

Part 4

WEKA K-means Clustering

Kimya Dhade Kalyani Deshmukh Sylvia Shun

K=2

=== Run information ===

Scheme: weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

Relation: user

Instances: 1000

Attributes: 4

 numFollowers

 numFriends

 createdAt

Ignored: iduser

Test mode: evaluate on training data

=== Clustering model (full training set) ===

kMeans

Number of iterations: 7

Within cluster sum of squared errors: 997.0673514490861

Initial starting points (random):

Cluster 0: 621,199,1/24/2014

Cluster 1: 3192,992,3/14/2014

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute	Full Data: (1000.0)	Cluster 0: (975.0)	Cluster 1: (25.0)
numFollowers	4459.005	1867.1672	105540.68
numFriends	2283.703	1218.3949	43830.72
createdAt	1/18/2012	1/18/2012	6/13/2011

Time taken to build model (full training data) : 0.1 seconds

=== Model and evaluation on training set ===

Clustered Instances

0 975 (98%)

1 25 (3%)

K=5

=== Run information ===

Scheme: weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 5 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

Relation: user

Instances: 1000

Attributes: 4

numFollowers

numFriends

createdAt

Ignored: iduser

Test mode: evaluate on training data

=== Clustering model (full training set) ===

kMeans

Number of iterations: 21

Within cluster sum of squared errors: 986.5047802266379

Initial starting points (random):

Cluster 0: 621,199,1/24/2014

Cluster 1: 3192,992,3/14/2014

Cluster 2: 382,310,3/19/2016

Cluster 3: 358,286,2/7/2016

Cluster 4: 128,556,7/21/2017

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute	Full Data (1000.0)	Cluster 0: (8.0)	Cluster 1: (17.0)	Cluster 2: (79.0)	Cluster 3: (874.0)	Cluster 4: (22.0)
numFollowers	4459.005	59507.5	119235	8261.2278	828.9336	26310.2273
numFriends	2283.703	271.5	53055.0588	5499.5949	572.0355	20234.9091
createdAt	1/18/2012	8/5/2017	6/13/2011	1/18/2012	10/28/2017	3/6/2012

Time taken to build model (full training data) : 0.09 seconds

=== Model and evaluation on training set ===

Clustered Instances

0 8 (1%)

1 17 (2%)

2 79 (8%)

3 874 (87%)

4 22 (2%)

K=10

=== Run information ===

Scheme: weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 10 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

Relation: user

Instances: 1000

Attributes: 4

numFollowers

numFriends

createdAt

Ignored: iduser

Test mode: evaluate on training data

=== Clustering model (full training set) ===

kMeans

=====

Number of iterations: 38

Within cluster sum of squared errors: 973.5305060509261

Initial starting points (random):

Cluster 0: 621,199,1/24/2014

Cluster 1: 3192,992,3/14/2014

Cluster 2: 382,310,3/19/2016

Cluster 3: 358,286,2/7/2016

Cluster 4: 128,556,7/21/2017

Cluster 5: 36,199,11/3/2017

Cluster 6: 95,24,1/11/2014

Cluster 7: 4,56,2/4/2017

Cluster 8: 276,183,4/21/2009

Cluster 9: 707,2145,9/16/2016

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute	Full Data 1000. 0	0:7.0	1:7.0	2:31.0	3:66.0	4:19.0	5:144. .0	6:65 4.0	7:41.0	8:23.0	9:8.0
numFollowers	4459. 005	30932 .4286	1712 16	13420 .0645	2906. 4545	39275 .6316	1247. 5764	326. 2829	4239. 4146	12998 .3478	10300 7.875
numFriends	2283. 703	1500. 2857	5897. 1429	10107 .9355	3898. 2424	28441 .4737	1537. 2014	282. 3532	559.8 78	1873. 087	81106 .375
createdAt	1/18/ 2012	2/16/ 2013	12/9/ 2010	1/18/ 2012	10/28 /2017	3/6/2 012	6/6/2 017	7/7/ 2009	11/28 /2017	4/21/ 2009	2/14/ 2015

Time taken to build model (full training data) : 0.22 seconds

=== Model and evaluation on training set ===

Clustered Instances

- 0 7 (1%)
- 1 7 (1%)
- 2 31 (3%)
- 3 66 (7%)
- 4 19 (2%)
- 5 144 (14%)
- 6 654 (65%)
- 7 41 (4%)
- 8 23 (2%)
- 9 8 (1%)