ctools - Support #2548

ctselect from python

07/12/2018 10:04 AM - Sadeh Iftach

Status:	Closed	**Start date:**	07/12/2018
Priority:	Normal	**Due date:**	
Assigned To:	Knödlseder Jürgen	**% Done:**	100%
Category:		**Estimated time:**	0.00 hour
Target version:			

Description

Dear Supporter,

I am using ctselect inside a loop in python, in order to compare different selection cuts on my events.
The code is something like:

```
for n in range(100):
    cut = foo(n)

    ctselect = ctools.ctselect()
    ...
    ctselect['inobs'] = input_file
    ctselect['expr'] = cut
    ...
    ctselect.run()
    events = ctselect.obs()
    ...
```

On some machines, in some runs of the code I get an error:

```
  File "/batch/25115154.1.std.q/test.py", line 288, in runCtlike
    ctlike = ctools.ctlike(events)
  File "/afs/ifh.de/group/cta/scratch/sadeh/software/ctools/ctools/lib/python3.6/site-packages/ctools/tools.py", line 1802, in __init__
    this = _tools.new_ctlike(*args)
ValueError: *** ERROR in GApplicationPars::operator[](std::string&): Invalid argument. Parameter "chatter" has not been found in
 parameter file. Please specify a valid parameter name.
```

I suspect that this is a memory issue - the ctools.ctselect() object is not deleted fast enough by the garbage collector as the loop iterates.

My questions:
1. Is this bug really memory running out?
2. Is there a way to only have a single ctools.ctselect(), but to somehow reset it in the beginning of the loop?
3. What is the recommended way to immediately release the memory taken by ctools.ctselect() in python?

Thanks for the advice,
Iftach.

History

#1 - 07/12/2018 10:15 AM - Tibaldo Luigi

Hi Iftach,
can you please specify the ctools version and OS you are using?

Why are you attributing the error to ctselect? The message

```
this = _tools.new_ctlike(*args)
ValueError: *** ERROR in GApplicationPars::operator[](std::string&): Invalid argument. Parameter "chatter" has not been found in parameter file.
 Please specify a valid parameter name.
```

makes me rather think that there is a problem in parsing the ctlike par file. Could you please post the whole block of code setting up and running ctselect and ctlike?

#2 - 07/12/2018 10:37 AM - Sadeh Iftach

Dear Luigi,

Sorry about the confusion. In previous tests (with the same code), I got the error

```
  File "/afs/ifh.de/group/cta/scratch/sadeh/software/ctools/ctools/lib/python3.6/site-packages/ctools/tools.py", line 1913, in __init__
    this = _tools.new_ctselect(*args)
ValueError: *** ERROR in GApplicationPars::operator[](std::string&): Invalid argument. Parameter "chatter" has not been found in parameter file.
 Please specify a valid parameter name.
```

but it's the same issue.

Here is the complete function (It's part of a larger class, which defines all the input parameters, but you get the deal):

```
import gammalib
import ctools

def runCtlike(self):
    for n in range(100):
        # get a string, eg 'ENERGY > 0.11'
        cut = foo(n)

        ctselect = ctools.ctselect()
        ctselect['inobs'] = self.conf['evtFileName']
        ctselect["rad"]  = self.conf['rad']
        ctselect["ra"]  = self.conf['ra']
        ctselect["dec"] = self.conf['dec']
        ctselect['tmin'] = self.conf['tmin']
        ctselect['tmax'] = self.conf['tmax']
        ctselect['emin'] = self.conf['e_min']
        ctselect['emax'] = self.conf['e_max']
        ctselect['expr'] = cut

        ctselect.run()
        events = ctselect.obs()

        ctlike = ctools.ctlike(events)
        ctlike['caldb'] = self.conf['caldb']
        ctlike['irf'] = self.conf['irf']
        ctlike['inmodel'] = self.conf['modelFit']
        ctlike['outmodel'] = self.conf['resultFileName']
        ctlike['logfile'] = self.conf['ctlikeLog']
        ctlike['debug'] = self.conf['debug']

        ctlike.execute()

        for model in ctlike.obs().models():
            if model.name() != self.conf['modelFitName']:
                continue

            ts = model.ts()
            # do something with ts ........
    return
```

The code is running on the batch farm at DESY:

```
$ uname -a
Linux wgs18.zeuthen.desy.de 2.6.32-696.20.1.el6.x86_64 #1 SMP Thu Jan 25 08:47:49 CST 2018 x86_64 x86_64 x86_64 GNU/Linux
```

I'm using python 3.6.4 (anaconda-5.1) and the dev versions of gammalib and ctools (repo cloned on July 6th).

#3 - 07/12/2018 10:41 AM - Sadeh Iftach

PS -
I not sure how the par file would come into this.
The loop works file for several iterations (does the fit etc.), and only crashes after a while.
I never directly manipulate the par file, but I don't know what happens behind the scenes.

#4 - 07/12/2018 10:55 AM - Tibaldo Luigi

- Assigned To set to Knödlseder Jürgen

The par file is manipulated behind the scenes every time you create an instance of a tool. In either case, ctselect or ctlike, the error you reported seems to be related to this step. The handling of par files was changed recently by Jürgen (#2513) to avoid issues arising when the structure was changed and old versions were trailing in the user home directory. We will need to understand whether your problem may be related to this.

On a side note: your ctselect configuration may not produce the intended results. Recently we modified the default behavior to select around the pointing direction of the observation (#2501). If you want to select around a given (ra, dec) direction you need to modify your code as follows

```
ctselect["usepnt"] = False
ctselect["rad"]  = self.conf['rad']
ctselect["ra"]  = self.conf['ra']
ctselect["dec"] = self.conf['dec']
```

#5 - 07/12/2018 11:15 AM - Sadeh Iftach

Thanks, Luigi.

I have been doing energy/time selection so far, so I don't think I have a problem with the pointing, but I've added you line to the code.

The ticket you referenced confuse me a bit.
I want to run this code on a batch farm, with multiple jobs running in parallel on different output directories. I assume this would imply a possible racing condition with modifying the par file.
How do I define (from the python interface) the directory in which I want the par file for a given job to be stored?

#6 - 07/12/2018 11:34 AM - Tibaldo Luigi

If you do not intend to do any spatial selection your code block should be replaced by

```
ctselect["rad"] = 'INDEF'
```

I'm not so familiar with the handling of par files, but I know there is a protection mechanism in place to avoid i/o by multiple processes. A quick and dirty way could be to redefine the environment variable $HOME for each process to point to a different directory, but I'm not sure it would solve your problem and therefore I think it's best to let Jürgen help you through this.

#7 - 07/12/2018 11:37 AM - Sadeh Iftach

Great.
Thanks again!

#8 - 07/12/2018 09:56 PM - Knödlseder Jürgen

I remember that there was an issue on the DESY batch farm with the file system and par files, Michael Maier run into this problem.

The issue is that there is a file locking mechanism when reading and writing into par files to make sure that no concurrent process is manipulating the par file. Now on the DESY batch farm this locking mechanism does not seem to work.

Nevertheless, in your code I see no parallel processing, so I'm wondering why there should be an issue. My guess is that in your case the ctselect or ctlike par file is empty, or at least the chatter parameter is missing.

Which ctools version are your using? My guess would be that with the development branch your problem will go away, since we changed the way how par files are handled.

#9 - 07/12/2018 10:22 PM - Sadeh Iftach

Hey Jürgen,

I'm using the dev version (as of July 6th, but I can update again).
I don't do parallel processing per job, but I am running multiple jobs at once.

Is there a way to set the par file to a local copy/directory for a particular job?
Thanks.

#10 - 07/13/2018 12:00 AM - Sadeh Iftach

I think I figured it out...
I've added the following to the top of my python routine:

```
new_PFILES_path = 'new_PFILES_path/'
```

```
os.makedirs(new_PFILES_path, exist_ok = True)
os.environ['PFILES'] = new_PFILES_path + ';' + os.environ['CTOOLS'] + '/syspfiles'
```

where new_PFILES_path is set to the tmp working directory of each job.

I'll run a few tests and close this issue if all is well.
Thank you both for your help.

#11 - 07/16/2018 02:27 PM - Sadeh Iftach

- Status changed from New to Closed

- % Done changed from 0 to 100