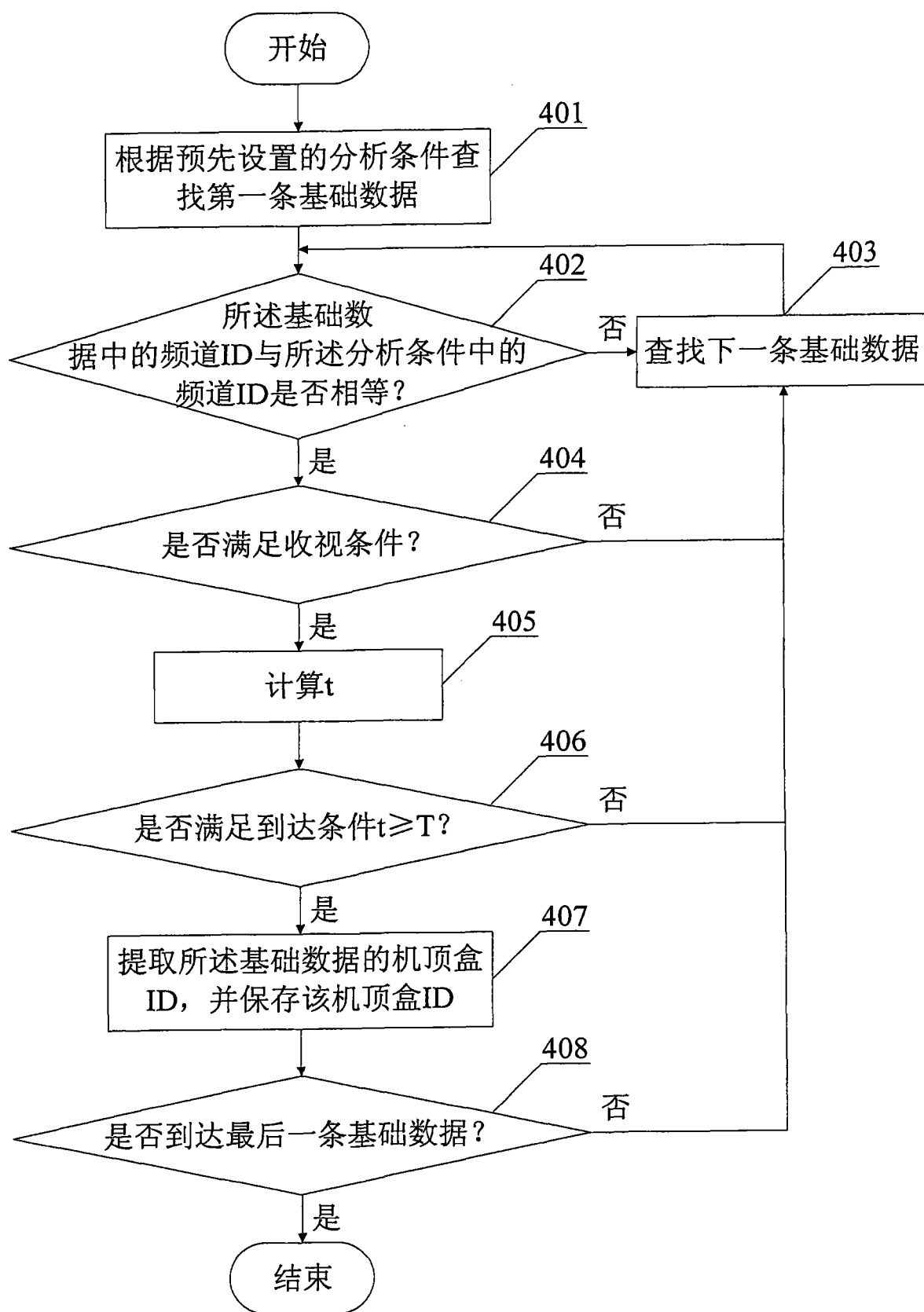


图 4

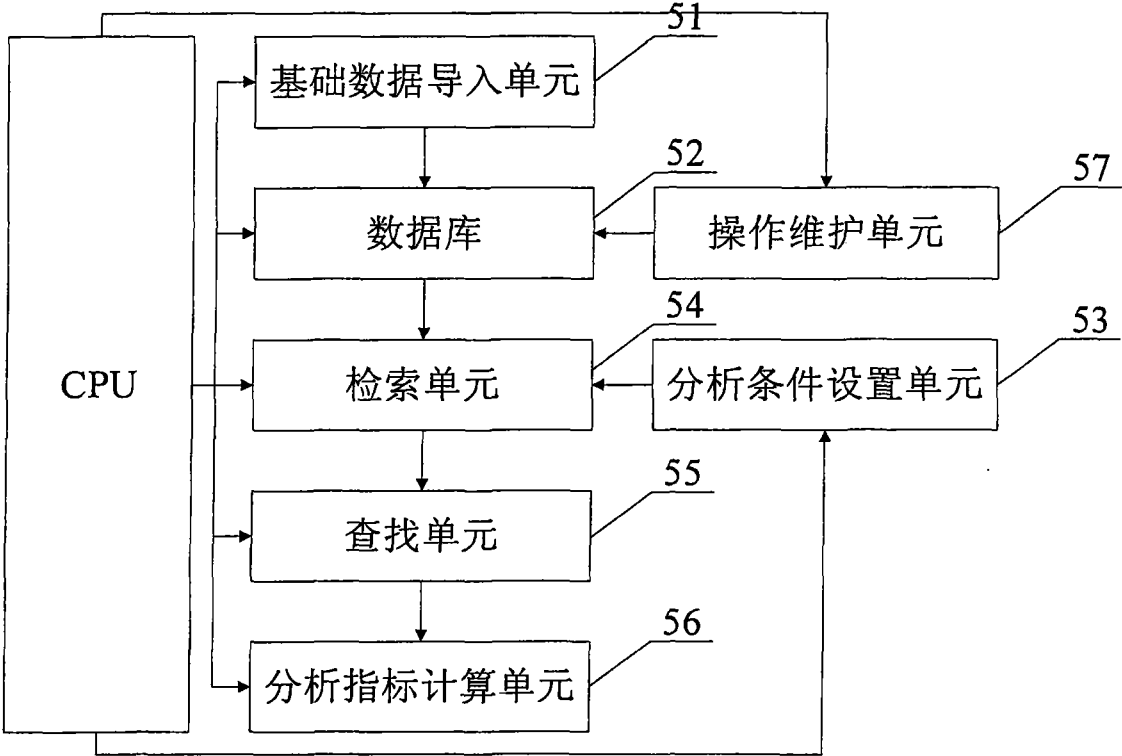


图 5

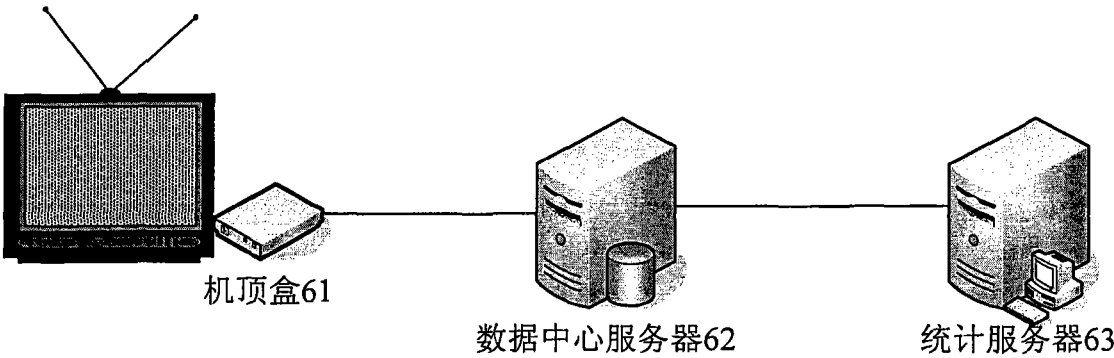


图 6

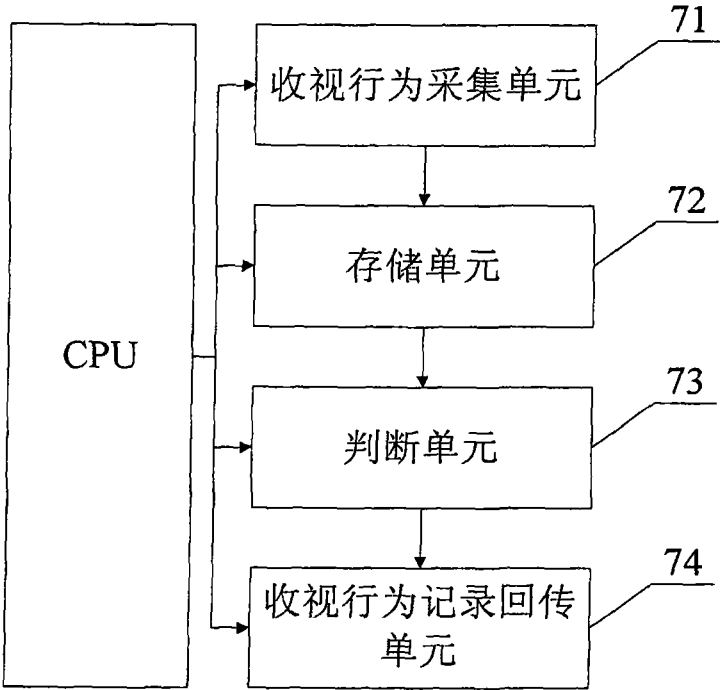


图 7

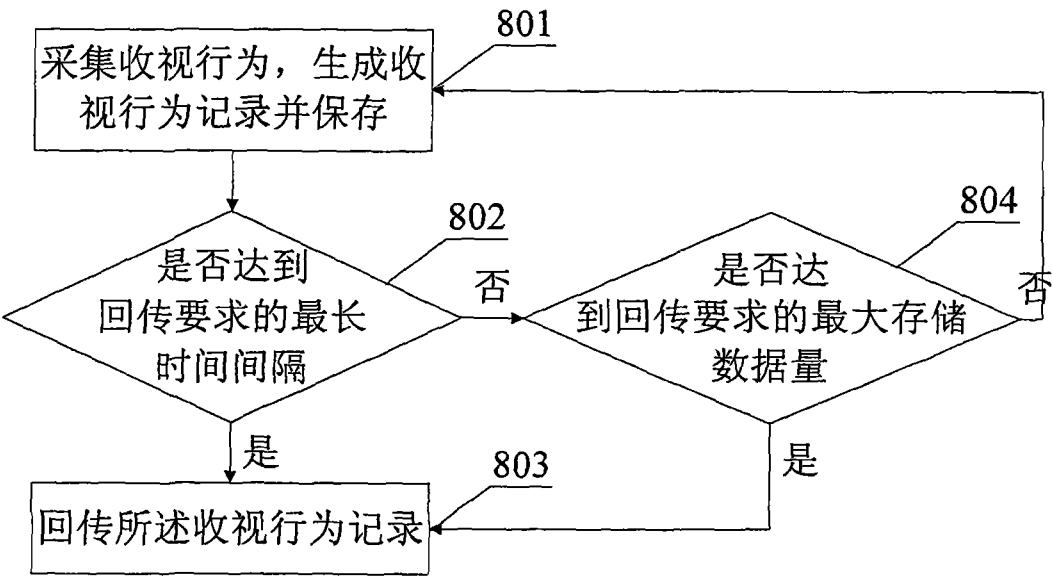


图 8



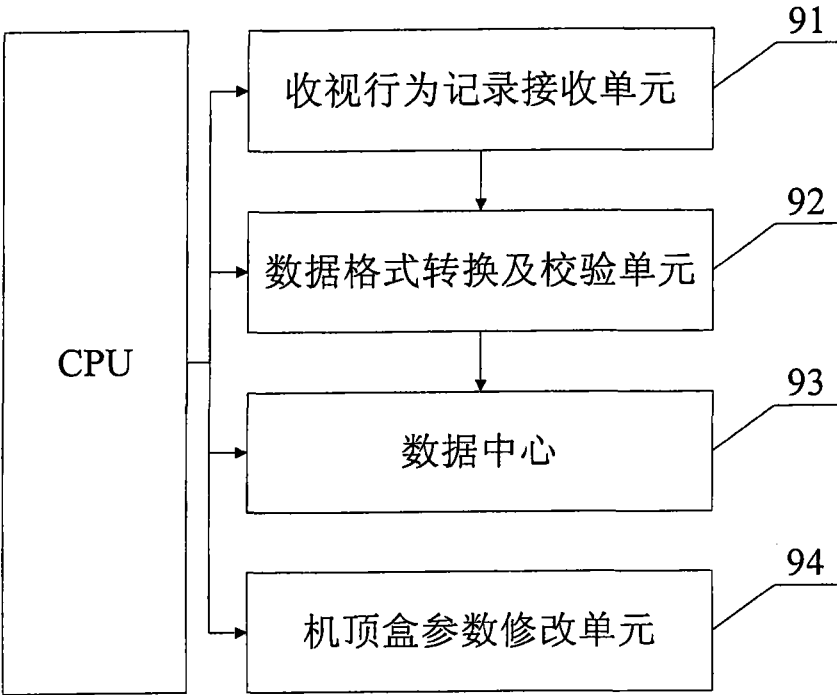


图 9

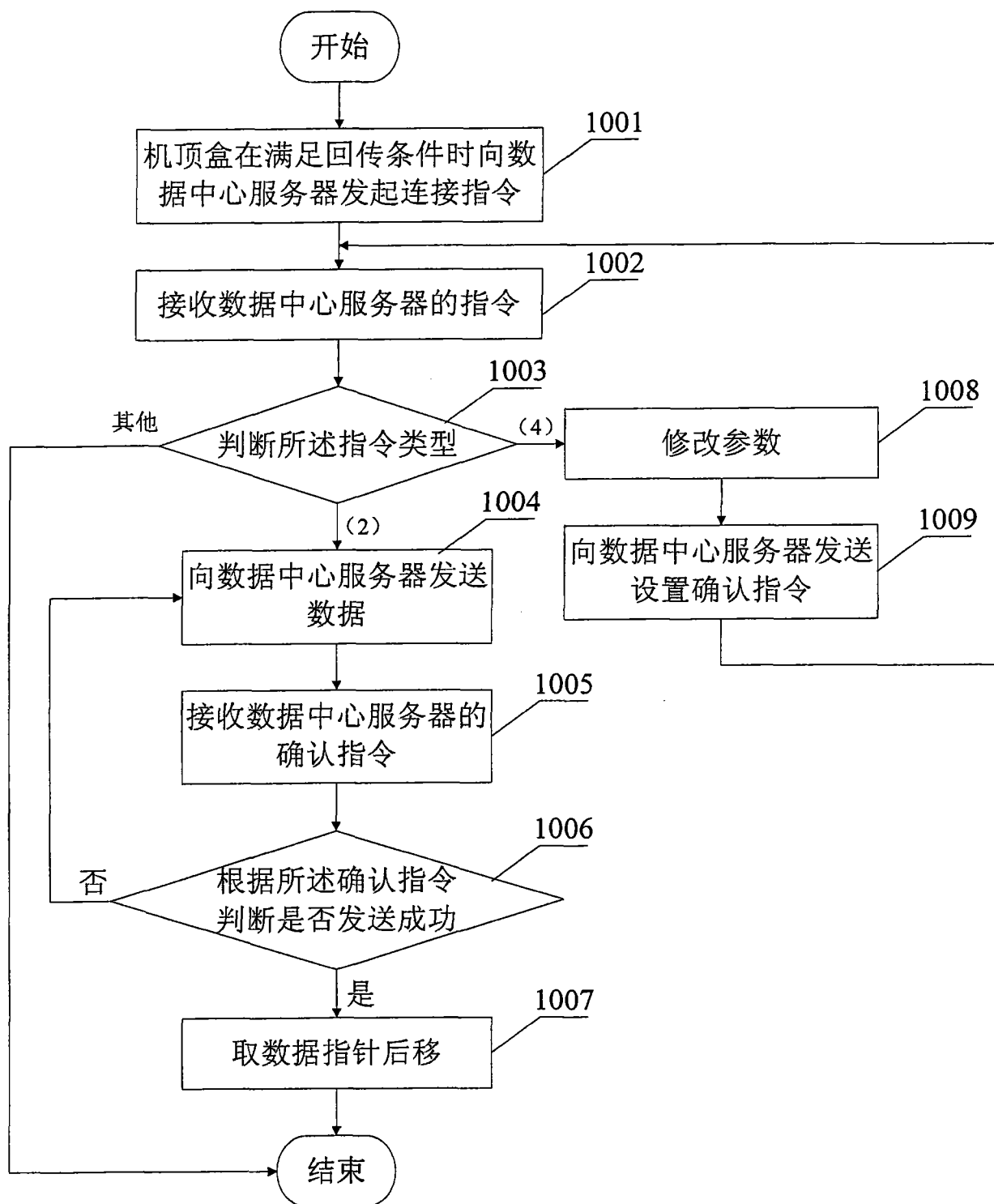


图 10