MASSIVE DATA PROCESSING TECHNIQUES

IN DATA SCIENCE (CS504048)

**Assignment: Recommendation System**

Trần Quốc Lĩnh – 51703124

*Khoa Công nghệ thông tin*

*Trường Đại học Tôn Đức Thắng*

**1a/**

Giả sử ta có

Mà,

Biểu diễn T dưới dạng đồ thị hai phía, ta được:

Qua đó, ta thấy

* Tii : Độ tương tự của user i với user i.
* Tij : Độ tương tự của user i với user j (i≠j).

**2b/**

Ta có biểu thức tính độ tương tự - cos similarity:

Với và áp dụng công thức trên ta được:

Item Similarity Matrix:

|  |  |  |  |
| --- | --- | --- | --- |
|  | I0 | I1 | I2 |
| I0 | 1 | 0,58 | 0,67 |
| I1 | 0,58 | 1 | 0,58 |
| I2 | 0,67 | 0,58 | 1 |

User Similarity Matrix:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | U0 | U1 | U2 | U3 |
| U0 | 1 | 0,71 | 0,82 | 0,71 |
| U1 | 0,71 | 1 | 0,58 | 0 |
| U2 | 0,82 | 0,58 | 1 | 0,58 |
| U3 | 0,71 | 0 | 0,58 | 1 |

Mặt khác:

Vậy và

**3c/**

**Similarity matrix**

Với

Dựa trên user-user:

Áp dụng công thức

|  |  |  |  |
| --- | --- | --- | --- |
|  | I0 | I1 | I2 |
| U0 | 1 | 0.82 | 1 |
| U1 | 1 | 0,58 | 1,29 |
| U2 | 1 | 1 | 1 |
| U3 | 1,29 | 0,58 | 1 |

Dựa trên item-item:

Áp dụng công thức

|  |  |  |  |
| --- | --- | --- | --- |
|  | I0 | I1 | I2 |
| U0 | 1 | 1,06 | 1 |
| U1 | 1 | 0,58 | 0,67 |
| U2 | 1 | 1 | 1 |
| U3 | 0,67 | 0,58 | 1 |

**4d/**

LIMIT\_USER = 1000

LIMIT\_MOVIE = 200

User-User:

MLB Spring Training

Pirates of the Caribbean: At World's End

Your World with Neil Cavuto

Transformers

Made of Honor

Movie-Movie:

Zoey 101

The Parent Trap

Presidential Press Conference

Blue's Clues

The Colbert Report