

ĐẠI HỌC QUỐC GIA TP.HCM
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN



Môn học: KHAI PHÁ DỮ LIỆU TRONG DOANH NGHIỆP

LỚP: DS317.P11

THỰC NGHIỆM

GVHD: ThS. Nguyễn Thị Anh Thư

Nhóm sinh viên thực hiện:

Nguyễn Hữu Nam	MSSV: 22520917
Nguyễn Khánh	MSSV: 22520641
Võ Đình Khánh	MSSV: 22520659
Nguyễn Minh Sơn	MSSV: 22521254
Bùi Hồng Sơn	MSSV: 22521246



Mục lục

1	Content-based Filtering	2
1.1	Feature Engineering	2
1.2	Training	2
1.3	Evaluation	2
2	Bayesian Personalized Ranking with Matrix Factorization	2
2.1	Feature Engineering	2
2.2	Training	3
2.3	Evaluation	3
3	Content-based Filtering	3
3.1	Feature Engineering	3
3.2	Training	3
3.3	Evaluation	3
4	Neural Factilization Machine (NFM)	3
4.1	Feature Engineering	3
4.2	Training	3
4.3	Evaluation	4
5	KGAT	4
5.1	Feature Engineering	4
5.2	Training	4
5.3	Evaluation	4



1. Content-based Filtering

1.1. Feature Engineering

Feature 1:

- Sử dụng trường name, about, field
- Vectorize các trường
- Tính độ tương đồng giữa các khóa học bằng cosine

Feature 2:

- Sử dụng thêm trường school, concept được nối từ các relations
- Vectorize các trường
- Tính độ tương đồng giữa các khóa học bằng cosine

1.2. Training

in progress...

1.3. Evaluation

in progress...

2. Bayesian Personalized Ranking with Matrix Factorization

2.1. Feature Engineering

Sử dụng các cặp user_id và course_id để tạo ra 1 bảng chứa các tương tác giữa người dùng và khóa học như sau:

user_id	course_id
U_24	C_55110
U_24	C_55231
...	...



2.2. Training

Đối với mỗi người dùng, loại bỏ khóa học cuối cùng mà người đó đã học, sau đó với mỗi khóa học chúng ta cặp theo một khóa chưa học hay còn gọi là negative sampling

2.3. Evaluation

Với các khóa học được để lại ở bước trước, cặp chúng với 99 khóa học khác mà người dùng đó chưa học, sau đó dùng mô hình đánh giá các khóa học trên rồi tính độ đo Recall@10 và NDCG@10

3. Content-based Filtering

3.1. Feature Engineering

3.2. Training

3.3. Evaluation

4. Neural Factorization Machine (NFM)

4.1. Feature Engineering

Trong phần này, nhóm sẽ tìm các bảng và thuộc tính có thể sử dụng để tạo ra các feature cho mô hình Neural Factorization Machine (NFM).

Các bảng được chọn bao gồm: **'user.json'**: - Sử dụng trường id, name, gender, school, course_order - Vectorize các trường

4.2. Training

Sử dụng các feature đã được tạo ra từ bước feature engineering, chúng em tiến hành chạy lại thuật toán Neural Factorization Machine để theo dõi kết quả.

Bảng kết quả:



Models	Feature Engineering	Score					
		Precision		Recal		NDCG	
NFM	last year feature + simple	0.1805	0.0616	0.1805	0.6155	0.1805	0.3727

4.3. Evaluation

- Started Evaluation and testing again feature engineering and data preprocessing from beginning.

5. KGAT

SAU MOT HOI SUA DI SUA LAI THI T QD DUNG CUA NAM
NGOAI :V

5.1. Feature Engineering

Feature 1: CHON GIONG NAM NGOAI NHUNG BO FIELD CON-
CEPT - CHAY THU DUOC 300K DUA DA ROI TINH

5.2. Training

5.3. Evaluation