Subject: Re: edit a large dataset
Posted by Peter Clinch on Wed, 03 Oct 2007 14:58:50 GMT

queiny wrote:
> Hi, I have a large dataset (>1e7 records). One field of the record is
> a string 'landuse'. Now I want to add one more field based on the
> 'landuse', i.e., if it is 'forest', then add 'aa', if it is 'urban',
> then add 'bb', etc.. Since there are more than 20 different landuse,
> use 'if-then', or 'case' for each record will be quite inefficient.

I generally find Perl is much better for hacking data sets than IDL,
because systematically fiddling with file contents is what it does very
well. Once it's been suitably butchered, then into IDL for the actual
crunching of numbers that IDL does very well.

Pete.
--
Peter Clinch   Medical Physics IT Officer
Tel 44 1382 660111 ext. 33637   Univ. of Dundee, Ninewells Hospital
Fax 44 1382 640177   Dundee DD1 9SY Scotland UK
net p.j.clinch@dundee.ac.uk   http://www.dundee.ac.uk/~pjclinch/

Subject: Re: edit a large dataset
Posted by Jean H. on Wed, 03 Oct 2007 16:29:31 GMT

> a=sort(data_orig)
> data_sorted=data_orig(a)
> ;processing based on 'data_sorted'
>
> but how could I go back the order of 'data_orig'? Is there a inversion
> of 'sort'.
>
> Thanks,

Hi,

well, you just have to do
data_orig[a] = data_sorted
or similarly
newLandUse[a] = new_sorted_land_use

Jean

PS: are you doing some kind of land-use change model?

Subject: Re: edit a large dataset
Posted by Vince Hradil on Wed, 03 Oct 2007 16:40:58 GMT

View Forum Message <> Reply to Message

On Oct 3, 9:48 am, queiny <quein...@yahoo.com> wrote:
> Hi, I have a large dataset (>1e7 records). One field of the record is
> a string 'landuse'. Now I want to add one more field based on the
> 'landuse', i.e., if it is 'forest', then add 'aa', if it is 'urban',
> then add 'bb', etc.. Since there are more than 20 different landuse,
> use 'if-then', or 'case' for each record will be quite inefficient.
>
> I am thinking of 'sort' the records based on 'landuse', and then use
> 'uniq' to locate the starting point and number of records for each
> 'landuse' type and add the new fields accordingly. But how could I go
> back to the original order afterward?
>
> a=sort(data_orig)
> data_sorted=data_orig(a)
> ;processing based on 'data_sorted'
>
> but how could I go back the order of 'data_orig'? Is there a inversion
> of 'sort'.
>
> Thanks,

a=sort(data_orig)
data_sorted=data_orig[a]
data_orig=data_sorted[sort(a)]

Cheers,
Vince

Subject: Re: edit a large dataset
Posted by Vince Hradil on Wed, 03 Oct 2007 16:45:04 GMT

View Forum Message <> Reply to Message

On Oct 3, 11:40 am, hradilv <hrad...@yahoo.com> wrote:
> On Oct 3, 9:48 am, queiny <quein...@yahoo.com> wrote:
>
>
>
>> Hi, I have a large dataset (>1e7 records). One field of the record is
>> a string 'landuse'. Now I want to add one more field based on the
>> 'landuse', i.e., if it is 'forest', then add 'aa', if it is 'urban',
>> then add 'bb', etc.. Since there are more than 20 different landuse,
>> use 'if-then', or 'case' for each record will be quite inefficient.
>

>> I am thinking of 'sort' the records based on 'landuse', and then use
>> 'uniq' to locate the starting point and number of records for each
>> 'landuse' type and add the new fields accordingly. But how could I go
>> back to the original order afterward?
>
>> a=sort(data_orig)
>> data_sorted=data_orig(a)
>> ;processing based on 'data_sorted'
>
>> but how could I go back the order of 'data_orig'? Is there a inversion
>> of 'sort'.
>
>> Thanks,
>
> a=sort(data_orig)
> data_sorted=data_orig[a]
> data_orig=data_sorted[sort(a)]
>
> Cheers,
> Vince

I was considering:
data_orig = data_orig[sort(data_orig[sort(data_orig)])]

But thought it was little to obfuscatory 8)

---

yeah, sort(a) will bring it back to original order. Thanks!!

>
> a=sort(data_orig)
> data_sorted=data_orig[a]
> data_orig=data_sorted[sort(a)]
>
> Cheers,
> Vince

Thanks. I don't know we can use

data_orig[a] = data_sorted.

that is the inversion of sort function.

We don't do landuse change modeling. The data we are using provide landuse infor for reference.

On Oct 3, 12:29 pm, Jean H <jghas...@DELTHIS.ucalgary.ANDTHIS.ca> wrote:

>
> well, you just have to do
> data_orig[a] = data_sorted
> or similarly
> newLandUse[a] = new_sorted_land_use
>
> Jean
>
> PS: are you doing some kind of land-use change model?