Subject: Re: remove duplicate elements from a multi-dimensional array efficiently in IDL
Posted by Jeremy Bailin on Fri, 01 May 2009 17:47:10 GMT
View Forum Message <> Reply to Message

On May 1, 12:13 pm, "chenb...@gmail.com" <chenb...@gmail.com> wrote:
> Hello, everyone!
>
> Is there anyone knows a routine in IDL that be capable to remove
> duplicate elements from a multi-dimensional array efficiently? I 'm
> now working with huge arrays, and I have written one by myself, it
> works but is with low efficiency.
>
> example of my problem:
> the input array:
> 1,10,9,100,200
> 2,11,8,101,201
> 2,11,8,101,201
> 3,10,9,100,200
> 4,7,12,99,199
> 2,11,8,101,201
>
> goal:
> remove the duplicate elements with the same values for the second and
> the third column.
>
> expected output:
> 1,10,9,100,200
> 2,11,8,101,201
> 4,7,12,99,199
>
> Thanks for your help!
>
> Bo

How's this:

```
input = [[1,10,9,100,200],[2,11,8,101,201],[2,11,8,101,201],$
  [3,10,9,100,200],[4,7,12,99,199],[2,11,8,101,201]]

; Step 1: Map your columns 2 and 3 into a single unique index
(requires ORD from JBIU):
col2ord = ord(input[1,*])
col3ord = ord(input[2,*])
index = col2ord + (max(col2ord)+1)*col3ord

; Step 2: Use histogram to find which ones have the same unique index
h = histogram(index, reverse_indices=ri)
```

```
; Step 3: Get the first one in each bin, and put back in sorted order
keep = ri[ri[where(h gt 0)]]
keep = keep[sort(keep)]

; Step 4: Print them out:
print, input[*,keep]

      1      10       9     100     200
      2      11       8     101     201
      4       7      12      99     199
```

-Jeremy.

Subject: Re: remove duplicate elements from a multi-dimensional array efficiently in IDL
Posted by Jeremy Bailin on Fri, 01 May 2009 18:36:18 GMT
View Forum Message <> Reply to Message

On May 1, 1:47 pm, Jeremy Bailin <astroco...@gmail.com> wrote:
> On May 1, 12:13 pm, "chenb...@gmail.com" <chenb...@gmail.com> wrote:
>
>
>
>> Hello, everyone!
>
>> Is there anyone knows a routine in IDL that be capable to remove
>> duplicate elements from a multi-dimensional array efficiently? I 'm
>> now working with huge arrays, and I have written one by myself, it
>> works but is with low efficiency.
>
>> example of my problem:
>> the input array:
>> 1,10,9,100,200
>> 2,11,8,101,201
>> 2,11,8,101,201
>> 3,10,9,100,200
>> 4,7,12,99,199
>> 2,11,8,101,201
>
>> goal:
>> remove the duplicate elements with the same values for the second and
>> the third column.
>
>> expected output:
>> 1,10,9,100,200

>> 2,11,8,101,201
>> 4,7,12,99,199
>
>> Thanks for your help!
>
>> Bo
>
> How's this:
>
> input = [[1,10,9,100,200],[2,11,8,101,201],[2,11,8,101,201],$
> [3,10,9,100,200],[4,7,12,99,199],[2,11,8,101,201]]
>
> ; Step 1: Map your columns 2 and 3 into a single unique index
> (requires ORD from JBIU):
> col2ord = ord(input[1,*])
> col3ord = ord(input[2,*])
> index = col2ord + (max(col2ord)+1)*col3ord
>
> ; Step 2: Use histogram to find which ones have the same unique index
> h = histogram(index, reverse_indices=ri)
>
> ; Step 3: Get the first one in each bin, and put back in sorted order
> keep = ri[ri[where(h gt 0)]]
> keep = keep[sort(keep)]
>
> ; Step 4: Print them out:
> print, input[*,keep]
>
> 1 10 9 100 200
> 2 11 8 101 201
> 4 7 12 99 199
>
> -Jeremy.

Incidentally, if you're dealing with huge arrays and run into memory problems with the histogram, you can replace:

index = col2ord + (max(col2ord)+1)*col3ord

with

index = ord(col2ord + (max(col2ord)+1)*col3ord)

which will make the histogram as compact as possible.

-Jeremy.

Subject: Re: remove duplicate elements from a multi-dimensional array efficiently in IDL
Posted by guillermo.castilla.ca on Sun, 03 May 2009 00:47:51 GMT
View Forum Message <> Reply to Message

On May 1, 12:36 pm, Jeremy Bailin <astroco...@gmail.com> wrote:
> On May 1, 1:47 pm, Jeremy Bailin <astroco...@gmail.com> wrote:
>
>
>
>> On May 1, 12:13 pm, "chenb...@gmail.com" <chenb...@gmail.com> wrote:
>
>>> Hello, everyone!
>
>>> Is there anyone knows a routine in IDL that be capable to remove
>>> duplicate elements from a multi-dimensional array efficiently? I 'm
>>> now working with huge arrays, and I have written one by myself, it
>>> works but is with low efficiency.
>
>>> example of my problem:
>>> the input array:
>>> 1,10,9,100,200
>>> 2,11,8,101,201
>>> 2,11,8,101,201
>>> 3,10,9,100,200
>>> 4,7,12,99,199
>>> 2,11,8,101,201
>
>>> goal:
>>> remove the duplicate elements with the same values for the second and
>>> the third column.
>
>>> expected output:
>>> 1,10,9,100,200
>>> 2,11,8,101,201
>>> 4,7,12,99,199
>
>>> Thanks for your help!
>
>>> Bo
>

If you don't have handy that ORD function Jeremy pointed out (I didn't know of it), and assuming your array is of byte type, you can do the following:

```
input = [[1,10,9,100,200],[2,11,8,101,201],[2,11,8,101,201],$
      [3,10,9,100,200],[4,7,12,99,199],[2,11,8,101,201]]
```

```
keep = Where(Histogram(1000L*input[1,*]+input[2,*], rev=r) GT 0)
keep = r[r[keep]]
print, input[*,keep[sort(keep)]]
      1      10       9     100     200
      2      11       8     101     201
      4       7      12      99     199
```

Cheers

Guillermo

Subject: Re: remove duplicate elements from a multi-dimensional array efficiently in IDL
Posted by Jeremy Bailin on Sun, 03 May 2009 00:57:38 GMT
View Forum Message <> Reply to Message

On May 2, 8:47 pm, guillermo.castilla.castell...@gmail.com wrote:
> On May 1, 12:36 pm, Jeremy Bailin <astroco...@gmail.com> wrote:
>
>
>
>> On May 1, 1:47 pm, Jeremy Bailin <astroco...@gmail.com> wrote:
>
>>> On May 1, 12:13 pm, "chenb...@gmail.com" <chenb...@gmail.com> wrote:
>
>>>> Hello, everyone!
>
>>>> Is there anyone knows a routine in IDL that be capable to remove
>>>> duplicate elements from a multi-dimensional array efficiently? I 'm
>>>> now working with huge arrays, and I have written one by myself, it
>>>> works but is with low efficiency.
>
>>>> example of my problem:
>>>> the input array:
>>>> 1,10,9,100,200
>>>> 2,11,8,101,201
>>>> 2,11,8,101,201
>>>> 3,10,9,100,200
>>>> 4,7,12,99,199
>>>> 2,11,8,101,201
>
>>>> goal:
>>>> remove the duplicate elements with the same values for the second and
>>>> the third column.
>
>>>> expected output:
>>>> 1,10,9,100,200

>>>> 2,11,8,101,201
>>>> 4,7,12,99,199
>
>>>> Thanks for your help!
>
>>>> Bo
>
> If you don't have handy that ORD function Jeremy pointed out (I didn't
> know of it), and assuming your array is of byte type, you can do the
> following:
>
> input = [[1,10,9,100,200],[2,11,8,101,201],[2,11,8,101,201],$
> [3,10,9,100,200],[4,7,12,99,199],[2,11,8,101,201]]
>
> keep = Where(Histogram(1000L*input[1,*]+input[2,*], rev=r) GT 0)
> keep = r[r[keep]]
> print, input[*,keep[sort(keep)]]
> 1 10 9 100 200
> 2 11 8 101 201
> 4 7 12 99 199
>
> Cheers
>
> Guillermo

You can find ord at:

http://web.astroconst.org/jbiu/jbiu-doc/math/ord.html

-Jeremy.

Subject: Re: remove duplicate elements from a multi-dimensional array efficiently in IDL
Posted by chenbo09@gmail.com on Sun, 03 May 2009 17:47:38 GMT
View Forum Message <> Reply to Message

On May 2, 7:57 pm, Jeremy Bailin <astroco...@gmail.com> wrote:
> On May 2, 8:47 pm, guillermo.castilla.castell...@gmail.com wrote:
>
>
>
>> On May 1, 12:36 pm, Jeremy Bailin <astroco...@gmail.com> wrote:
>
>>> On May 1, 1:47 pm, Jeremy Bailin <astroco...@gmail.com> wrote:
>
>>>> On May 1, 12:13 pm, "chenb...@gmail.com" <chenb...@gmail.com> wrote:
>

>>>> > Hello, everyone!
>
>>>> > Is there anyone knows a routine in IDL that be capable to remove
>>>> > duplicate elements from a multi-dimensional array efficiently? I 'm
>>>> > now working with huge arrays, and I have written one by myself, it
>>>> > works but is with low efficiency.
>
>>>> > example of my problem:
>>>> > the input array:
>>>> > 1,10,9,100,200
>>>> > 2,11,8,101,201
>>>> > 2,11,8,101,201
>>>> > 3,10,9,100,200
>>>> > 4,7,12,99,199
>>>> > 2,11,8,101,201
>
>>>> > goal:
>>>> > remove the duplicate elements with the same values for the second and
>>>> > the third column.
>
>>>> > expected output:
>>>> > 1,10,9,100,200
>>>> > 2,11,8,101,201
>>>> > 4,7,12,99,199
>
>>>> > Thanks for your help!
>
>>>> > Bo
>
>> If you don't have handy that ORD function Jeremy pointed out (I didn't
>> know of it), and assuming your array is of byte type, you can do the
>> following:
>
>> input = [[1,10,9,100,200],[2,11,8,101,201],[2,11,8,101,201],$
>>   [3,10,9,100,200],[4,7,12,99,199],[2,11,8,101,201]]
>
>> keep = Where(Histogram(1000L*input[1,*]+input[2,*], rev=r) GT 0)
>> keep = r[r[keep]]
>> print, input[*,keep[sort(keep)]]
>>   1   10   9   100   200
>>   2   11   8   101   201
>>   4   7   12   99   199
>
>> Cheers
>
>> Guillermo
>
> You can find ord at:

>
> http://web.astroconst.org/jbiu/jbiu-doc/math/ord.html
>
> -Jeremy.

Jeremy,

Thanks for your kind and prompt help!
It took my own routine 18 hours to do the job. I have just plug the codes you kindly offered into my codes, I'll let you know how efficient your routine is. Thanks!

Bo

Subject: Re: remove duplicate elements from a multi-dimensional array efficiently in IDL
Posted by chenbo09@gmail.com on Sun, 03 May 2009 17:50:49 GMT
View Forum Message <> Reply to Message

On May 2, 7:47 pm, guillermo.castilla.castell...@gmail.com wrote:
> On May 1, 12:36 pm, Jeremy Bailin <astroco...@gmail.com> wrote:
>
>
>
>> On May 1, 1:47 pm, Jeremy Bailin <astroco...@gmail.com> wrote:
>
>>> On May 1, 12:13 pm, "chenb...@gmail.com" <chenb...@gmail.com> wrote:
>
>>>> Hello, everyone!
>
>>>> Is there anyone knows a routine in IDL that be capable to remove
>>>> duplicate elements from a multi-dimensional array efficiently? I 'm
>>>> now working with huge arrays, and I have written one by myself, it
>>>> works but is with low efficiency.
>
>>>> example of my problem:
>>>> the input array:
>>>> 1,10,9,100,200
>>>> 2,11,8,101,201
>>>> 2,11,8,101,201
>>>> 3,10,9,100,200
>>>> 4,7,12,99,199
>>>> 2,11,8,101,201
>
>>>> goal:
>>>> remove the duplicate elements with the same values for the second and
>>>> the third column.

>
>>>> expected output:
>>>> 1,10,9,100,200
>>>> 2,11,8,101,201
>>>> 4,7,12,99,199
>
>>>> Thanks for your help!
>
>>>> Bo
>
> If you don't have handy that ORD function Jeremy pointed out (I didn't
> know of it), and assuming your array is of byte type, you can do the
> following:
>
> input = [[1,10,9,100,200],[2,11,8,101,201],[2,11,8,101,201],$
> [3,10,9,100,200],[4,7,12,99,199],[2,11,8,101,201]]
>
> keep = Where(Histogram(1000L*input[1,*]+input[2,*], rev=r) GT 0)
> keep = r[r[keep]]
> print, input[*,keep[sort(keep)]]
> 1 10 9 100 200
> 2 11 8 101 201
> 4 7 12 99 199
>
> Cheers
>
> Guillermo
Hi Guillermo,

Thanks for your suggestion! Have a nice weekend!

Bo

Subject: Re: remove duplicate elements from a multi-dimensional array efficiently in IDL
Posted by chenbo09@gmail.com on Sun, 03 May 2009 17:54:43 GMT
View Forum Message <> Reply to Message

On May 2, 7:47 pm, guillermo.castilla.castell...@gmail.com wrote:
> On May 1, 12:36 pm, Jeremy Bailin <astroco...@gmail.com> wrote:
>
>
>
>> On May 1, 1:47 pm, Jeremy Bailin <astroco...@gmail.com> wrote:
>
>>> On May 1, 12:13 pm, "chenb...@gmail.com" <chenb...@gmail.com> wrote:
>

>>>> Hello, everyone!
>
>>>> Is there anyone knows a routine in IDL that be capable to remove
>>>> duplicate elements from a multi-dimensional array efficiently? I 'm
>>>> now working with huge arrays, and I have written one by myself, it
>>>> works but is with low efficiency.
>
>>>> example of my problem:
>>>> the input array:
>>>> 1,10,9,100,200
>>>> 2,11,8,101,201
>>>> 2,11,8,101,201
>>>> 3,10,9,100,200
>>>> 4,7,12,99,199
>>>> 2,11,8,101,201
>
>>>> goal:
>>>> remove the duplicate elements with the same values for the second and
>>>> the third column.
>
>>>> expected output:
>>>> 1,10,9,100,200
>>>> 2,11,8,101,201
>>>> 4,7,12,99,199
>
>>>> Thanks for your help!
>
>>>> Bo
>
> If you don't have handy that ORD function Jeremy pointed out (I didn't
> know of it), and assuming your array is of byte type, you can do the
> following:
>
> input = [[1,10,9,100,200],[2,11,8,101,201],[2,11,8,101,201],$
> [3,10,9,100,200],[4,7,12,99,199],[2,11,8,101,201]]
>
> keep = Where(Histogram(1000L*input[1,*]+input[2,*], rev=r) GT 0)
> keep = r[r[keep]]
> print, input[*,keep[sort(keep)]]
> 1 10 9 100 200
> 2 11 8 101 201
> 4 7 12 99 199
>
> Cheers
>
> Guillermo
Hi Guillermo,

Thanks for your kind suggestion! Have a nice weekend!

Bo

Subject: Re: remove duplicate elements from a multi-dimensional array efficiently in IDL
Posted by chenbo09@gmail.com on Mon, 04 May 2009 15:12:53 GMT
View Forum Message <> Reply to Message

On May 3, 12:47 pm, "chenb...@gmail.com" <chenb...@gmail.com> wrote:
> On May 2, 7:57 pm, Jeremy Bailin <astroco...@gmail.com> wrote:
>
>
>
>> On May 2, 8:47 pm, guillermo.castilla.castell...@gmail.com wrote:
>
>>> On May 1, 12:36 pm, Jeremy Bailin <astroco...@gmail.com> wrote:
>
>>>> On May 1, 1:47 pm, Jeremy Bailin <astroco...@gmail.com> wrote:
>
>>>> > On May 1, 12:13 pm, "chenb...@gmail.com" <chenb...@gmail.com> wrote:
>
>>>> > > Hello, everyone!
>
>>>> > > Is there anyone knows a routine in IDL that be capable to remove
>>>> > > duplicate elements from a multi-dimensional array efficiently? I 'm
>>>> > > now working with huge arrays, and I have written one by myself, it
>>>> > > works but is with low efficiency.
>
>>>> > > example of my problem:
>>>> > > the input array:
>>>> > > 1,10,9,100,200
>>>> > > 2,11,8,101,201
>>>> > > 2,11,8,101,201
>>>> > > 3,10,9,100,200
>>>> > > 4,7,12,99,199
>>>> > > 2,11,8,101,201
>
>>>> > > goal:
>>>> > > remove the duplicate elements with the same values for the second and
>>>> > > the third column.
>
>>>> > > expected output:
>>>> > > 1,10,9,100,200
>>>> > > 2,11,8,101,201
>>>> > > 4,7,12,99,199
>

>>>> > > Thanks for your help!
>
>>>> > > Bo
>
>>> If you don't have handy that ORD function Jeremy pointed out (I didn't
>>> know of it), and assuming your array is of byte type, you can do the
>>> following:
>
>>> input = [[1,10,9,100,200],[2,11,8,101,201],[2,11,8,101,201],$
>>> [3,10,9,100,200],[4,7,12,99,199],[2,11,8,101,201]]
>
>>> keep = Where(Histogram(1000L*input[1,*]+input[2,*], rev=r) GT 0)
>>> keep = r[r[keep]]
>>> print, input[*,keep[sort(keep)]]
>>> 1 10 9 100 200
>>> 2 11 8 101 201
>>> 4 7 12 99 199
>
>>> Cheers
>
>>> Guillermo
>
>> You can find ord at:
>
>> http://web.astroconst.org/jbiu/jbiu-doc/math/ord.html
>
>> -Jeremy.
>
> Jeremy,
>
> Thanks for your kind and prompt help!
> It took my own routine 18 hours to do the job. I have just plug the
> codes you kindly offered into my codes, I'll let you know how
> efficient your routine is. Thanks!
>
> Bo

Hi Jeremy,

Your code helps me save 7 hours! That's a lot. Thanks!

Bo

Subject: Re: remove duplicate elements from a multi-dimensional array efficiently in IDL
Posted by Jeremy Bailin on Mon, 04 May 2009 18:15:33 GMT

On May 4, 11:12 am, "chenb...@gmail.com" <chenb...@gmail.com> wrote:
>> Jeremy,
>
>> Thanks for your kind and prompt help!
>> It took my own routine 18 hours to do the job. I have just plug the
>> codes you kindly offered into my codes, I'll let you know how
>> efficient your routine is. Thanks!
>
>> Bo
>
> Hi Jeremy,
>
> Your code helps me save 7 hours! That's a lot. Thanks!
>
> Bo

No problem! Glad it helped.

-Jeremy.

Subject: Re: remove duplicate elements from a multi-dimensional array efficiently in IDL
Posted by vino on Wed, 06 May 2009 11:52:59 GMT

Hello Jeremy,

Thanks for your idea here... just when i was thinking how to do this,
i found this post....

Thanks to the OP as well!!

Regards,

Vino

On May 4, 7:15 pm, Jeremy Bailin <astroco...@gmail.com> wrote:
> On May 4, 11:12 am, "chenb...@gmail.com" <chenb...@gmail.com> wrote:
>
>>> Jeremy,
>
>>> Thanks for your kind and prompt help!
>>> It took my own routine 18 hours to do the job. I have just plug the

>>> codes you kindly offered into my codes, I'll let you know how
>>> efficient your routine is. Thanks!
>
>>> Bo
>
>> Hi Jeremy,
>
>> Your code helps me save 7 hours! That's a lot. Thanks!
>
>> Bo
>
> No problem! Glad it helped.
>
> -Jeremy.
