package com.twitter.simclusters\_v2.scalding.embedding.tfg

import com.twitter.dal.client.dataset.KeyValDALDataset

import com.twitter.scalding.\_

import com.twitter.scalding\_internal.multiformat.format.keyval.KeyVal

import com.twitter.simclusters\_v2.hdfs\_sources.EntityEmbeddingsSources

import com.twitter.simclusters\_v2.thriftscala.{

EmbeddingType,

ModelVersion,

SimClustersEmbeddingId,

UserToInterestedInClusterScores,

SimClustersEmbedding => ThriftSimClustersEmbedding

}

import com.twitter.wtf.scalding.jobs.common.{AdhocExecutionApp, ScheduledExecutionApp}

/\*\*

\* Apps to generate fav-based Topic-Follow-Graph (TFG) topic embeddings from inferred languages

\* The fav-based embeddings are built from topic followers' fav-based InterestedIn

\*/

/\*\*

./bazel bundle src/scala/com/twitter/simclusters\_v2/scalding/embedding/tfg:fav\_inferred\_lang\_tfg\_topic\_embeddings-adhoc

scalding remote run \

--user cassowary \

--keytab /var/lib/tss/keys/fluffy/keytabs/client/cassowary.keytab \

--principal service\_acoount@TWITTER.BIZ \

--cluster bluebird-qus1 \

--main-class com.twitter.simclusters\_v2.scalding.embedding.tfg.FavInferredLanguageTfgBasedTopicEmbeddingsAdhocApp \

--target src/scala/com/twitter/simclusters\_v2/scalding/embedding/tfg:fav\_inferred\_lang\_tfg\_topic\_embeddings-adhoc \

--hadoop-properties "scalding.with.reducers.set.explicitly=true mapreduce.job.reduces=4000" \

-- --date 2020-06-28

\*/

object FavInferredLanguageTfgBasedTopicEmbeddingsAdhocApp

extends InferredLanguageTfgBasedTopicEmbeddingsBaseApp

with AdhocExecutionApp {

override val isAdhoc: Boolean = true

override val embeddingType: EmbeddingType = EmbeddingType.FavInferredLanguageTfgTopic

override val embeddingSource: KeyValDALDataset[

KeyVal[SimClustersEmbeddingId, ThriftSimClustersEmbedding]

] = EntityEmbeddingsSources.FavInferredLanguageTfgTopicEmbeddingsDataset

override val pathSuffix: String = "fav\_inferred\_lang\_tfg\_topic\_embeddings"

override val modelVersion: ModelVersion = ModelVersion.Model20m145kUpdated

override def scoreExtractor: UserToInterestedInClusterScores => Double = scores =>

scores.favScore.getOrElse(0.0)

}

/\*\*

./bazel bundle src/scala/com/twitter/simclusters\_v2/scalding/embedding/tfg:fav\_inferred\_lang\_tfg\_topic\_embeddings

capesospy-v2 update --build\_locally --start\_cron fav\_inferred\_lang\_tfg\_topic\_embeddings src/scala/com/twitter/simclusters\_v2/capesos\_config/atla\_proc3.yaml

\*/

object FavInferredLanguageTfgBasedTopicEmbeddingsScheduledApp

extends InferredLanguageTfgBasedTopicEmbeddingsBaseApp

with ScheduledExecutionApp {

override val isAdhoc: Boolean = false

override val embeddingType: EmbeddingType = EmbeddingType.FavInferredLanguageTfgTopic

override val embeddingSource: KeyValDALDataset[

KeyVal[SimClustersEmbeddingId, ThriftSimClustersEmbedding]

] = EntityEmbeddingsSources.FavInferredLanguageTfgTopicEmbeddingsDataset

override val pathSuffix: String = "fav\_inferred\_lang\_tfg\_topic\_embeddings"

override val modelVersion: ModelVersion = ModelVersion.Model20m145kUpdated

override def scoreExtractor: UserToInterestedInClusterScores => Double = scores =>

scores.favScore.getOrElse(0.0)

override val firstTime: RichDate = RichDate("2020-07-04")

override val batchIncrement: Duration = Days(1)

}