package com.twitter.timelines.prediction.features.simcluster

import com.twitter.dal.personal\_data.thriftjava.PersonalDataType.\_

import com.twitter.finagle.stats.StatsReceiver

import com.twitter.ml.api.Feature.\_

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

import com.twitter.timelines.data\_processing.ml\_util.aggregation\_framework.TypedAggregateGroup

import scala.collection.JavaConverters.\_

class SimclusterFeaturesHelper(statsReceiver: StatsReceiver) {

import SimclusterFeatures.\_

private[this] val scopedStatsReceiver = statsReceiver.scope(getClass.getSimpleName)

private[this] val invalidSimclusterModelVersion = scopedStatsReceiver

.counter("invalidSimclusterModelVersion")

def fromUserClusterInterestsPair(

userInterestClustersPair: (Long, ClustersUserIsInterestedIn)

): Option[SimclusterFeatures] = {

val (userId, userInterestClusters) = userInterestClustersPair

if (userInterestClusters.knownForModelVersion == SIMCLUSTER\_MODEL\_VERSION) {

val userInterestClustersFavScores = for {

(clusterId, scores) <- userInterestClusters.clusterIdToScores

favScore <- scores.favScore

} yield (clusterId.toString, favScore)

Some(

SimclusterFeatures(

userId,

userInterestClusters.knownForModelVersion,

userInterestClustersFavScores.toMap

)

)

} else {

// We maintain this counter to make sure that the hardcoded modelVersion we are using is correct.

invalidSimclusterModelVersion.incr

None

}

}

}

object SimclusterFeatures {

// Check http://go/simclustersv2runbook for production versions

// Our models are trained for this specific model version only.

val SIMCLUSTER\_MODEL\_VERSION = "20M\_145K\_dec11"

val prefix = s"simcluster.v2.$SIMCLUSTER\_MODEL\_VERSION"

val SIMCLUSTER\_USER\_INTEREST\_CLUSTER\_SCORES = new SparseContinuous(

s"$prefix.user\_interest\_cluster\_scores",

Set(EngagementScore, InferredInterests).asJava

)

val SIMCLUSTER\_USER\_INTEREST\_CLUSTER\_IDS = new SparseBinary(

s"$prefix.user\_interest\_cluster\_ids",

Set(InferredInterests).asJava

)

val SIMCLUSTER\_MODEL\_VERSION\_METADATA = new Text("meta.simcluster\_version")

}

case class SimclusterFeatures(

userId: Long,

modelVersion: String,

interestClusterScoresMap: Map[String, Double])