Subject: Cluster analysis
Posted by mph410 on Fri, 27 Feb 1998 08:00:00 GMT
View Forum Message <> Reply to Message

I have a 3D image array containing a series of clusters of pixels of value 1, in a sea of value zero pixels.

Has anyone an idea on how to produce an algorithm to count the number of pixels in each of these clusters?

Many thanks.

John Dickson

Subject: Re: Cluster analysis
Posted by David Foster on Mon, 09 Mar 1998 08:00:00 GMT
View Forum Message <> Reply to Message

Alex Schuster wrote:
>
> John Dickson wrote:
>
>> I have a 3D image array containing a series of clusters of pixels
>> of value 1, in a sea of value zero pixels.
>>
>> Has anyone an idea on how to produce an algorithm to count the
>> number of pixels in each of these clusters?
>
> Are the clusters connected? If not, I would look at the SEARCH3D
> routine.
>
> If you need more than that, maybe the routines NN_LEARN and NN_CLUST
> could help, this is an implementations of the k-means cluster analysis.
> I don't recall where I got them from, they were a bit hard to find, so
> if anyone is interested, just ask me and I mail them.
>
> Alex

If you haven't found what you need, email me and I will dig up a
C routine that does exactly what you need. It orders the clusters
in decreasing size, and assumes that they are NOT connected. It is
much faster than search3d.pro .

Dave

~~~~~~ ~~~~~~
David S. Foster Univ. of California, San Diego
~~~~~~

Programmer/Analyst   Brain Image Analysis Laboratory
foster@bial1.ucsd.edu Department of Psychiatry
(619) 622-5892   8950 Via La Jolla Drive, Suite 2240
                      La Jolla, CA 92037
~~~~~~ ~~~~~~~

Subject: Re: Cluster analysis
Posted by James Kuyper on Fri, 03 Dec 2004 14:44:01 GMT

View Forum Message <> Reply to Message

Harald wrote:

> The cluster analysis in IDL with clust_wts and cluster finds only 8
> useful clusters. I can set n_clusters to a greater number, however all
> cluster more than 8 useful ones are at locations, that do not contain
> any of the samples. Is there any statistical reason that the number of
> clusters can never exceed 8?
>
> I have a data set with 24 variables and 7680 samples and would expect
> that I can easily find 20 or so different clusters.
>
> Does anybody have better IDL software that does the cluster analysis
> properly?

Could you post a copy of your dataset, so we can try to duplicate your (lack of) results?

Subject: Re: Cluster analysis
Posted by Chris.Jacobsen on Fri, 03 Dec 2004 23:30:42 GMT

View Forum Message <> Reply to Message

hfrey@ssl.berkeley.edu (Harald) wrote in message news:<51c9fb99.0412021550.41c2d53d@posting.google.com>...

> The cluster analysis in IDL with clust_wts and cluster finds only 8
> useful clusters. I can set n_clusters to a greater number, however all
> cluster more than 8 useful ones are at locations, that do not contain
> any of the samples. Is there any statistical reason that the number of
> clusters can never exceed 8?
>
> I have a data set with 24 variables and 7680 samples and would expect
> that I can easily find 20 or so different clusters.
>
> Does anybody have better IDL software that does the cluster analysis
> properly?

In our work we started out using the "stock" IDL
~~~~~~

cluster routine but we have added to it a bit.
Still, we have not changed the basic algorithm.
We've found that preparation of the data can
make a big difference. If the variation in
variable X is 100 times bigger than the variation
in variable Y, then the clustering (which looks
at simple Euclidian distance) will not see the
variation in Y very well. One approach is
to subtract the mean of each variable, and
apply a scale factor to the data in variable
Y so that it is spread out over the same distance
as in variable X.

With data preparation of that sort, the stock
IDL routine can certainly find more than 8 clusters.

A paper on our work is at
http://xray1.physics.sunysb.edu/~micros/publications/papers/ lerotic_ultramic_2004.pdf

---

Subject: Re: Cluster analysis
Posted by George N. White III on Sun, 05 Dec 2004 14:43:58 GMT

View Forum Message <> Reply to Message

On Fri, 3 Dec 2004, Chris Jacobsen wrote:

> In our work we started out using the "stock" IDL
> cluster routine but we have added to it a bit.
> Still, we have not changed the basic algorithm.
> We've found that preparation of the data can
> make a big difference. If the variation in
> variable X is 100 times bigger than the variation
> in variable Y, then the clustering (which looks
> at simple Euclidian distance) will not see the
> variation in Y very well. One approach is
> to subtract the mean of each variable, and
> apply a scale factor to the data in variable
> Y so that it is spread out over the same distance
> as in variable X.

Sound advice. You should also consider whether a
non-linear transform (e.g. alog()) should be applied
to some variables. Many people overlook the rank
transform, which gives you a distance measure that
is essentially the number of observations with values
between the two points. This is a way to make sense
of comparisons between different physical quantities.

--
George N. White III <aa056@chebucto.ns.ca>

Subject: Re: Cluster analysis
Posted by Harald Frey on Mon, 06 Dec 2004 22:21:00 GMT

View Forum Message <> Reply to Message

My original data set it too large to be posted but I used an earlier
posting about cluster analysis to create a little demonstration. My
program follows this posting. It uses a function gauss2 by Craig
Markwardt that I include in case you don't have it. In case my program
does not make it though the email, it can be accessed at our anonymous
ftp site

```
ftp sprite.ssl.berkeley.edu
cd pub/hfrey/idl/
get cluster_play_10.pro
```

I create 10 clusters, but IDL find only 8 correctly and puts the
remaining 2 clusters close to [0,0] where there are no original data
points.

Harald

```
 ;========================================================== ======
;+
; NAME:
;   GAUSS2
;
; AUTHOR:
;   Craig B. Markwardt, NASA/GSFC Code 662, Greenbelt, MD 20770
;   craigm@lheamail.gsfc.nasa.gov
;
; PURPOSE:
;   Compute Gaussian curve given the mean, sigma and area.
;
; MAJOR TOPICS:
;   Curve and Surface Fitting
;
; CALLING SEQUENCE:
;   YVALS = GAUSS2(X, Y, [XCENT, YCENT, SIGMA, PEAK])
;
; DESCRIPTION:
;
;  This routine computes the values of a Gaussian function whose
;  X-values, mean, sigma, and total area are given.  It is meant to be
;  a demonstration for curve-fitting.
```

```
;
;  XVALS can be an array of X-values, in which case the returned
;  Y-values are an array as well.  The second parameter to GAUSS1
;  should be an array containing the MEAN, SIGMA, and total AREA, in
;  that order.
;
; INPUTS:
;   X - 2-dimensional array of "X"-values.
;   Y - 2-dimensional array of "Y"-values.
;
;   XCENT - X-position of gaussian centroid.
;   YCENT - Y-position of gaussian centroid.
;
;   SIGMA - sigma of the curve (X and Y widths are the same).
;
;   PEAK - the peak value of the gaussian function.
;
; RETURNS:
;
;   Returns the array of Y-values.
;
; EXAMPLE:
;
;   p = [2.2D, -0.7D, 1.4D, 3000.D]
;   x = (dindgen(200)*0.1 - 10.) # (dblarr(200) + 1)
;   y = (dblarr(200) + 1) # (dindgen(200)*0.1 - 10.)
;   z = gauss2(x, y, p)
;
;   Computes the values of the Gaussian at equispaced intervals in X
;   and Y (spacing is 0.1).  The gaussian has a centroid position of
;   (2.2, -0.7), standard deviation of 1.4, and peak value of 3000.
;
; REFERENCES:
;
; MODIFICATION HISTORY:
;   Written, 02 Oct 1999, CM
;
;-

function gauss2, x, y, p, _EXTRA=extra

u = ((x-p(0))/p(2))^2 + ((y-p(1))/p(2))^2
mask = u LT 100
f = p(3) * mask * exp(-0.5D * temporary(u) * mask)
mask = 0

return, f
end
```

```
 ;=========================================================== =======
pro Cluster_play_10
; program to create 10 clusters and try to find them all
; Harald Frey, December 6, 2004

FORWARD_FUNCTION gauss2

x = findgen(1000)*0.1 - 50. & y = x
xx = x # (y*0 + 1)
yy = (x*0 + 1) # y

; create 10 two-dimensional clusters
z = 30 * gauss2(xx, yy, [ 20D, 33D, .2, 1]) + $
10 * gauss2(xx, yy, [-30D,-13D, .2, 1]) + $
40 * gauss2(xx, yy, [-20D, 21D, .2, 1]) + $
10 * gauss2(xx, yy, [-10D,-11D, .2, 1]) + $
20 * gauss2(xx, yy, [-13D,  1D, .2, 1]) + $
10 * gauss2(xx, yy, [ 23D, 21D, .2, 1]) + $
30 * gauss2(xx, yy, [ 33D,-31D, .2, 1]) + $
10 * gauss2(xx, yy, [  3D,-41D, .2, 1]) + $
50 * gauss2(xx, yy, [ -3D, 11D, .2, 1]) + $
20 * gauss2(xx, yy, [ 10D, -2D, .2D, 1])
zi = floor(z)   ;; Convert to integer

;; Find the positions of significant data points
wh = where(z GT 5, ct)
if ct EQ 0 then message, 'ERROR: no signif points!'
xi = x(wh MOD 1000)
yi = y(floor(wh/1000))
xy = transpose([[xi],[yi]])

; input for cluster analysis
array=xy
; number of clusters
nc=10
; number of iterations
ni=10

; do cluster analysis
weights=clust_wts(array,n_clusters=nc,n_iterations=ni)

; display original distribution
window,0
plot, xi, yi, psym=3,title='Original distribution'

; display distribution with centers of clusters
window,1
```

```
plot, xi, yi, psym=3,title='10 clusters'
oplot, weights(0,*), weights(1,*), psym=1, symsize=3
print,'Cluster centers'
print,weights

END
```

Subject: Re: Cluster analysis
Posted by Chris[2] on Tue, 07 Dec 2004 19:32:13 GMT

View Forum Message <> Reply to Message

Hi Harald,

Have you tried the new CLUSTER_TREE and DENDROGRAM/DENDRO_PLOT routines, which were introduced in IDL6.1? These give better control over which method is used to cluster, and might work well for your data.

Cheers,

Chris
Research Systems, Inc.

"Harald" <hfrey@ssl.berkeley.edu> wrote in message
news:51c9fb99.0412021550.41c2d53d@posting.google.com...

> The cluster analysis in IDL with clust_wts and cluster finds only 8
> useful clusters. I can set n_clusters to a greater number, however all
> cluster more than 8 useful ones are at locations, that do not contain
> any of the samples. Is there any statistical reason that the number of
> clusters can never exceed 8?
>
> I have a data set with 24 variables and 7680 samples and would expect
> that I can easily find 20 or so different clusters.
>
> Does anybody have better IDL software that does the cluster analysis
> properly?
>
> Thank you,
> Harald (hfrey@ssl.berkeley.edu)