cuckoo_hash

configuration of machine

> CPU: i7-12700
> GPU: GTX-1660s

Algorithm

> On the base of original cuckoo hash, I separate the table into funcutin_num * 2 parts and each function only take charge of two of these part. In this way, we can efficiently avoid hash conflict and insert elements fast.
>
> I make insert and search part into cuda version. In insert, I add share variable need_rehash_share to terminate kernal when some value's evict chain has surpassed the evict_bound. In search, I terminate the kernal as long as it find the pos so that it will perform well if most to-find elements are in the range of first hash function.
>
> I use cpp's standard random creater so I don't need seed.

Benchmark

> Each benchmark of insert and search record the time before and after their class function is called. Insert function will return the rehash-times. So the speed is computed by (data_size * rehash_times / used_time). Because search won't cause rehash, the speed is computed by (data_size / used_time).

Improvement and problems

> There is no direct improve on insert and search function instead of the mode I use.

> I make some improvement on search data if most data are not in the hash table. In order to fast get the answer of each hash function, I launch multiple kernals at the same time to find them. It acutually speed up the situation I hope to improve, but the performance of the situation that most to-find are in the hash table. As I don't know the way to communicate between kernal functions.
>
> Another problem is about the benchmark. As I rehash more, the speed it gets is slowly down greatly. I don't know it's the problem of my algrithm or the benchmark method.

The answer to the task4

> The better evict bound is when the ratio is about 5 or 6. Because it ensures enough space to insert elements and limits the infinit loop.

Compile and run

> There's a compile.sh which should be a compile file to get task1/2/3/4. And a run.sh can automatically run the experiments task.
>
> The basic input director is done instead of seed because of what i have said.
>
> Besides basic input director, I add --function_num to specify the number of hash function the table use. And --file-in to specify user-specified file, which may have some display error which is a little hard to fix.