Subject: Re: Clustering
Posted by Jeremy Bailin on Tue, 01 Nov 2011 19:16:53 GMT
View Forum Message <> Reply to Message

On 10/31/11 11:40 AM, Kai Muehlbauer wrote:
> Hi all,
>
> I took a big step forward.
>
> I slightly changed my histograms. I reduced the number of bins by
> increasing the binsize. I cut off noise before the histograms which also
> reduces number of bins. Then I fill the histograms in an array similar
> to Mario is doing.
>
> FOR K=0L, 1999 DO BEGIN
> FOR J=0L, 359 DO BEGIN
> array = REFORM(source[K,*,J)]
> hist_arr = HISTOGRAM(array,BINSIZE=0.5, MAX=7.5, MIN=0)
> Array[*,K*360L+J] = hist_arr
> ENDFOR
> ENDFOR
>
> Then the weights for 10 Clusters are calculated and CLUSTER is called
>
> weights = CLUST_WTS(array2, N_CLUSTERS = 10)
> tmp_result = CLUSTER(array1, weights, N_CLUSTERS = 10)
>
> Then the data needs REFORMing
>
> result1 = REFORM(tmp_result,360,2000)
>
> and in my case the dimensions need to be interchanged
>
> FOR I=0,range1 - 1 DO BEGIN
> result[I,*] = result1[*,I]
>
> Anyway the results were not useful. I noticed that a great deal (about
> 90 percent) of the histograms could be grouped into one cluster. So I
> reduced the histograms used to calculate the weights to a reasonable
> amount, to get better weights also for the remaining 10 percent.
>
> After that I get quite usable clusters of my data. I think with a little
> tweaking there should be even better results.
>
> There should also be a speedup possible in the above code. But I'am
> still in the learning phase, so a little help is appreciated. I still
> struggle with those dimensions.
>

> Thanks Mario for providing your example. I tried this but got an out of
> memory error while calculating the distance matrix. But that was before
> my reduction of histogram number of bins. I will test this later and
> come back with some results in november ;-)
>
> Cheers,
> Kai
>

You can probably use JD's HIST_ND to get rid of those for loops, which should speed things up.

-Jeremy.
