AI2

Assignment 1

Lazar Popov (s3340473)
Sander Kaatee (s3150798)
Group 19

October 12, 2020

6 Final Questions - Sander Kaatee

6.1 Parameters

Prefetch threshold

A lower prefetch threshold will result in a higher amount of hits and thus a higher hitrate, however the accuracy will decrease as the result of the decrease in this parameter is that a larger chunk of URL's will be prefetched.

Number of clusters

The number of clusters increases the hitrate of both algorithms, however this comes at a cost of the computing resources and accuracy, that is, the accuracy goes down as the number of clusters increases.

Clustering algorithm

The results of the K-Means and the Kohonen algorithm seem very similar. When set to the same amount of clusters (K-Means: 9, Kohonen 3x3), K-Means performs better:

K-Means:
Hitrate: 0.65
Accuracy: 0.51
Kohonen:
Hitrate: 0.42
Accuracy: 0.57

However, another factor at play here is the amount of training epochs for the Kohonen algorithm, on this test it was set to 100, going much lower will deteriorate the performance of Kohonen even more, an higher amount epochs does not seem to have effect for a map of 3x3 clusters.
With larger amount of clusters we can see that K-means has in general a higher hitrate than Kohonen, but Kohonen can have a higher accuracy. However it also becomes clear that the Kohonen algorithm is a lot more computationally intensive.

6.2 Algorithms

Workings

Kohonen's algorithm cluster centers do not depend on it's members (in fact the way we implemented it the clusters only gain members once the set number of epochs has passed), while the K-means

cluster centers do depend on its members. In a sense the Kohonen's centers move towards the clients, while for K-means it is reversed: the clients jump towards the clusters.

Computational complexity

K-means:
K + (K*N + N)*E = (K*N + N)*E, where K is the amount of clusters, N is the amount of input data and E is the amount of epochs.
Kohonen:
K + (K + 0.01 - 0.5 K) * E = 2K*E.

Final performance

Discussed above.

6 Final Questions - Lazar Popov

6.1 Influence of parameters

6.1.1 Prefetch threshold

A higher prefetch threshold results in a higher accuracy but lower hitrate. Lowering the threshold has the opposite result.

6.1.2 Number of clusters/size of map

When increasing the number of clusters the hitrate increases, the accuracy decreases, and the computational time increases as well.

6.2 Clustering algorithm

At small values such as 9 clusters (3x3 map), the K-Means has better hitrate than Kohone (0.58 vs 0.44) but lower accuracy (0.49 vs 0.56). On bigger values 49 clusters (7x7 map) both algorithms have similar accuracy (0.49) but Kohones had better Hitrate(0.69 vs 0.65)

6.3 Comparison of algorithms

6.3.1 Workings

Kohones relies on a network and uses learning rule to identify the clusters. Also the clusters centroids do not rely explicitly on the members of the clusters. The K-Means centroid rely on the members of the clusters and the creation of the cluster centroids is done by averaging all the members.

6.3.2 Computational complexity

K-means:
K + (K*N + N)*E = (K*N + N)*E, where K is the amount of clusters, N is the amount of input data and E is the amount of epochs.
Kohonen:
K + (K + 0.01 - 0.5 K) * E = 2K*E.

Final performance

K-Means is faster with hitrates and accuracies of the two algorithms are very close. It seems that K-means is the better algorithm but I believe with very large values for Kohonen on the number of iterations in very large in difficult data sets Kohonen will outperform K-Means