(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 107305501 A
(43)申请公布日 2017.10.31

(21)申请号 201610262701.1

(22)申请日 2016.04.25

(71)申请人 百度在线网络技术(北京)有限公司
地址 100085 北京市海淀区上地十街10号百度大厦三层

(72)发明人 项连志

(74)专利代理机构 北京品源专利代理有限公司 11332
代理人 孟金喆 胡彬

(51)Int.Cl.
G06F 9/46(2006.01)
G06F 17/30(2006.01)

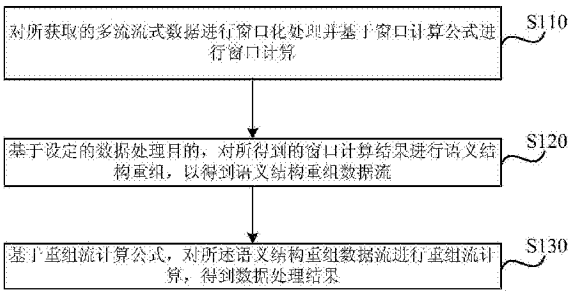
权利要求书3页 说明书14页 附图5页

(54)发明名称

一种多流流式数据的处理方法和系统

(57)摘要

本发明公开了一种多流流式数据的处理方法和系统。该处理方法包括:对所获取的多流流式数据进行窗口化处理并基于窗口计算公式进行窗口计算;基于设定的数据处理目的,对所得到的窗口计算结果进行语义结构重组,以得到语义结构重组数据流;基于重组流计算公式,对所述语义结构重组数据流进行重组流计算,得到数据处理结果。利用该处理方法,不仅解决大数据处理中耗时长,异常处理成本高的问题,还解决了不能对多数据流进行处理计算的问题,进而更加完善了对海量数据流进行处理计算的实现过程,从而满足了用户基于数据处理结果实时探索深层次信息或原因的诉求。



CN 107305501 A

1.一种多流流式数据的处理方法，其特征在于，包括：

对所获取的多流流式数据进行窗口化处理并基于窗口计算公式进行窗口计算；

基于设定的数据处理目的，对所得到的窗口计算结果进行语义结构重组，以得到语义结构重组数据流；

基于重组流计算公式，对所述语义结构重组数据流进行重组流计算，得到数据处理结果。

2.根据权利要求1所述的方法，其特征在于，对所获取的多流流式数据进行窗口化处理并基于窗口计算公式进行窗口计算之前，所述方法还包括：

基于用户输入的指令，确定所述窗口计算公式、所述重组流计算公式和语义结构。

3.根据权利要求2所述的方法，其特征在于，所述方法还包括：

将所述语义结构、所述窗口计算公式和所述重组流计算公式存放至任务配置文件中。

4.根据权利要求1所述的方法，其特征在于，对所获取的多流流式数据进行窗口化处理包括：

基于所获取的多流流式数据的数据属性，将所获取的多流流式数据划分到各个数据计算窗口中；

为所述各个数据计算窗口赋予窗口标记，其中被赋予窗口标记的数据计算窗口被设定为拒绝所述数据流中属于所述数据计算窗口的后续数据流入。

5.根据权利要求1-4任一项所述的方法，其特征在于，基于窗口计算公式进行窗口计算之后，所述方法还包括：

修改第一语义协作元数据的取值，所述第一语义协作元数据的取值用于标识是否触发语义结构重组操作。

6.根据权利要求5所述的方法，其特征在于，基于设定的数据处理目的对所得到的窗口计算结果进行语义结构重组，以得到语义结构重组数据流，包括：

在所述第一语义协作元数据的取值符合语义结构重组操作的触发条件时，以流形式获取所述窗口计算结果；

基于设定的数据处理目的，对所获取的窗口计算结果流进行语义结构重组，以得到重组数据流。

7.根据权利要求6所述的方法，其特征在于，所述方法还包括：

修改第二语义协作元数据的取值，所述第二语义协作元数据的取值用于标识是否触发重组流计算操作，以及

其中，在所述第二语义协作元数据的取值符合重组流计算操作的触发条件时，基于重组流计算公式，对所述语义结构重组数据流进行重组流计算，得到数据处理结果。

8.根据权利要求4所述的方法，其特征在于，在为所述各个数据计算窗口赋予窗口标记之后，所述方法还包括：

收集后续请求流入所述各个数据计算窗口的数据，作为所述各个数据计算窗口的过期数据；

对所述过期数据进行窗口计算，并对窗口计算结果进行语义结构重组得到语义结构重组数据补偿流；

对所述语义结构重组数据补偿流进行重组流计算，得到补偿计算结果；

根据所述补偿计算结果对所述数据处理结果进行修正。

9.一种多流流式数据的处理系统，其特征在于，包括：

窗口化处理模块，用于对所获取的多流流式数据进行窗口化处理；

窗口计算模块，用于基于窗口计算公式对窗口化处理后的多流流式数据进行窗口计算；

数据流重组模块，用于基于设定的数据处理目的，对所得到的窗口计算结果进行语义结构重组，以得到语义结构重组数据流；

重组流计算模块，用于基于重组流计算公式，对所述语义结构重组数据流进行重组流计算，得到数据处理结果。

10.根据权利要求9所述的系统，其特征在于，所述系统还包括：

第一预处理模块，用于在对所获取的多流流式数据进行窗口化处理并基于窗口计算公式进行窗口计算之前，基于用户输入的指令，确定所述窗口计算公式、所述重组流计算公式和语义结构。

11.根据权利要求10所述的系统，其特征在于，所述系统还包括：

第二预处理模块，用于将所述语义结构、所述窗口计算公式和所述重组流计算公式存放至任务配置文件中。

12.根据权利要求9所述的系统，其特征在于，所述窗口化处理模块，具体用于：

基于所获取的多流流式数据的数据属性，将所获取的多流流式数据划分到各个数据计算窗口中；

为所述各个数据计算窗口赋予窗口标记，其中被赋予窗口标记的数据计算窗口被设定为拒绝所述数据流中属于所述数据计算窗口的后续数据流入。

13.根据权利要求9-12任一所述的系统，其特征在于，所述系统还包括：

第一元数据值修改模块，用于在基于窗口计算公式进行窗口计算之后，修改第一语义协作元数据的取值，所述第一语义协作元数据的取值用于标识是否触发语义结构重组操作。

14.根据权利要求13所述的系统，其特征在于，所述数据流重组模块，具体用于：

在所述第一语义协作元数据的取值符合语义结构重组操作的触发条件时，以流形式获取所述窗口计算结果；

基于设定的数据处理目的，对所获取的窗口计算结果流进行语义结构重组，以得到重组数据流。

15.根据权利要求14所述的系统，其特征在于，所述系统还包括：

第二元数据值修改模块，用于修改第二语义协作元数据的取值，所述第二语义协作元数据的取值用于标识是否触发重组流计算操作；

相应的，所述重组流计算模块具体用于：

在所述第二语义协作元数据的取值符合重组流计算操作的触发条件时，基于重组流计算公式，对所述语义结构重组数据流进行重组流计算，得到数据处理结果。

16.根据权利要求12所述的系统，其特征在于，所述系统还包括过期数据补偿处理模块，用于：

在为所述各个数据计算窗口赋予窗口标记之后，收集后续请求流入所述各个数据计算

窗口的数据，作为所述各个数据计算窗口的过期数据；

对所述过期数据进行窗口计算，并对窗口计算结果进行语义结构重组得到语义结构重组数据补偿流；

对所述语义结构重组数据补偿流进行重组流计算，得到补偿计算结果；根据所述补偿计算结果对所述数据处理结果进行修正。

一种多流流式数据的处理方法和系统

技术领域

[0001] 本发明实施例涉及大数据处理技术，尤其涉及一种多流流式数据的处理方法和系统。

背景技术

[0002] 随着时代的进步和经济的发展，人们日常生活中对信息的需求量越来越大，尤其是随着互联网的日益普及，每天都有海量的信息在互联网上发布和传播，对于数据计算和分析的技术人员来说，传统的数据计算分析系统已不能承受对海量数据的计算分析，由此出现了对大规模数据的处理系统。

[0003] 目前，常见的大数据处理技术有两种：一种是将已获取的整体数据集作为输入，然后通过批量分析计算模型进行处理计算，最终输出所需的结果集；另一种是以数据流的形式输入，然后对流形式的数据进行实时的分析计算，并实时获取计算结果。上述两种数据处理技术，尽管可以处理海量数据，但存在一定的不足：针对第一种数据处理技术，存在耗时较长，异常处理成本高等问题，且也不能满足实时性分析计算大数据的要求。针对后一种数据处理方法，尽管能够满足实时性分析，但仅擅长处理单条数据流的计算处理(如对数据流的统计求和计算)，却不能很好地支持多条数据流的分析计算(如，多条数据流之间进行计算，或单条数据流进行环比计算等)。

[0004] 综上所述，现有技术中实时处理多条海量数据流的问题依旧不能解决，无法满足实际需求。

发明内容

[0005] 本发明提供了一种多流流式数据的处理方法和系统，以实现对海量的多流流式数据的实时处理。

[0006] 本发明实施例采用以下技术方案：

[0007] 第一方面，本发明实施例提供了一种多流流式数据的处理方法，该方法包括：

[0008] 对所获取的多流流式数据进行窗口化处理并基于窗口计算公式进行窗口计算；

[0009] 基于设定的数据处理目的，对所得到的窗口计算结果进行语义结构重组，以得到语义结构重组数据流；

[0010] 基于重组流计算公式，对所述语义结构重组数据流进行重组流计算，得到数据处理结果。

[0011] 第二方面，本发明实施例还提供了一种多流流式数据的处理系统，该系统包括：

[0012] 窗口化处理模块，用于对所获取的多流流式数据进行窗口化处理；

[0013] 窗口计算模块，用于基于窗口计算公式对经过窗口化处理后的多流流式数据进行窗口计算；

[0014] 数据流重组模块，用于基于设定的数据处理目的，对所得到的窗口计算结果进行语义结构重组，以得到语义结构重组数据流；

[0015] 重组流计算模块，用于基于重组流计算公式，对所述语义结构重组数据流进行重组流计算，得到数据处理结果。

[0016] 本发明提供了一种多流流式数据的处理方法和系统，该处理方法，首先在获取到数据流后，对数据流中的数据进行窗口化处理，并基于窗口计算公式进行窗口计算；然后对所得的窗口计算结果基于数据处理目的进行语义结构重组形成语义结构重组流；之后对语义结构重组流基于重组流计算公式进行计算得到数据处理结果。利用该处理方法，有效的实现了实时处理涉及海量实体数据的多流流式计算，由此不仅解决大数据处理中耗时长，异常处理成本高的问题，还解决了不能对多数据流进行处理计算的问题，进而更加完善了对海量数据流进行处理计算的实现过程，从而满足了用户基于数据处理结果实时探索深层次信息或原因的诉求。

附图说明

[0017] 图1为本发明实施例一提供的一种多流流式数据的处理方法的流程图；

[0018] 图2为本发明实施例二提供的一种多流流式数据的处理方法的流程图；

[0019] 图3为本发明实施例三提供的一种多流流式数据的处理方法的流程图；

[0020] 图4为本发明实施例四提供的一种多流流式数据的处理方法的流程图；

[0021] 图5a为本发明实施例五提供的一种多流流式数据的处理方法的优选实施例；

[0022] 图5b为本发明实施例五提供的基于多流流式数据处理方法在关键词消费诊断中体现出的应用效果图；

[0023] 图6a为本发明实施例六提供的一种多流流式数据的处理系统的结构框图；

[0024] 图6b为本发明实施例六提供的一种多流流式数据处理的构架图。

具体实施方式

[0025] 下面结合附图并通过具体实施方式来进一步说明本发明的技术方案。可以理解的是，此处所描述的具体实施例仅仅用于解释本发明，而非对本发明的限定。

[0026] 另外还需要说明的是，为了便于描述，附图中仅示出了与本发明相关的部分而非全部内容。在更加详细地讨论示例性实施例之前应当提到的是，一些示例性实施例被描述成作为流程图描绘的处理或方法。虽然流程图将各项操作(或步骤)描述成顺序的处理，但是其中的许多操作可以被并行地、并发地或者同时实施。此外，各项操作的顺序可以被重新安排。当其操作完成时所述处理可以被终止，但是还可以具有未包括在附图中的附加步骤。所述处理可以对应于方法、函数、规程、子例程、子程序等等。

[0027] 实施例一

[0028] 图1为本发明实施例一提供的一种多流流式数据的处理方法的流程图，该方法可适用于对海量的多流流式数据进行数据处理的情况，可以由多流流式数据的处理系统执行，其中该处理系统可由软件和/或硬件实现，并一般可集成在流式数据分析的框架平台上。

[0029] 如图1所示，本发明实施例一提供的一种多流流式数据的处理方法，具体包括：

[0030] S110、对所获取的多流流式数据进行窗口化处理并基于窗口计算公式进行窗口计算。

[0031] 在本实施例中，所述多流流式数据具体可理解为来自不同数据流的、作用于同一个数据处理计算的流式数据。在本实施例中，在对多流流式数据进行处理时，需要有数据流的输入，一般地，不同数据流的输入过程相互独立。此外，本实施例中的数据流一般可看作无序数据流，即所述数据流的输入时序与其业务时序存在不同的情况。

[0032] 所述窗口化处理的目的为将获取的数据流中的数据划分到数据计算窗口中。具体的，所述数据计算窗口可指具有相同数据标记(数据标记主要由数据的一个或多个属性值决定)的数据集合。对于流式数据而言，基于数据自身的属性值就可以确定自身的数据标记，基于确定的数据标记，就可以确定数据自身所属的数据计算窗口，在确定出数据所属的数据计算窗口后，就可以将数据流中数据划分到相应的数据计算窗口。

[0033] 在本实施例中，将数据流中的数据划分到相应的数据计算窗口后，可以基于窗口计算公式对所述数据计算窗口中的数据实时进行窗口计算，并输出窗口计算结果。其中，所述窗口计算公式具体可理解为在数据计算窗口中对所划分的数据进行处理计算时所依据的公式。一般的，常见的窗口计算公式有：对数据的指定属性进行累计求和、对数据的指定属性进行最大值计算、对数据的指定属性进行最小值计算、又或者对数据的指定属性进行的均指计算等。在本实施例例中，所输出的窗口计算结果可以存储在高速缓存中，以便于在后续计算中使用。

[0034] 在本实施例中，对于多流流式数据的数据处理，其过程可看作先对多个数据流分别进行单独处理计算，以便于后续将单独处理计算的结果进行重组，最终基于设定的多流计算公式实现多流流式数据的数据处理。

[0035] 以将数据的时间属性值作为数据标记为例，对所获取的一条数据流进行单流处理操作的过程为：所获取的数据流基于时间属性值划分到不同的数据计算窗口(此时该数据计算窗口可简称为时间窗口)；且可基于时间点进行时间窗口划分，实现数据的窗口化(常见的时间点划分可以是0点～1点、1点至2点，...，23点～0点的划分)；在任一个时间窗口(如1点～2点)中，又可基于设定的窗口计算公式(如累计求和)实时计算数据的某个属性(如消费属性)在该时间窗口内的对应值(如消费累计值)。

[0036] S120、基于设定的数据处理目的，对所得到的窗口计算结果进行语义结构重组，以得到语义结构重组数据流。

[0037] 在本实施例中，S110仅完成了多流流式数据处理中对各数据流进行单独处理的计算，为了实现多数据流的相互计算，需要将多数据流的窗口计算结果以流形式基于特定的条件进行汇合重组，以得到一条基于特定语义结构的数据流。

[0038] 在本实施例中，所述数据处理目的具体可理解为多流流式数据处理的语义结构重组阶段所需要达到的目的。该数据处理目的一般包括语义结构重组操作时所需的处理参数以及基于用户输入指令获得的处理信息，其中，所述处理参数具体可理解为合成语义结构重组流时所需的从实际应用场景的外部环境中动态获取的相关上下文；所述处理信息可以在多流流式数据处理的预处理阶段确定，所述预处理阶段具体可理解为对实际应用场景中用户输入的指令进行分析拆解的过程。

[0039] 在实际的多流流式数据处理过程中，获取到高速缓存中存放的窗口计算结果后，首先确定数据处理目的中的处理信息，然后基于所述处理信息来确定进行语义结构重组所需要的数据流，并在确定出所需要的多条数据流之后，将上述各数据流对应的窗口计算结

果以流形式基于所述数据处理目的中的处理参数进行结构重组，最终得到语义结构重组流。

[0040] S130、基于重组流计算公式，对所述语义结构重组数据流进行重组流计算，得到数据处理结果。

[0041] 在本实施例中，基于S120得到语义结构重组流之后，可以对语义结构重组流进行计算，在进行计算时，需要基于重组流计算公式。一般地所述重组流计算公式在进行多流流式数据处理的预处理阶段提前获得，只需在进行语义结构重组流计算操作时直接调用即可。

[0042] 本实施例的技术方案，首先在获取到数据流后，对数据流中的数据进行窗口化处理，并基于窗口计算公式进行窗口计算；然后对所得的窗口计算结果基于数据处理目的进行语义结构重组形成语义结构重组流；之后对语义结构重组流基于重组流计算公式进行计算得到数据处理结果。利用该处理方法，有效的实现了实时处理涉及海量实体数据的多流流式计算，由此不仅解决大数据处理中耗时长，异常处理成本高的问题，还解决了不能对多条数据流进行处理计算的问题，进而更加完善了对海量数据流进行处理计算的实现过程，从而满足了用户基于数据处理结果实时探索深层次信息或原因的诉求。

[0043] 实施例二

[0044] 图2为本发明实施例二提供的一种多流流式数据的处理方法的流程图，本发明实施例以上述实施例一为基础进行优化，在本实施例中，在“对所获取的多流流式数据进行窗口化处理并基于窗口计算公式进行窗口计算”之前，还优化包括了：基于用户输入的指令，确定所述窗口计算公式、所述重组流计算公式和语义结构。

[0045] 如图2所示，本发明实施例二提供的一种多流流式数据的处理方法，具体包括：

[0046] S210、基于用户输入的指令，确定所述窗口计算公式、所述重组流计算公式和语义结构。

[0047] 一般的，在进行流式数据处理(无论是单流流式数据处理还是多流流式数据处理)之前，首先需要对上述流式数据处理进行预处理，以获得流式数据处理所需的计算公式和语义结构。

[0048] 在本实施例中，可获取用户或技术人员输入的指令，并基于所获取的指令对多流流式数据处理进行预处理。具体的，所述指令可以是代码指令，该预处理过程可以概括为：1)基于获取的代码指令通过语义分析器提取多流流式处理所需的多流计算公式以及语义结构；2)基于公式拆解策略对所述多流计算公式进行公式拆解，得到用于单流计算的所述窗口计算公式和用于多流计算的所述重组流计算公式。

[0049] 在本实施例中，所述多流流式计算公式具体可用于多流流式数据的计算；所述语义结构可以作为所述数据处理目的中的处理信息，以用于得到语义结构重组流。

[0050] 在基于上述步骤1)确定出多流计算公式后，由于多流计算公式所需的数据信息包含在多个数据流中，不能对所述多流计算公式直接处理计算，需要对所述多流计算公式进行拆解，将其拆分成多个计算公式，然后分别对多个计算公式进行处理计算。

[0051] 上述步骤2)实现了对多流计算公式的拆解，其中，所述公式拆解策略具体可理解为具有设定优先计算规则的公式拆解算法。其优先计算规则可理解为：先考虑单条数据流的计算，再考虑多条数据流的计算。基于所述公式拆分策略，可将确定的多流计算公式拆分

成单条数据流计算所需的窗口计算公式，以及多条数据流计算所需的重组流计算公式。上述所得到的窗口计算公式和重组流计算公式均存放至任务配置文件中，以便于后续步骤中使用。

[0052] 进一步的，在确定所述窗口计算公式、所述重组流计算公式和语义结构之后，还可以优选包括：将所述语义结构、所述窗口计算公式和所述重组流计算公式存放至任务配置文件中。

[0053] 在本实施例中，所述任务配置文件具体可理解为存储多流流式数据处理过程中所需要的相关操作数据信息的配置文件。在整个多流流式数据处理过程中可基于任务配置文件中的操作数据信息进行任务调度。

[0054] S220、对所获取的多流流式数据进行窗口化处理并基于窗口计算公式进行窗口计算。

[0055] S230、基于设定的数据处理目的，对所得到的窗口计算结果进行语义结构重组，以得到语义结构重组数据流。

[0056] 在本实施例中，所述数据处理目的具体可以包括语义结构以及处理参数，其中，所述语义结构基于S210的预处理获得，预先存储在任务配置文件中，在进行语义结构重组时可以从任务配置文件中直接调用；所述处理参数可以在在进行语义结构重组时从外部环境中动态获取。具体的，首先基于所述语义结构确定进行语义结构重组所需要的数据流，并在确定所需要的多条数据流之后，将上述各数据流对应的窗口计算结果以流形式基于所述数据处理目的中的处理参数进行结构重组，最终得到语义结构重组流。

[0057] S240、基于重组流计算公式，对所述语义结构重组数据流进行重组流计算，得到数据处理结果。

[0058] 本实施例的技术方案，进一步增加了对多流流式数据处理的预处理操作。利用该处理方法，可以在对多数据流进行处理之前，确定多流流式数据处理所需的计算公式和语义结构，以实现多流计算公式的预处理，进而保证多流流式数据处理的正常进行。

[0059] 实施例三

[0060] 图3为本发明实施例三提供的一种多流流式数据的处理方法的流程图，本发明实施例以上述实施例一为基础进行优化，在本实施例中，将“对所获取的多流流式数据进行窗口化处理”优化为：基于所获取的多流流式数据的数据属性，将所获取的多流流式数据划分到各个数据计算窗口中；为所述各个数据计算窗口赋予窗口标记，其中被赋予窗口标记的数据计算窗口被设定为拒绝所述数据流中属于所述数据计算窗口的后续数据流入。

[0061] 在上述优化的基础上，在“为所述各个数据计算窗口赋予窗口标记”之后，还优化包括了：收集后续请求流入所述各个数据计算窗口的数据，作为所述各个数据计算窗口的过期数据；对所述过期数据进行窗口计算，并对窗口计算结果进行语义结构重组得到语义结构重组数据补偿流；对所述语义结构重组数据补偿流进行重组流计算，得到补偿计算结果；根据所述补偿计算结果对所述数据处理结果进行修正。

[0062] 如图3所示，本发明实施例三提供的一种多流流式数据的处理方法，具体包括：

[0063] S310、基于所获取的多流流式数据的数据属性，将所获取的多流流式数据划分到各个数据计算窗口中。

[0064] 在本实施例中，可以将所述多流流式数据理解为多条数据流的数据，且所获取的

数据流主要为实时输入的无序数据流。所述数据属性具体可指能够表示数据流中的数据特性的信息。示例性的，如数据所具有的日期、时间或者ID号等。

[0065] 所述数据属性一般用来划分数据流中的数据，可将所获取的数据流划分成多个具有不同数据标记的数据集，由此就可以形成多个包含不同数据的数据计算窗口。

[0066] S320、为所述各个数据计算窗口赋予窗口标记，其中被赋予窗口标记的数据计算窗口被设定为拒绝所述数据流中属于所述数据计算窗口的后续数据流入。

[0067] 在本实施例中，可以基于标记策略对所述各个数据计算窗口赋予窗口标记。所述标记策略具体可理解为基于数据流的流入动态设定的当前数据计算窗口计算操作的结束条件。一般地，所述标记策略可基于时间设置，也可基于设定时间内的数据流流量值设置。示例性的，所述标记策略可设置为在T时间点+10分钟，当到达T时间点+10分钟这个时间点后，T时间段窗口就可拒绝接收属于该窗口的数据流入到该窗口中。

[0068] 具体的，由于所获取的数据流为实时输入的无序数据流，所以将数据流中数据划分到相应数据计算窗口也是实时进行的，但是任一数据计算窗口对所流入的数据并不是无条件接收的，其接收时机是有条件限定的。该条件限定就是本实施例所提供的标记策略，即在基于标记策略对任一数据计算窗口进行标记后，该数据计算窗口不再接收数据流中属于该数据计算窗口的后续数据。

[0069] 在本实施例中，进行S320的操作可以保证多流流式数据处理的整体时效性。因为对数据计算窗口中数据的实时处理计算仅是整个多流流式数据处理的一部分操作，后续的语义结构重组操作需要基于窗口计算结果才可开始执行，若一直执行将数据流的数据划分到相应的数据计算窗口的操作，则就需要一直对数据计算窗口中的数据进行实时窗口计算，进而延误整个数据处理的时间。

[0070] S330、基于窗口计算公式进行窗口计算。

[0071] 在本实施例中，可以基于所述窗口计算公式，对所述数据计算窗口中的数据进行实时窗口计算，并实时输出窗口计算结果到高速缓存中。

[0072] 其中，所述窗口计算公式基于上述实施例二中提出的公式拆分策略确定，并预先存储在任务配置文件中。此外，对所述数据计算窗口中数据进行实时窗口计算的窗口计算结果是实时存放到高速缓存的相应位置上的，所述窗口计算结果可作为语义结构重组操作部分的数据流，将所述窗口计算结果存放至高速缓存中可以在后续使用时实现快速调用。

[0073] S340、基于设定的数据处理目的，对所得到的窗口计算结果进行语义结构重组，以得到语义结构重组数据流；

[0074] S350、基于重组流计算公式，对所述语义结构重组数据流进行重组流计算，得到数据处理结果。

[0075] 在本实施例中，上述S340和S350与实施例一中的步骤S120和S130完全相同，这里不再详述。本发明实施例三中的S360～S390体现了本发明实施例的又一个优化，即增加了对已得出的数据处理结果的修正操作。

[0076] S360、收集后续请求流入所述各个数据计算窗口的数据，作为所述各个数据计算窗口的过期数据。

[0077] 在本实施例中，由于对各个数据计算窗口进行窗口计算操作时输入的数据流为无序数据流，所以就会出现数据延迟输入的情况，由此在为所述各个数据计算窗口赋予窗口

标记后仍存在数据流入对应数据计算窗口的情况。由于被赋予窗口标记的数据计算窗口被设定为拒绝所述数据流中属于所述数据计算窗口的后续数据流入，所以在出现上述情况时，无法对上述新流入的数据继续在原来的数据计算窗口中进行增量处理，然而，为了保证最终数据处理结果的正确性，需要对新流入的延迟数据进行特殊处理。

[0078] 本实施例将窗口标记后请求流入数据计算窗口的延迟数据称为过期数据，下述步骤S370～S390体现出了对所述过期数据进行的特殊处理。

[0079] S370、对所述过期数据进行窗口计算，并对窗口计算结果进行语义结构重组得到语义结构重组数据补偿流。

[0080] 在本实施例中，在确定过期数据后，对过期数据再次重新进行窗口计算，并在设定时间内输出窗口计算结果，以保证后续步骤的进行。可以理解的是，窗口标记后，需要实时监控所述数据计算窗口，并对所述数据计算窗口拒绝接收的过期数据进行实时收集和处理，以确保对最终数据处理结果的实时修正。

[0081] 在本实施例中，对所述过期数据的处理过程可相当于对过期数据重新进行本发明实施例一中描述的三个操作步骤。由此，在对过期数据进行窗口计算操作数据窗口计算结果后，仍基于设定的数据处理目的对窗口计算结果进行语义结构重组，以合成语义结构重组数据补偿流。

[0082] S380、对所述语义结构重组数据补偿流进行重组流计算，得到补偿计算结果。

[0083] 在本实施例中，基于所合成的语义结构重组数据补偿流以及相应的重组计算公式，可以进一步实现重组流计算，并达到补偿计算结果。

[0084] S390、根据所述补偿计算结果对所述数据处理结果进行修正。

[0085] 在本实施例中，处理计算得出的补偿计算结果可用于对之前所得出的数据处理结果的结果修正，具体的，所述修正的实现可基于实际情况设定相应的修正公式。

[0086] 此外，在本实施例中，基于补偿计算结果对数据处理结果的修正时机可理解为在得到数据处理结果之后进行。具体的，对于任意两窗口计算结果流合成的语义结构重组流而言，在对该语义结构重组流中的数据处理结果进行修正时，无须等到该语义结构重组流中的所有数据都计算得到数据处理结果后才开始进行修正操作，而是只要该语义结构重组流中存在已得到数据处理结果的数据，且该数据又存在相应的补偿计算结果，就可以将该数据的数据处理结果进行实时修正。由此既保证了多流流式数据处理结果的准确性，还保证了多流流式数据处理的时效性。

[0087] 本实施例的技术方案，提出了对数据处理结果进行修正的优化方案，由此在使得对海量多数据流进行高效处理计算的同时，实现对数据处理结果的修正，进而保证了多流流式数据处理方法的高效性和准确性。

[0088] 实施例四

[0089] 图4为本发明实施例四提供的一种多流流式数据的处理方法的流程图，本发明实施例以上述实施例一为基础进行优化，在本实施例中，在“基于窗口计算公式进行窗口计算”之后，还优化包括了：修改第一语义协作元数据的取值，所述第一语义协作元数据的取值用于标识是否触发语义结构重组操作。

[0090] 进一步的，本实施例将“基于设定的数据处理目的对所得到的窗口计算结果进行语义结构重组，以得到语义结构重组数据流”优化为：在所述第一语义协作元数据的取值符

合语义结构重组操作的触发条件时，以流形式从获取高速缓存中获取所述窗口计算结果；基于设定的数据处理目的，对所获取的窗口计算结果流进行语义结构重组，以得到重组数据流。

[0091] 此外，还优化包括了：修改第二语义协作元数据的取值，所述第二语义协作元数据的取值用于标识是否触发重组流计算操作。进一步的，在所述第二语义协作元数据的取值符合重组流计算操作的触发条件时，基于重组流计算公式，对所述语义结构重组数据流进行重组流计算，得到数据处理结果。

[0092] 如图4所示，本发明实施例四提供的一种多流流式数据的处理方法，具体包括：

[0093] S410、对所获取的多流流式数据进行窗口化处理并基于窗口计算公式进行窗口计算。

[0094] 在本实施例中，上述S410与实施例一中的步骤S110完全相同，这里不再详述。

[0095] S420、修改第一语义协作元数据的取值，所述第一语义协作元数据的取值用于标识是否触发语义结构重组操作。

[0096] 在本实施例中，由于对数据计算窗口中的数据进行的是实时处理计算，且实时处理计算所得窗口计算结果直接存放于所述数据计算窗口在高速缓存中对应的位置，并可作为后续语义结构重组时的输入流。然而，对于语义结构重组的操作，该操作需要相应的触发条件才能进行。

[0097] 在本实施例中，其触发条件可认为是在数据计算窗口被标记后，所设定的第一语义协作元数据的取值。在本实施例中，当数据计算窗口被标记后，不再有数据流入该窗口，相应的窗口计算结果值也不再实时改变，需要对多流流式数据进一步处理，即进行后续的语义结构重组，但是，在进行多流流式数据的语义结构重组之前，需要特定的触发操作来启动多流流式数据的语义结构重组。本实施例可以基于修改第一语义协作元数据的取值来确定是否触发启动语义结构重组。

[0098] 在本实施例中，所述第一语义协作元数据的取值具体可看作一个触发下一处理步骤启动的标记位。示例性的，可允许所述第一语义协作元数据的取值为0或1，当取值为0时，则可认为无法触发下一步骤的启动，当取值为1时，则可认为可以触发下一步骤的启动。

[0099] S430、在所述第一语义协作元数据的取值符合语义结构重组操作的触发条件时，以流形式获取所述窗口计算结果。

[0100] 在本实施例中，当S420中的第一语义协作元数据的取值被修改后，满足语义结构重组的触发条件时，以流形式获取所述窗口计算结果。

[0101] 具体的，所述以流形式获取所述窗口计算结果可概述为：确定进行多流流式数据处理所需的至少两条输入流，所述至少两条输入流具体可看作通过数据计算窗口计算得到的至少两条数据流的窗口计算结果流。

[0102] 需要说明的是，不同数据流的窗口计算结果分别存放在高速缓存的指定位置中，且不同的数据流除了指由不同源程序集在同一时间或不同时间产生的数据集合外，还可以指由一个源程序集在不同日期产生的数据集，示例性的，如同一源程序集在昨天产生的数据集合与今天产生的数据集合就可以属于不同数据流。

[0103] S440、基于设定的数据处理目的，对所获取的窗口计算结果流进行语义结构重组，以得到重组数据流。

[0104] 在本实施例中，所述设定的数据处理目的具体可理解为当前所处理的多流流式数据需要达到的目的。所述对所获取的窗口计算结果流进行语义结构重组具体可理解为：对所述至少两条数据流的窗口计算结果进行语义结构重组。

[0105] 在本实施例中，获取到至少两条数据流的窗口计算结果后，需要基于数据处理目的进行语义结构的重组，由此将至少两条数据流基于窗口计算结果合成一条重组数据流，所述重组数据流可用于计算多流流式数据的最终数据结果。

[0106] S450、修改第二语义协作元数据的取值，所述第二语义协作元数据的取值用于标识是否触发重组流计算操作。

[0107] 在本实施例中，所述第二语义协作元数据与所述第一语义协作元数据的作用相同，都可理解为用于触发下一处理步骤的启动。不同的是，所述第二语义元数据具体可用于触发重组流计算操作。需要说明的是，所述窗口计算的操作、语义结构重组的操作以及重组流计算操作之间是相互关联，相互协作的，且只有在完成前一个操作后，才能触发后一个操作的执行。

[0108] 示例性的，所述第二语义协作元数据的取值同样可看作一个触发标记位，其取值也可优选为0或1，且值为0时，表明不触发下一步操作，值为1时表明可以触发下一步操作。

[0109] S460、在所述第二语义协作元数据的取值符合重组流计算操作的触发条件时，基于重组流计算公式，对所述语义结构重组数据流进行重组流计算，得到数据处理结果。

[0110] 在本实施例中，当第二语义协作元数据的取值符合重组流计算操作的触发条件后，就会启动对语义结构重组数据流的重组流计算操作。具体的，所述重组流计算操作主要基于重组流计算公式，所述重组流计算公式具体为基于公式拆解策略所得。在进行重组流计算后，就可得到多流流式数据处理期望得到的数据处理结果。

[0111] 本实施例的技术方案，具体化了实施例一中的实现步骤，基于本实施例提供的处理方法，既能够实现海量数据的实时处理，还能实现多数据流的处理计算，进而更加完善了对海量数据流进行处理计算的实现过程，从而满足了用户基于数据处理结果实时探索深层次信息或原因的诉求。

[0112] 此外，需要说明的是，本实施例中的S410、S420～S440、以及S450～S460这3个大操作步骤之间是相互触发相互协作的，具体的，各大操作步骤之间可以按照各自间关联的语义协作元数据进行触发，并按照预处理阶段获取的语义结构进行相互间的协作，共同协作完成多流流式数据处理。其中，最初的窗口计算操作的数据流为无条件运行流，只要有数据流输入，就会持续运行，实时计算窗口计算结果，且窗口计算结果会实时存放至高速缓存中；语义结构重组流基于第一语义协作元数据的取值触发并启动，之后重组流计算操作基于第二语义协作元数据的取值触发并启动，最终得到所需的数据处理结果。

[0113] 实施例五

[0114] 图5a为本发明实施例五提供的一种多流流式数据的处理方法的优选实施例，本实施例五的应用场景是对百度罗盘产品中的关键词消费诊断功能，该应用场景所需要的数据是网民基于百度推广显示的关键词进行点击所产生的点击消费数据，该数据被称为凤巢点击消费数据。对于凤巢点击消费数据，每条数据所包含的信息内容可概括为：关键词、点击日期、点击时间以及点击产生的消费金额等。

[0115] 对于应用场景中所提的关键词消费诊断，基于本发明实施例五所提供的处理方

法，可以确定其进行数据处理的目的是：获得所有关键词在今日整点T累计消费环比昨日整点T累计消费的突降值，以使用户基于得到的数据处理结果可以实时查询自身账户下所有关键词在整点T累计消费情况以及环比情况。

[0116] 需要说明的是，本实施例五提供的优选实施例，最初的输入流为凤巢点击消费数据流，该数据流的特点是：1)数据量大和2)实时性强。作为全球最大的中文搜索引擎，每时每刻都有网民在百度搜索引擎中基于关键词搜索，因此，时刻就有用户基于搜索的关键词点击百度推广中所显示的广告，用户在基于关键词点击时，后台就产生了点击消费数据。由此可以看出作为数据流的凤巢点击数据的海量的实时的数据流。此外，本实施例所提供的优选实施例还体现了对多数据流处理的特点，即，对于环比计算，需要基于今天和昨天的数据信息，因此符合多数据流处理的范畴。

[0117] 如图5a所示，本发明实施例五提供的一种优选的多流流式数据的处理方法，具体包括：

[0118] S510、基于公式拆解策略拆解多流计算的计算公式。

[0119] 在本实施例中，所述多流计算公式的拆解过程可表述为：

[0120] 1)基于获取的关键词整点T累计消费环比突降值这个输入指令，确定用于多流流式数据处理的多流计算公式和语义结构；2)对所述环比突降值的计算公式进行公式拆解，以得到窗口计算公式和重组流计算公式。

[0121] 示例性的，由于本实施例进行的是多流流式数据处理，需要在对多流流式数据进行处理之前，先基于获取的输入指令确定用于多流流式数据处理的多流计算公式和语义结构，在本实施例中，可以确定其语义结构是计算关键词的今日整点T累计消费值和昨日整点T累计消费值。之后还可基于输入指令确定其多流计算公式是：计算关键词的整点T累计消费值，然后计算关键词的今日整点T累计消费值与昨日整点T累计消费值的差。

[0122] 分析上述确定的多流计算公式，可知上述计算公式需要分步计算才能实现，因此基于公式拆分策略将上述多流计算公式拆分为用于单数据流计算的窗口计算公式和用于多数据流计算的重组流计算公式。示例性的，上述公式拆分的窗口计算公式为：对整点T时间段流入的关键词进行消费值的累计求和；其拆分出的重组流计算公式为：计算关键词今日整点T时间段的累计消费值与关键词昨日整点T时间段的累计消费值。需要说明的是，本实施例中所提的关键词消费值的计算具体指同一关键词消费值的计算。此外，基于输入指令确定的语义结构和基于公式拆分策略拆分出的计算公式均对应存储到任务配置文件中，当基于处理方法进行数据处理时可以在进行相应的操作步骤时快速调用该操作步骤所需的计算公式。

[0123] S520、按小时级时间属性划分凤巢点击消费数据流中的数据到小时级窗口。

[0124] 在本实施例中，在对基于输入指令确定多流计算公式进行公式拆分操作后，可以将凤巢点击消费数据流作为输入流进行获取，并在获取到凤巢点击消费数据流之后按小时级的时间属性对数据流的数据进行数据计算窗口划分。

[0125] 示例性的，所述小时级时间属性具体可指一天划分为24个时间段，每个时间段为一个小时级窗口，然后基于凤巢点击数据的点击时间确定将数据划分到哪个小时级窗口。例如，当前获取的关键词点击时间为0点24份，则可将该关键词及其数据信息划分到0点这个小时级窗口中。

[0126] S530、对小时级窗口中的数据实时进行关键词整点累计消费值计算，并输出计算结果到高速缓存中。

[0127] 示例性的，在进行数据计算窗口划分后，需要基于窗口计算公式对数据计算窗口中的数据进行实时计算，并将窗口计算结果实时输出到高速缓存中。例如，在0点这个小时级窗口中，对划分至这个窗口的关键词进行消费值的累计计算，对于关键词A，在0点1分，0点4分，以及0点8分产生了点击消费金额分别为1元，1元，和1元，则截止到0点9分，关键词A的消费值累计计算值为3元，此时高速缓存中存放的计算结果就为3元。对于其他关键词，其他小时级窗口的计算过程与上述描述相同，这里不再详述。

[0128] S540、判断是否对小时级窗口进行标记，若是，则分别执行S550和S551；若否，则返回执行S530。

[0129] 在本实施例中，为了节省多流流式数据处理时间，可以设定冻结小时级窗口的冻结时机(所述冻结可理解为拒绝数据流中属于该窗口的后续数据的流入请求)，而该冻结时机的设定可以基于窗口标记进行。示例性的，由于是实时的数据计算，对0点小时级窗口来说，输入的数据流划分到该小时级窗口的时间可以大致确定为0点到1点之间，在过了1点之后，所获取的绝大部分数据流将会基于时间属性划分到1点小时级窗口中。由此，可以通过设定标记策略对0点小时级窗口进行窗口标记，以标识可以冻结该小时级窗口。

[0130] 一般地，在设定标记策略时通常基于本实施例的处理方法所处的应用场景设定，例如，针对本实施例中提到的关键词消费诊断这个应用场景，可以将标记策略设置为定时标记，例如，设置标记策略为：在1点+5分钟时，即1点05分时对0点小时级窗口进行窗口标记，由此冻结该小时级窗口。

[0131] 在本实施例中，判断小时级窗口是否基于标记策略的进行了窗口标记，如果没有进行窗口标记，则需要返回S530继续进行实时增量计算，如果符合标记策略进行了窗口标记，则可分别执行两个分支步骤，这两个分支步骤可概括为基于S551进行过期数据的补偿处理计算，以及基于S550～S580进行的实时数据的多流处理计算。

[0132] S551、将窗口标记后请求流入小时级窗口的数据确定为过期数据，对过期数据进行实时分形补偿计算，得到实施补偿计算结果，之后执行S590。

[0133] 在本实施例中，在基于标记策略对小时级窗口进行标记后，由于存在数据延迟的情况，所以还会有属于所述小时级窗口的数据没有被及时划分到其中，进而在窗口标记后又无法流入到该小时级窗口中，可称这些数据为过期数据。然而，由于这些过期数据不能在原来的小时级窗口中继续进行进行窗口计算，但为了保证多流流式数据处理结果的准确性，也不能直接舍弃这些过期数据，因此需要设置特定的步骤对这些过期数据进行分形补偿处理。

[0134] 示例性的，对过期数据进行实时分形补偿计算的过程可概括为：获取过期数据，对过期数据重新基于窗口计算公式进行实时增量计算，并得到实时增量计算的窗口计算结果；然后，对过期数据流的今日关键词整点累计消费值以及上述过期数据流对应的昨日关键词整点累计消费值进行基于数据处理目的进行语义结构重组，合成语义结构重组补偿流；最后基于重组计算公式对所述语义结构重组补偿流进行重组流计算，并最终得到补偿计算结果。在本实施例中，补偿计算结果的获得过程是实时进行的，只要有过期数据出现，就进行上述操作。

[0135] S550、获取高速缓存中的今日关键词整点消费累计值数据流和昨日关键词整点消费累计值数据流。

[0136] S560、基于数据处理目的，将获取的两条数据流进行语义结构重组，合成语义结构重组流。

[0137] S570、基于拆分出的环比计算公式，对语义结构重组流进行重组流计算，获得数据处理结果。

[0138] 在本实施例中，S550～S570是小时级窗口标记后需要执行的另一个分支步骤，首先，在基于S530确定小时级窗口进行标记后，需要修改与语义结构重组相关联的第一语义协作元数据的取值，以基于第一语义协作元数据的取值确定是否触发进行语义结构重组的操作，并在触发启动执行语义结构重组后，再获取高速缓存中存放的今日关键词整点消费累计值数据流和昨日关键词整点消费累计值数据流，基于应用场景中的数据处理目的，将获取的两条数据流进行语义结构重组，获得合成后的语义结构重组流；然后，在合成语义结构重组流之后，还需要修改与重组流计算操作相关联的第二语义协作元数据的取值，以基于第二语义协作元数据的取值确定是否触发进行重组流计算的操作，并在触发启动执行重组流计算之后，对合成的语义结构重组流进行重组流计算。示例性的，重组流计算就是进行今日关键词整点消费累计值与昨日关键词整点消费累计值的差计算，计算出的差值就可看作该应用场景下进行多流流式数据处理的数据处理结果。

[0139] 需要注意的是，本实施例中提出的第一语义协作元数据和第二语义协作元数据并没有具体体现在本实施例五的流程图中，而是以语义协作元数据与相应操作步骤之间的数据信息的交流来体现，但所起的作用和要表达的意思并不因文字表述的改变而改变。

[0140] S580、在数据处理结果输出后，基于实时补偿计算结果修正数据处理结果，并实时输出修正后处理结果。

[0141] 在本实施例中，基于窗口计算操作、语义结构重组流的合成操作以及重组流计算操作进行多流流式数据处理之后，就可以数据初始的数据处理结果。所述数据处理结果可看作该应用场景下的多流流式数据的处理结果，然而，为了保证数据处理结果的准确性，需要基于实时计算得出的补偿计算结果进行数据处理结果修正，并最终输出修正后的处理结果。需要说明的是，对数据处理结果的修正操作，也需要基于与结果修正操作关联的语义协作元数据的触发，示例性的，其触发过程可参考图5a中语义协作元数据与重组计算操作之间的数据信息交流来完成。

[0142] 在关键词消费诊断这个应用场景中，基于上述提供的多流流式数据处理方法，最终输出的结果是各关键词在整点T的累计消费值，以及各关键词在整点T的累计消费值的环比突降值。

[0143] 图5b为本发明实施例五基于多流流式数据处理方法在关键词消费诊断中体现出的应用效果图。如图5b所示，其显示结果为任一用户ID中任一指定关键词在整点T时间段内的数据累计消费值的曲线分布，以及该关键词在整点T时间段累计消费值的环比突降值的曲线分布，其中，T∈[0点，23点]。需要说明的是，图5b中所显示的信息内容仅可看作在某个时间点对某一关键词的数据处理结果的截取片段，在实际应用中，若该关键词存在补偿计算结果，则在不同时刻，该关键词所得到的数据处理结果会基于补偿计算结果的修正而产生变化。

[0144] 本实施例的技术方案，提供了一种多流流式数据的处理方法的优选实施例，基于该优选实施例，更清楚了多流流式数据的处理方法的操作步骤，基于该处理方法，高效的实现了海量数据的实时处理计算，节省了数据的收集时间，加快了数据处理进度，且通过设定补偿处理计算，进一步保证多流计算结果的准确性。

[0145] 实施例六

[0146] 图6a为本发明实施例六提供的一种多流流式数据的处理系统的结构框图，本实施例的处理系统可由软件和/或硬件实现，可适用于对海量的多流流式数据进行数据处理的情况，并一般可集成在流式数据分析的框架平台上。如图6a所示，该处理系统包括：窗口化处理模块60、窗口计算模块61、数据流重组模块62以及重组流计算模块63。

[0147] 其中，窗口化处理模块60用于对所获取的多流流式数据进行窗口化处理。

[0148] 窗口计算模块61，用于基于窗口计算公式对经过窗口化处理后的多流流式数据进行窗口计算。

[0149] 数据流重组模块62用于基于设定的数据处理目的，对所得到的窗口计算结果进行语义结构重组，以得到语义结构重组数据流。

[0150] 重组流计算模块63用于基于重组流计算公式，对所述语义结构重组数据流进行重组流计算，得到数据处理结果。

[0151] 在本实施例中，该处理系统首先通过窗口化处理模块60将对所获取的多流流式数据进行窗口化处理，窗口计算模块61基于窗口计算公式进行窗口计算；然后，通过数据流重组模块62基于设定的数据处理目的，对所得到的窗口计算结果进行语义结构重组，以得到语义结构重组数据流；最终，通过重组流计算模块63基于重组流计算公式，对所述语义结构重组数据流进行重组流计算，得到数据处理结果。

[0152] 本实施例的技术方案，有效的实现了实时处理涉及海量实体的多流流式计算，由此不仅解决大数据处理中耗时长，异常处理成本高的问题，还解决了不能对多数据流进行处理计算的问题，进而更加完善了对海量数据流进行处理计算的实现过程，从而满足了用户基于数据处理结果实时探索深层次信息或原因的诉求。

[0153] 进一步的，所述处理系统还可以包括：

[0154] 第一预处理模块，用于在对所获取的多流流式数据进行窗口化处理并基于窗口计算公式进行窗口计算之前，基于用户输入的指令，确定所述窗口计算公式、所述重组流计算公式和语义结构。

[0155] 进一步的，所述处理系统还可以包括：

[0156] 第二预处理模块，用于将所述语义结构、所述窗口计算公式和所述重组流计算公式存放至任务配置文件中。

[0157] 在上述实施例的技术上，所述窗口化处理模块60具体可以用于：

[0158] 基于所获取的多流流式数据的数据属性，将所获取的多流流式数据划分到各个数据计算窗口中；为所述各个数据计算窗口赋予窗口标记，其中被赋予窗口标记的数据计算窗口被设定为拒绝所述数据流中属于所述数据计算窗口的后续数据流入。

[0159] 进一步的，所述处理系统还可以包括：

[0160] 第一元数据值修改模块，用于在基于窗口计算公式进行窗口计算之后，修改第一语义协作元数据的取值，所述第一语义协作元数据的取值用于标识是否触发语义结构重组

操作。

[0161] 在上述实施例的基础上，所述数据流重组模块62具体可以用于：

[0162] 在所述第一语义协作元数据的取值符合语义结构重组操作的触发条件时，以流形式获取所述窗口计算结果；基于设定的数据处理目的，对所获取的窗口计算结果流进行语义结构重组，以得到重组数据流。

[0163] 进一步的，所述处理系统还可以包括：

[0164] 第二元数据值修改模块，用于修改第二语义协作元数据的取值，所述第二语义协作元数据的取值用于标识是否触发重组流计算操作。

[0165] 相应的，所述重组流计算模块63具体用于：

[0166] 在所述第二语义协作元数据的取值符合重组流计算操作的触发条件时，基于重组流计算公式，对所述语义结构重组数据流进行重组流计算，得到数据处理结果。

[0167] 进一步的，所述处理系统还可以包括过期数据补偿处理模块，用于：在为所述各个数据计算窗口赋予窗口标记之后，收集后续请求流入所述各个数据计算窗口的数据，作为所述各个数据计算窗口的过期数据；对所述过期数据进行窗口计算，并对窗口计算结果进行语义结构重组得到语义结构重组数据补偿流；对所述语义结构重组数据补偿流进行重组流计算，得到补偿计算结果；根据所述补偿计算结果对所述数据处理结果进行修正。

[0168] 此外，图6b为本发明实施例七提供的一种多流流式数据处理的构架图，图6b所示的构架图，将本发明实施例提供的多流流式数据处理系统的工作流程及其之间的协作关系更加清晰化。如图6b所示，该构架图具体可分为4条相互关联的语义流，分别是：窗口统计分析流64、语义结构重组流65、语义结构计算流66、以及分形补偿流67。上述4条语义流，窗口统计分析流64对应多流流式数据处理系统的窗口化处理模块60和窗口计算模块61，其余3条语义流分别对应多流流式数据处理系统的数据流重组模块62、重组流计算模块63以及过期数据补偿处理模块。图6b可以看出，各语义流在相互协作的基础上又独立完成各自对应模块的功能步骤，最终实现对多流流式数据的处理。

[0169] 注意，上述仅为本发明的较佳实施例及所运用技术原理。本领域技术人员会理解，本发明不限于这里所述的特定实施例，对本领域技术人员来说能够进行各种明显的变化、重新调整和替代而不会脱离本发明的保护范围。因此，虽然通过以上实施例对本发明进行了较为详细的说明，但是本发明不仅仅限于以上实施例，在不脱离本发明构思的情况下，还可以包括更多其他等效实施例，而本发明的范围由所附的权利要求范围决定。

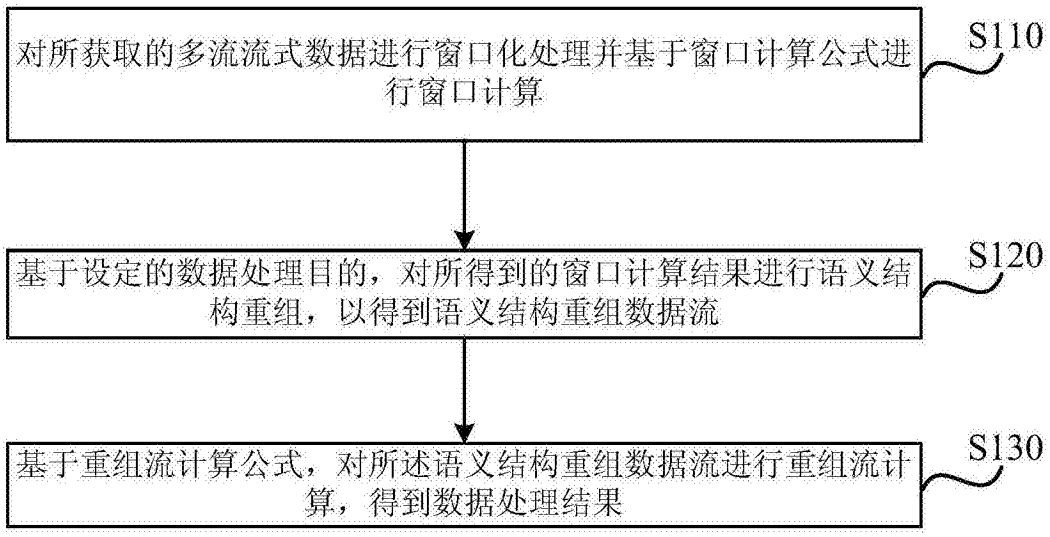

对所获取的多流流式数据进行窗口化处理并基于窗口计算公式进行窗口计算
S110
基于设定的数据处理目的，对所得到的窗口计算结果进行语义结构重组，以得到语义结构重组数据流
S120
基于重组流计算公式，对所述语义结构重组数据流进行重组流计算，得到数据处理结果
S130

图1

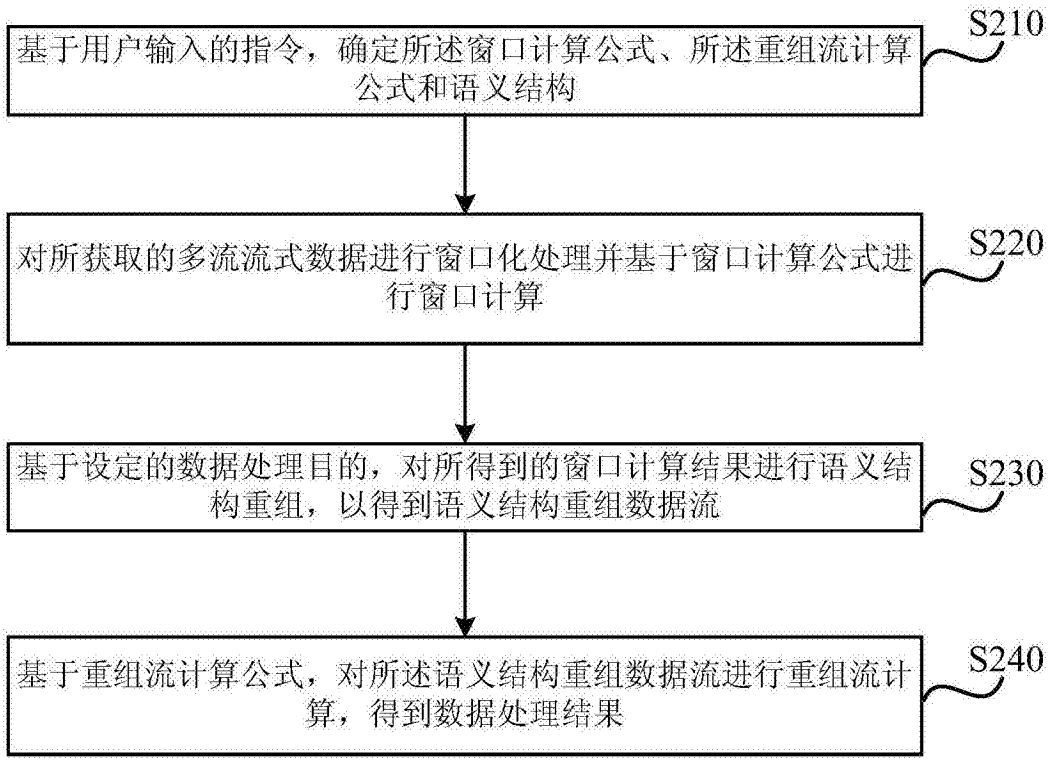

基于用户输入的指令，确定所述窗口计算公式、所述重组流计算公式和语义结构
S210
对所获取的多流流式数据进行窗口化处理并基于窗口计算公式进行窗口计算
S220
基于设定的数据处理目的，对所得到的窗口计算结果进行语义结构重组，以得到语义结构重组数据流
S230
基于重组流计算公式，对所述语义结构重组数据流进行重组流计算，得到数据处理结果
S240

图2

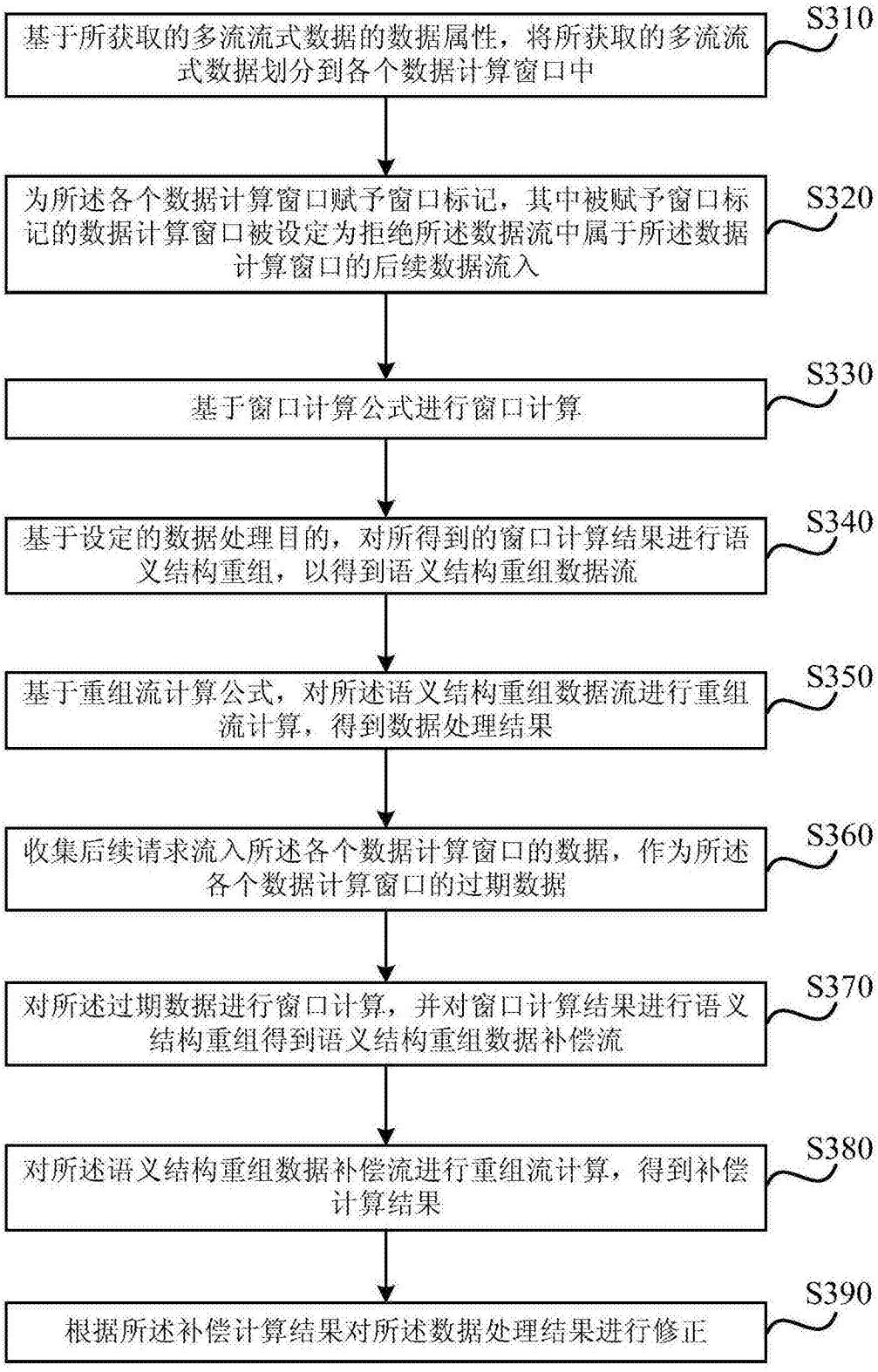

基于所获取的多流流式数据的数据属性，将所获取的多流流式数据划分到各个数据计算窗口中
S310
为所述各个数据计算窗口赋予窗口标记，其中被赋予窗口标记的数据计算窗口被设定为拒绝所述数据流中属于所述数据计算窗口的后续数据流入
S320
基于窗口计算公式进行窗口计算
S330
基于设定的数据处理目的，对所得到的窗口计算结果进行语义结构重组，以得到语义结构重组数据流
S340
基于重组流计算公式，对所述语义结构重组数据流进行重组流计算，得到数据处理结果
S350
收集后续请求流入所述各个数据计算窗口的数据，作为所述各个数据计算窗口的过期数据
S360
对所述过期数据进行窗口计算，并对窗口计算结果进行语义结构重组得到语义结构重组数据补偿流
S370
对所述语义结构重组数据补偿流进行重组流计算，得到补偿计算结果
S380
根据所述补偿计算结果对所述数据处理结果进行修正
S390

图3

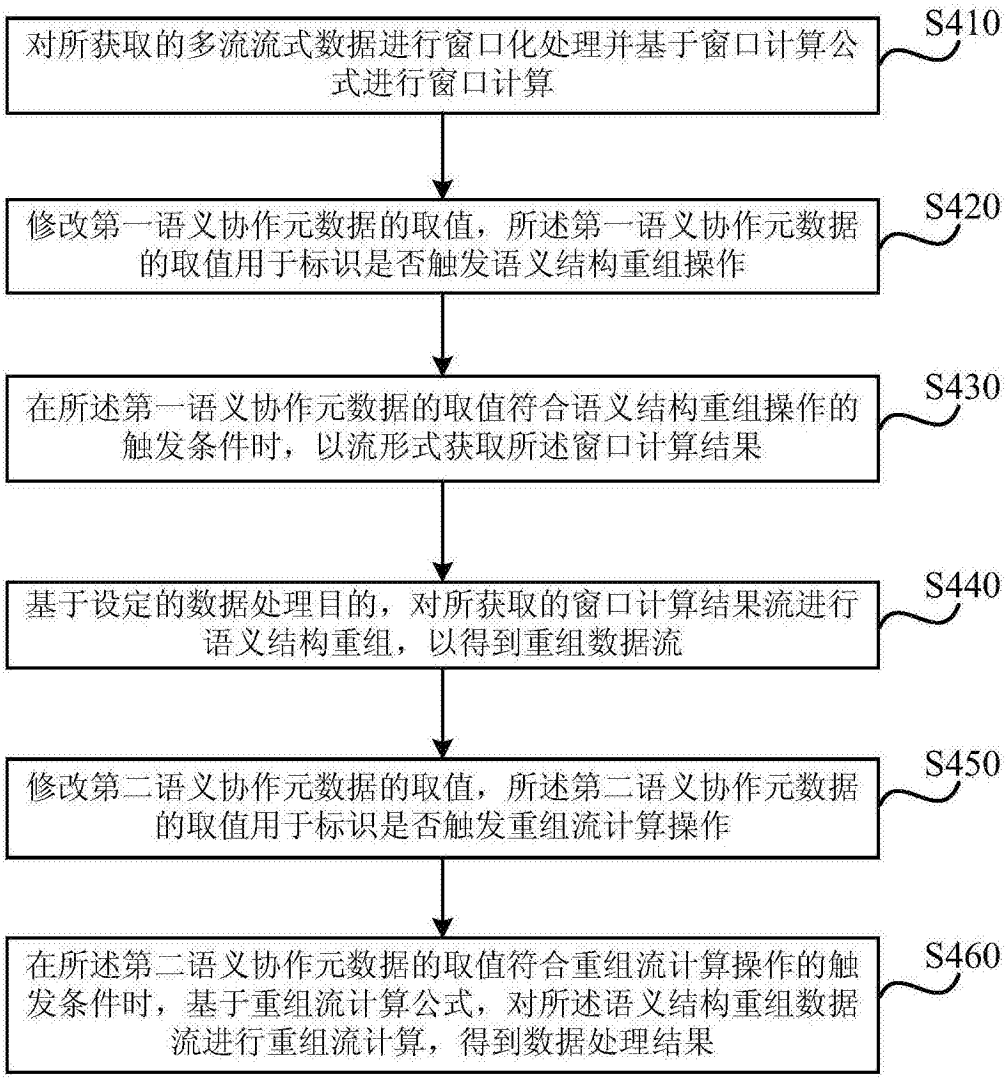

对所获取的多流流式数据进行窗口化处理并基于窗口计算公式进行窗口计算
S410
修改第一语义协作元数据的取值，所述第一语义协作元数据的取值用于标识是否触发语义结构重组操作
S420
在所述第一语义协作元数据的取值符合语义结构重组操作的触发条件时，以流形式获取所述窗口计算结果
S430
基于设定的数据处理目的，对所获取的窗口计算结果流进行语义结构重组，以得到重组数据流
S440
修改第二语义协作元数据的取值，所述第二语义协作元数据的取值用于标识是否触发重组流计算操作
S450
在所述第二语义协作元数据的取值符合重组流计算操作的触发条件时，基于重组流计算公式，对所述语义结构重组数据流进行重组流计算，得到数据处理结果
S460

图4

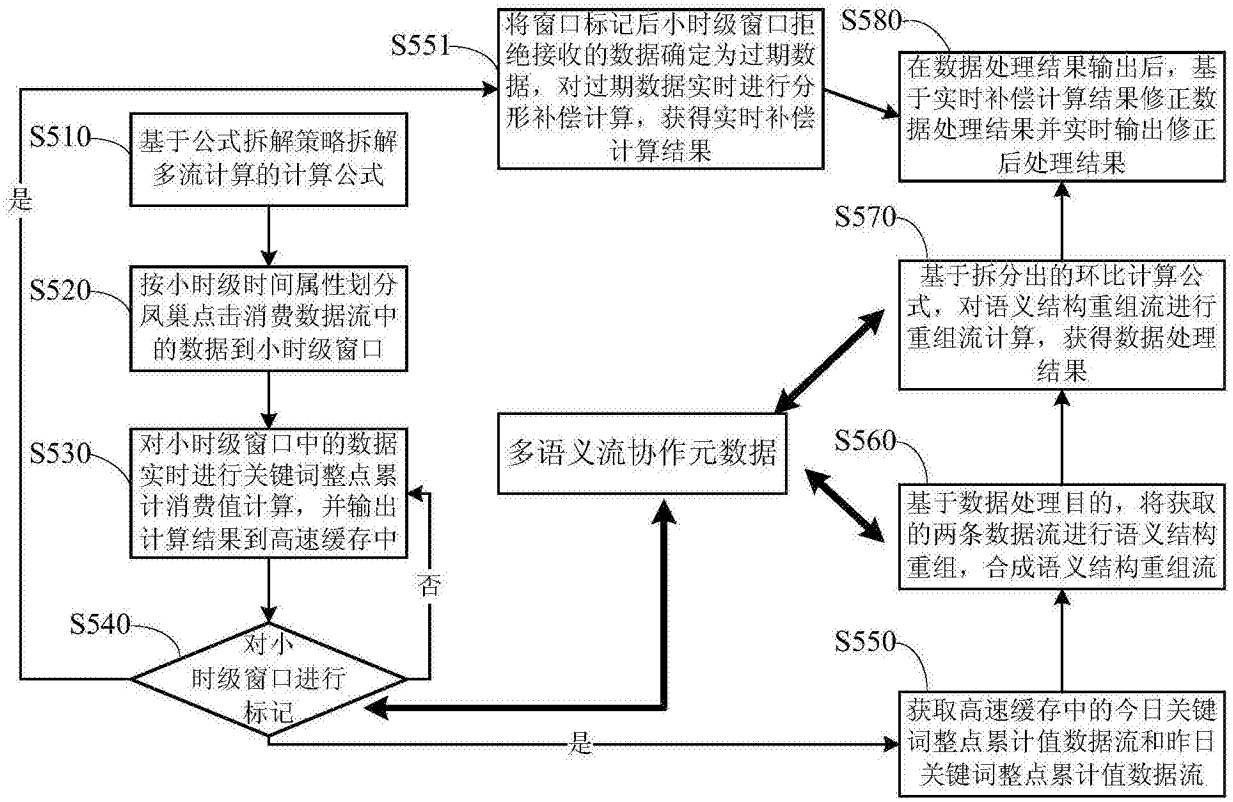

S551
将窗口标记后小时级窗口拒绝接收的数据确定为过期数据，对过期数据实时进行分形补偿计算，获得实时补偿计算结果
S580
在数据处理结果输出后，基于实时补偿计算结果修正数据处理结果并实时输出修正后处理结果
S510
基于公式拆解策略拆解多流计算的计算公式
是
S570
基于拆分出的环比计算公式，对语义结构重组流进行重组流计算，获得数据处理结果
S520
按小时级时间属性划分凤巢点击消费数据流中的数据到小时级窗口
多语义流协作元数据
S530
对小时级窗口中的数据实时进行关键词整点累计消费值计算，并输出计算结果到高速缓存中
S560
基于数据处理目的，将获取的两条数据流进行语义结构重组，合成语义结构重组流
否
S540
对小时级窗口进行标记
S550
获取高速缓存中的今日关键词整点累计值数据流和昨日关键词整点累计值数据流
是

图5a

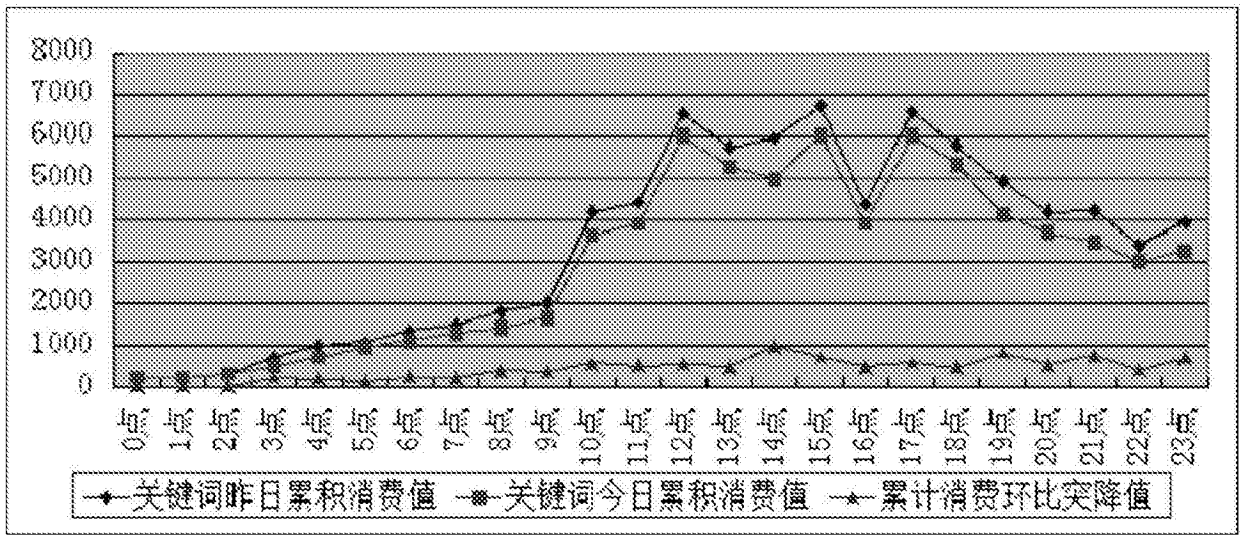

8000
7000
6000
5000
4000
3000
2000
1000
0
关键词昨日累积消费值
关键词今日累积消费值
累计消费环比实降值

图5b

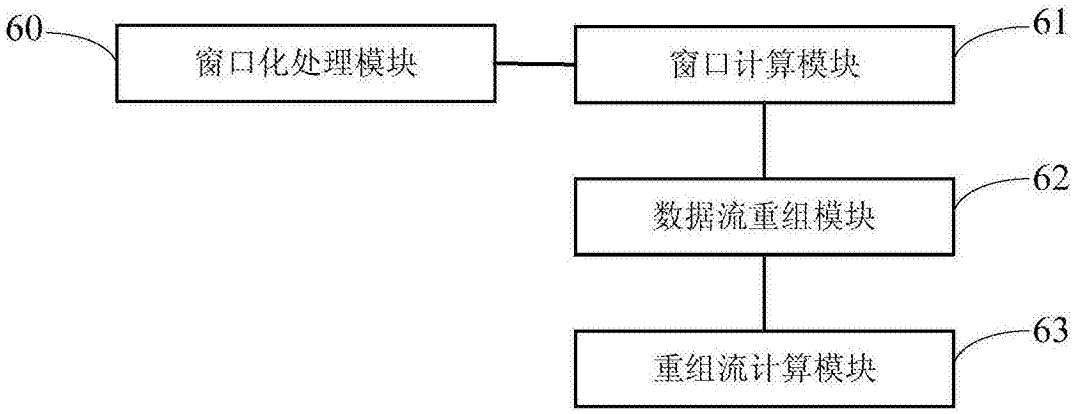

60
窗口化处理模块
61
窗口计算模块
62
数据流重组模块
63
重组流计算模块

图6a

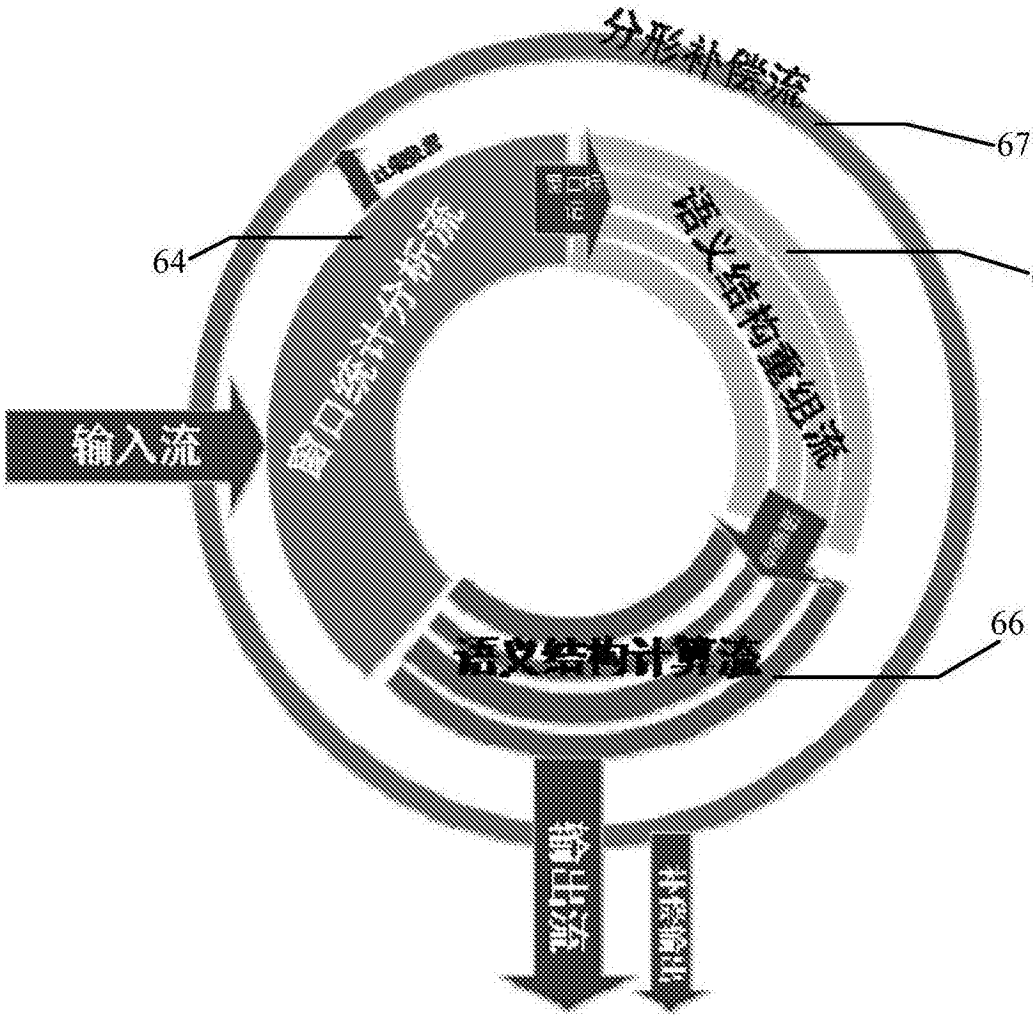

分形补偿流
67
64
65
语义结构重组流
输入流
窗口语义分析流
66
语义结构计算流

图6b