[0005] Accordingly, to generate the necessary amount of data to properly train a model to make accurate predictions, systems may need to use data aggregation. Data aggregation refers to the process of combining and summarizing individual data points or records into a more concise and meaningful representation. It involves collecting and merging data from multiple sources, grouping it based on certain criteria, and applying mathematical or statistical operations to derive summary statistics or insights. However, data aggregation also presents numerous technical challenges. For example, aggregating large volumes of data can be computationally intensive and may require efficient algorithms and infrastructure to handle the scale. As the dataset size increases, the processing time and resource requirements may grow significantly. Additionally, aggregating data from multiple sources may involve dealing with different data formats, structures, or schemas. Ensuring proper data integration and compatibility can be challenging, especially when the sources have inconsistencies or conflicting data representations. These technical problems are only further exacerbated when dealing with time-series data.

SUMMARY

[0005a] In an aspect, the present disclosure refers to a method for aggregating time-series data streams based on potential state characteristics following aggregation, the method comprising:

receiving a first model profile, wherein the first model profile is populated based on a first plurality of time-series data streams, and wherein the first model profile corresponds to a required state characteristic;

determining a first feature input based on the first model profile;

retrieving a first plurality of time-series data stream clusters, wherein the first plurality of time-series data stream clusters is generated by a first artificial intelligence model that is trained to cluster a plurality of available time-series data streams into the first plurality of time-series data stream clusters by aggregating a subset of the plurality of available time-series data streams;

determining a second feature input based on the first plurality of time-series data stream clusters;

determining, based on the first feature input and the second feature input, a first similarity between the first plurality of time-series data streams and each of the first plurality of time-series data stream clusters;

selecting a first time-series data stream cluster from the first plurality of time-series data stream clusters based on the first similarity exceeding a first similarity threshold;

generating a second time-series data stream cluster based on aggregating the first time-series data stream cluster and the first plurality of time-series data streams;

determining a first state characteristic for the second time-series data stream cluster;

comparing the first state characteristic to the required state characteristic; and

generating, at a user interface, a first recommendation for the first time-series data stream cluster based on comparing the first state characteristic to the required state characteristic.

[0005b] In another aspect, the present disclosure refers to a system arranged to carry out the method of the above aspect.

[0005c] In another aspect, the present disclosure refers to a computer program which, when executed by one or more processors, causes the one or more processors to carry out the method of the above aspect.

[0006] Systems and methods are described herein for novel uses and/or improvements to data aggregation related to artificial intelligence applications, specifically applications related to aggregating time-series data. As one example, systems and methods are described herein for predicting effects when aggregating time-series data and modifying the one or more data streams used to populate a model profile and/or feed an artificial intelligence application with the time-series data. That is, the systems and methods may aggregate time-series data streams based on potential state characteristics following aggregation. However, determining whether to aggregate data based on potential state characteristics following aggregation (as opposed to current characteristics of data that may be aggregated) introduces numerous technical challenges.

[0007] For example, time-series data is often collected at irregular intervals, making it difficult to align and aggregate the data points. Different time series may have different sampling frequencies or may not be synchronized, leading to challenges in combining them effectively. Second, time-series data frequently contains missing values or gaps due to various reasons such as sensor malfunction, data transmission errors, or simply unavailability of data during specific time periods. Aggregating the data requires handling these missing values appropriately to ensure accurate

results. Finally, time-series data from different sources or sensors may have temporal

example, time-series data can be susceptible to noise, outliers, and/or measurement errors. Identifying and filtering out such anomalies while aggregating the data is important to ensure reliable and accurate results. Similarly, aggregating time-series data involves considering the statistical dependencies and relationships between data points. Simple averaging or summation may not be appropriate if there are complex interactions or dependencies within the data.

[0011] Accordingly, the systems and methods may determine whether the time-series data resulting from the aggregation has prerequisite state characteristics (e.g., an acceptable level of noise, outliers, and/or measurement errors). After comparing any resulting state characteristic to the required state characteristic of the model profile, the system may then generate a first recommendation for the first time-series data stream cluster based on comparing the first state characteristic to the required state characteristic. For example, the recommendation may indicate that aggregating particular data, from particular data streams, will or will not result in potential state characteristics following that aggregation.

[0012] In some examples, systems and methods for aggregating time-series data streams based on potential state characteristics following aggregation are described. For example, the system may receive a first model profile, wherein the first model profile is populated based on a first plurality of time-series data streams, and wherein the first model profile corresponds to a required state characteristic. The system may retrieve a first plurality of time-series data stream clusters, wherein the first plurality of data stream clusters is generated by a first artificial intelligence model that is trained to cluster a plurality of available time-series data streams into the first plurality of time-series data stream clusters by aggregating a subset of the plurality of available time-series data streams. The system may determine a first similarity between the first plurality of time-series data streams and each of the first plurality of time-series data stream clusters. The system may select a first time-series data stream cluster from the first plurality of time-series data stream clusters based on the first similarity exceeding a first similarity threshold. The system may generate a second time-series data stream cluster based on aggregating the first time-series data stream cluster and the first plurality of time-series data streams. The system may determine a first state characteristic for the second time-series data stream cluster. The system may compare the first state characteristic to the required state characteristic. The system may generate, at a user interface, a first recommendation for the first time-series data stream cluster based on comparing the first state characteristic to the required state characteristic.