

Trabajo final Topics Especiales en Telemática

Sebastián García Valencia

Simon escobar

Pablo restrepo

Universidad EAFIT

Medellín, Colombia, Suramérica

Resumen

El presente documento contiene el análisis y diseño de un programa paralelo para recomendación de películas.

Palabras clave.

Mpi, c, sistema de recomendación, programación paralela

1. Introducción

Los problemas actualmente, ya sean científicos o comerciales, cada vez necesitan más procesamiento, ya sea en datos en instrucciones o en ambos. Por esto la programación paralela es cada vez más importante, en el presente programa se pretende usar mpi para solucionar un problema de programación intensiva en procesamiento sobre??????????

2. Algoritmo secuencial

Antes de iniciar se debe tener un punto de partida, al hacer este algoritmo secuencialmente para un conjunto de datos de 1000 en las matrices se obtuvo un tiempo de 21,059443 segundos.

3. PCAM (aquí falta todo)

3.1 Particionamiento

A cada tarea se enviara primero toda la matriz de destino, y luego por medio de un round robin se les ira rotando cada macrobloque de la matriz original, cada tarea procesara y devolverá el índice y mejor match de la determinada posición de la matriz de referencias.

3.2 Comunicación

Se tendrá una tarea master encargada de llenar las matrices y esta le enviara a los n trabajadores los datos, el control se dara enviando primero el índice de la matriz de referencias que se esta trabajando, la comunicación será solo de 2 niveles y se dara

Master a slaves: 1 a n

Slaves a master: 1 a 1

3.3 Aglomeracion

Por el momento no se tendrán submaster que distribuyan los datos

3.4 Mapeo

Se tienen 4 nodos, uno será el master (para este caso no se pondrá a procesar) y en los otros 3 se tendrán los n trabajadores repartidos, se dejara que mpi gestione los procesos multiples en cada nodo y por ahora será transparente para nosotros.

4. Análisis de tiempos.

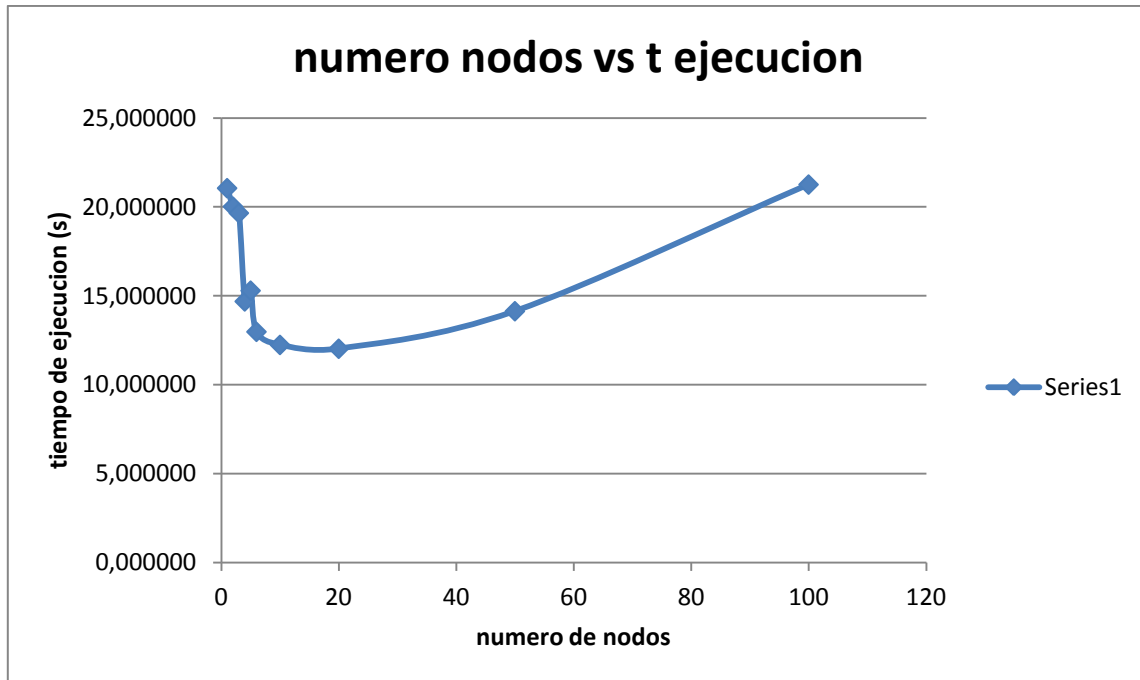
Se tomo como base matrices de 1000000 posiciones, 1000 usuarios por 1000 películas

```
#define NUM_USERS 1000
```

```
#define NUM_MOVIES 1000
```

```
#define m 5
```

num nodos	t ejecucion
1	21,059443
2	20,023174
3	19,658452
4	14,690265
5	15,294607
6	12,979850
10	12,242615
20	12,027465
50	14,138212
100	21,261076



```
mpi@cluster:~/programs/gruposps
Processing time: 14.690265
[mpi@cluster gruposps]$ ^C
[mpi@cluster gruposps]$ mpirun -np 3 sisse
Matriz tomada desde el fichero de texto:
Matriz de Correlacion:
Matriz de recomendación:
recomendaciones para el usuario 3:
2
16
4
4
13
Processing time: 15.294607
[mpi@cluster gruposps]$ mpirun -np 5 sisse
Matriz tomada desde el fichero de texto:
Matriz de Correlacion:
Matriz de recomendación:
recomendaciones para el usuario 3:
2
16
4
4
13
Processing time: 15.623455
[mpi@cluster gruposps]$ mpirun -np 4 sisse
Matriz tomada desde el fichero de texto:
Matriz de Correlacion:
Matriz de recomendación:
recomendaciones para el usuario 3:
2
16
4
4
13
Processing time: 14.565765
[mpi@cluster gruposps]$ ^C
[mpi@cluster gruposps]$ mpirun -np 6 sisse
Matriz tomada desde el fichero de texto:
Matriz de Correlacion:
Matriz de recomendación:
recomendaciones para el usuario 3:
2
```

```

mpi@cluster:~/programs/gruposps
Matriz tomada desde el fichero de texto:
Matriz de Correlacion:
Matriz de recomendación:
recomendaciones para el usuario 3:
2
16
4
4
13
Processing time: 12.979850
[mpi@cluster gruposps]$ ^C
[mpi@cluster gruposps]$ mpirun -np 10 sierre
Matriz tomada desde el fichero de texto:
Matriz de Correlacion:
Matriz de recomendación:
recomendaciones para el usuario 3:
2
16
4
4
13
Processing time: 12.242615
[mpi@cluster gruposps]$ ^C
[mpi@cluster gruposps]$ mpirun -np 20 sierre
Matriz tomada desde el fichero de texto:
Matriz de Correlacion:
Matriz de recomendación:
recomendaciones para el usuario 3:
2
16
4
4
13
Processing time: 12.027465
[mpi@cluster gruposps]$ ^C
[mpi@cluster gruposps]$ mpirun -np 50 sierre
Matriz tomada desde el fichero de texto:
Matriz de Correlacion:
Matriz de recomendación:
recomendaciones para el usuario 3:
2
16

```

```

mpi@cluster:~/programs/gruposps
13
Processing time: 14.138212
[mpi@cluster gruposps]$ ^C
[mpi@cluster gruposps]$ mpirun -np 50 sierre
Matriz tomada desde el fichero de texto:
Matriz de Correlacion:
Matriz de recomendación:
recomendaciones para el usuario 3:
2
16
4
4
13
Processing time: 13.996180
[mpi@cluster gruposps]$ mpirun -np 100 sierre
Matriz tomada desde el fichero de texto:
Matriz de Correlacion:
Matriz de recomendación:
recomendaciones para el usuario 3:
2
16
4
4
13
Processing time: 21.261076
[mpi@cluster gruposps]$ ^C
[mpi@cluster gruposps]$ mpirun -np 2 sierre
Matriz tomada desde el fichero de texto:
Matriz de Correlacion:
Matriz de recomendación:
recomendaciones para el usuario 3:
2
16
4
4
13
Processing time: 20.871079
[mpi@cluster gruposps]$ mpirun -np 3 sierre
Matriz tomada desde el fichero de texto:
Matriz de Correlacion:
Matriz de recomendación:
recomendaciones para el usuario 3:

```

```

mpi@cluster:~/programs/gruposps
recomendaciones para el usuario 3:
2
16
4
4
13
Processing time: 19.658452
[mpi@cluster gruposps]$ ^C
[mpi@cluster gruposps]$ ^C
[mpi@cluster gruposps]$ mpirun -np 4 sistr
Matriz tomada desde el fichero de texto:
Matriz de Correlacion:
Matriz de recomendación:
recomendaciones para el usuario 3:
2
16
4
4
13
Processing time: 14.321200
[mpi@cluster gruposps]$ mpirun -np 6 sistr
Matriz tomada desde el fichero de texto:
Matriz de Correlacion:
Matriz de recomendación:
recomendaciones para el usuario 3:
2
16
4
4
13
Processing time: 13.031520
[mpi@cluster gruposps]$ mpirun -np 5 sistr
Matriz tomada desde el fichero de texto:
Matriz de Correlacion:
Matriz de recomendación:
recomendaciones para el usuario 3:
2
16
4
4
13
Processing time: 15.320354

```

5. Funcionamiento

Para facilidad de la ejecución se tiene un makefile

Para compilar, una ves ubicado en la carpeta del proyecto se debe usar el comando

`$make`

El comando run viene por defecto para ejecutar en 4 nodos, para esto se da

`$make run`

Para ejecutar en otro numero de nodos se usa

`$mpirun -np <numNodos> vectoresMov`

makefile

```

mpi@cluster:~/programs/gruposps
$make
mpiflags=
all:
    mpicc sistrRecom.c -lm -o sistr

compSec:
    mpicc sistrRecomSec.c -lm -o sistr

run:
    mpirun -np 4 sistr

clean:
    rm -f *.o *.C~ Makefile~ sistr

```

ejecucion

```

$ mpicc -xcluster -lmp -xgroups
mpicc sistRecom.c -lm -o sistre
[mpr@cluster gruposps]$ make run
mpirun -np 4 sistre
Matriz tomada desde el fichero de texto:
Matriz de Correlacion:
1.000000 0.020515 0.000000 -0.471779 -0.103729 -0.059601 0.272560 -0.572172 0.087
039 0.106711
0.020515 1.000000 0.124226 -0.583723 -0.541736 0.553372 -0.166667 0.352148 -0.766
032 -0.315244
0.000000 0.124226 1.000000 -0.295527 -0.089730 0.000000 -0.212959 0.440959 -0.421
637 0.058742
-0.471779 -0.583723 -0.295527 1.000000 0.326168 -0.223978 -0.084962 -0.279247 0.5
3396 0.140615
-0.103729 -0.541736 -0.089730 0.326168 1.000000 -0.008327 0.670721 -0.305234 0.25
8377 0.692605
-0.059601 0.553372 0.000000 -0.223978 -0.008327 1.000000 -0.009882 0.233842 -0.44
0204 -0.310734
0.272560 -0.166667 -0.212959 -0.084962 0.670721 -0.009882 1.000000 -0.543313 0.30
3046 0.484125
-0.572172 0.352148 0.440959 -0.279247 -0.305234 0.233842 -0.543313 1.000000 -0.59
7614 -0.199822
0.087039 -0.766032 -0.421637 0.513996 0.255377 -0.440204 0.303046 -0.597614 1.00
000 -0.111456
0.106711 -0.315244 0.058742 0.140615 0.692605 -0.310734 0.484125 -0.199822 -0.11
456 1.000000
Matriz de recomendación:
6 9 8 1 2
5 7 2 0 6
7 1 9 0 5
4 9 6 5
9 6 9 8 5
3 7 2 4 6
4 9 8 0 5
2 1 5 9 3
3 6 4 0 9
4 6 3 0 2
recomendaciones para el usuario 3:
0
0
0
0
4
Processing time: 0.073309

```

6. Posibles mejoras

- En el momento el master no hace ejecución, dado que este es un nodo completo podría aprovecharse también

??????????????????????”

7. Enlaces

Repositorio GIT: <https://bitbucket.org/prestrepoh/parallel-pearsons-recommendation-algorithm>

8. Bibliografía

??????????????????????'

<http://hpcc.usc.edu/support/documentation/examples-of-mpi-programs/>