Subject: set difference, with duplicates
Posted by Russell[1] on Wed, 10 Feb 2016 20:03:33 GMT
View Forum Message <> Reply to Message

Hi everyone...

I have two long-integer arrays, one of which contains duplicate entries, and I'd like to find the elements in a but not in b. For example:

a=[2,3,4,5,2,3,5,2,4,3,10,100]
b=[2,10]

I'd like set operator that computes a dif b, but preserves the duplicate entries (and order if possible) for a. For example,

c = set_difference(a,b)

and I would want

c=[3,4,5,3,5,4,3,100]

I'm aware of the Coyote's cgsetdifference, but that does not preserve duplicates (or I didn't realize the right set of options).

Any ideas? If it helps, the a-array may be very long 10^5 elements and the b array will be 10^3 elements. I also expect the values to be very high, but I can compress them to the lowest possible integer arrays.

Thanks for any advice,
Russell

Subject: Re: set difference, with duplicates
Posted by Burch on Wed, 10 Feb 2016 20:23:02 GMT
View Forum Message <> Reply to Message

On Wednesday, February 10, 2016 at 2:03:36 PM UTC-6, rrya...@gmail.com wrote:
> Hi everyone...
>
> I have two long-integer arrays, one of which contains duplicate entries, and I'd like to find the elements in a but not in b. For example:
>
> a=[2,3,4,5,2,3,5,2,4,3,10,100]
> b=[2,10]
>
>
> I'd like set operator that computes a dif b, but preserves the duplicate entries (and order if

possible) for a. For example,
>
> c = set_difference(a,b)
>
> and I would want
>
> c=[3,4,5,3,5,4,3,100]
>
> I'm aware of the Coyote's cgsetdifference, but that does not preserve duplicates (or I didn't realize the right set of options).
>
> Any ideas? If it helps, the a-array may be very long 10^5 elements and the b array will be 10^3 elements. I also expect the values to be very high, but I can compress them to the lowest possible integer arrays.
>
> Thanks for any advice,
> Russell

One option is to use match2 from the IDL Astronomy Library:
  http://idlastro.gsfc.nasa.gov/ftp/pro/misc/match2.pro
  http://idlastro.gsfc.nasa.gov

For your example:

```
IDL> a = [2,3,4,5,2,3,5,2,4,3,10,100]
IDL> b = [2,10]
IDL> match2, a, b, a_in_b, b_in_a
```

Note that match 2 finds matching elements and returns -1 for elements with no match.

```
IDL> print, a_in_b
         0        -1        -1        -1         0        -1        -1         0        -1        -1         1
-1
IDL> c = a[where(a_in_b eq -1)]
IDL> print, c
      3       4       5       3       5       4       3     100
```

-Jeff