

ABSTRACT

Provided is a method for clustering a data stream. The data stream comprises a plurality of data partitions having a temporal relationship. The method comprises: acquiring N resulting models of preceding data partitions prior to a current data partition, wherein N is a natural number not less than 2, each of the resulting models is generated according to a clustering result of a corresponding preceding data partition, and each of the resulting models comprises one or more representative parameters in different categories; according to the N resulting models, determining a starting model of the current data partition, wherein one or more representative parameters in different categories in the starting model are determined based on representative parameters of the same category in the N resulting models; and using the starting model to cluster data records in the current data partition. The technical solution of the specification enables a clustering result of a current data partition to have both long-term efficacy and timeliness, thus preventing severe jitters of the clustering result, and improving the stability of a business associated with the data stream while providing support to timeliness of the business.