Overview

========

The \*\*aggregation framework\*\* is a set of libraries and utilities that allows teams to flexibly

compute aggregate (counting) features in both batch and in real-time. Aggregate features can capture

historical interactions between on arbitrary entities (and sets thereof), conditional on provided features

and labels.

These types of engineered aggregate features have proven to be highly impactful across different teams at Twitter.

What are some features we can compute?

--------------------------------------

The framework supports computing aggregate features on provided grouping keys. The only constraint is that these keys are sparse binary features (or are sets thereof).

For example, a common use case is to calculate a user's past engagement history with various types of tweets (photo, video, retweets, etc.), specific authors, specific in-network engagers or any other entity the user has interacted with and that could provide signal. In this case, the underlying aggregation keys are `userId`, `(userId, authorId)` or `(userId, engagerId)`.

In Timelines and MagicRecs, we also compute custom aggregate engagement counts on every `tweetId`. Similary, other aggregations are possible, perhaps on `advertiserId` or `mediaId` as long as the grouping key is sparse binary.

What implementations are supported?

-----------------------------------

Offline, we support the daily batch processing of DataRecords containing all required input features to generate

aggregate features. These are then uploaded to Manhattan for online hydration.

Online, we support the real-time aggregation of DataRecords through Storm with a backing memcache that can be queried

for the real-time aggregate features.

Additional documentation exists in the [docs folder](docs)

Where is this used?

--------------------

The Home Timeline heavy ranker uses a varierty of both [batch and real time features](../../../../src/scala/com/twitter/timelines/prediction/common/aggregates/README.md) generated by this framework.

These features are also used for email and other recommendations.