

**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**KHOA KHOA HỌC MÁY TÍNH**

**A blue logo with a black background

Description automatically generated with low confidence**

**BÀI TẬP CHƯƠNG 4**

Giảng Viên: Nguyễn Thị Anh Thư

Mã Lớp: CS313.O21

Nhóm 04

| *Trương Văn Khải* | *21520274* |
| --- | --- |
| *Đoàn Nhật Sang* | *21522542* |
| *Lê Ngô Minh Đức* | *21520195* |
| *Lê Minh Quang* | *21522510* |
| *Lê Yến Nhi* | *21522427* |
| *Hoàng Thị Mỹ Hạnh* | *21522044* |
| *Hoàng Tiến Đạt* | *21520696* |

**TP. HỒ CHÍ MINH, 3/2024**

**NHẬN XÉT CỦA GIẢNG VIÊN HƯỚNG DẪN**

........................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................ ............................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................................ .................................................................................................................................... .................................................................................................................................... .................................................................................................................................... .................................................................................................................................... ...................................................................................................................................

**BẢNG PHÂN CÔNG CÔNG VIỆC**

| **Chi tiết công việc** | | **Trương Văn Khải (NT)** | **Đoàn Nhật Sang** | **Lê Ngô Minh Đức** | **Lê Minh Quang** | **Lê Yến Nhi** | **Hoàng Thị Mỹ Hạnh** | **Hoàng Tiến Đạt** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Phân công** | |  |  |  |  |  |  |  |
| **Mục 1, 2, 3** | |  |  |  |  |  |  |  |
| **Mục 4** | |  |  |  |  |  |  |  |
| **Mục 5** | |  |  |  |  |  |  |  |
| **Mục 6 - Nội dung 1** | |  |  |  |  |  |  |  |
| **Mục 6 - Nội dung 2** | |  |  |  |  |  |  |  |
| **Mục 6 - Nội dung 3** | |  |  |  |  |  |  |  |
| **Mục 7, 8** | |  |  |  |  |  |  |  |
| **Format docs** | |  |  |  |  |  |  |  |
| **Kiểm tra** | |  |  |  |  |  |  |  |
| **Mức độ hoàn thành** | | **100%** | **100%** | **100%** | **100%** | **100%** | **100%** | **100%** |

Bảng 0.1. Phân công công việc

**MỤC LỤC**

[1. Tên đề tài, thời gian thực hiện, tổng kinh phí. 1](#_1fob9te)

[2. Nhóm thực hiện: chủ nhiệm, nhân lực. 1](#_1fob9te)

[3. Mô tả đề tài: giới thiệu, ứng dụng, các dự án liên quan trong cùng lĩnh vực. 1](#_1fob9te)

[4. Tổng quan: ý tưởng, tính cấp thiết, tính mới. 2](#_v2sabddwptt7)

[5. Input-ouput: 3](#_ma6tqogrv586)

[6. Mục tiêu đề tài. 5](#_1fob9te)

[6.1. Nội dung 1 5](#_1fob9te)

[6.2. Nội dung 2. 5](#_1fob9te)

[6.3. Nội dung 3. 6](#_o5kzi7b93yax)

[7. Kết quả dự kiến, sản phẩm đề tài. 6](#_1fob9te)

[8. Tài liệu tham khảo. 6](#_l22de3xezqio)

# **1. Tên đề tài, thời gian thực hiện, tổng kinh phí.**

* Tên đề tài: **HỆ THỐNG KHUYẾN NGHỊ KHÓA HỌC CHO BỘ DỮ LIỆU MOOCCUBEX**.
* Thời gian thực hiện: 8 tuần.
* Tổng kinh phí: Chưa biết.

# **2. Nhóm thực hiện: chủ nhiệm, nhân lực.**

* Chủ nhiệm: **Trương Văn Khải - 21520274**
* Thành viên tham gia:  **Đoàn Nhật Sang - 21522542**

**Lê Ngô Minh Đức - 21520195**

**Lê Yến Nhi - 21522427**

**Lê Minh Quang - 21522510**

**Hoàng Thị Mỹ Hạnh - 21522044**

**Hoàng Tiến Đạt - 21520696**

* Giảng viên hướng dẫn: **ThS**. Nguyễn Thị Anh Thư

# **3. Mô tả đề tài: giới thiệu, ứng dụng, các dự án liên quan trong cùng lĩnh vực.**

* **Giới thiệu**:

Các nền tảng học tập trực tuyến (MOOCs - Massive Open Online Courses) ngày nay càng dần được phổ biến, được xây dựng với quy mô ngày càng lớn và đúng một vai trò quan trong trong giáo dục cho học sinh. Để duy trì được sự quan tâm của người dùng đối với MOOCs, các hệ thống khuyến nghị đã được nghiên cứu và triển khai để khuyến nghị các khóa học, video hoặc các khái niệm trong bài học mà một người dùng có thể quan tâm.

Ở bài này, chúng tôi sẽ tập trung vào nhiệm vụ khuyến nghị các khóa học mà người dùng có thể quan tâm, nhằm giảm được phần nào đó tỷ lệ bỏ các khóa học đăng ký hoặc thúc đẩy hiệu quả hiệu suất học tập của người dùng.

* **Ứng dụng**:

Xây dựng hệ thống khuyến nghị các nguồn tài nguyên học tập. Chúng tôi thực hiện nhiệm vụ này trên MOOCCubeX để đánh giá khả năng của tập dữ liệu này và thảo luận về cách nó có thể được sử dụng để tiến hành các nghiên cứu liên quan hoặc mở rộng trên các bộ dữ liệu MOOCs khác.

Sau đó có thể sử dụng các bộ dữ liệu học tập của các trường Đại học ở Việt Nam để xây dựng các ứng dụng có liên đến “Cố vấn học tập thông minh tại các trường đại học” hay “Cố vấn học tập thông minh cho các nền tảng học tập trực tuyến”.

* **Các dự án liên quan**:

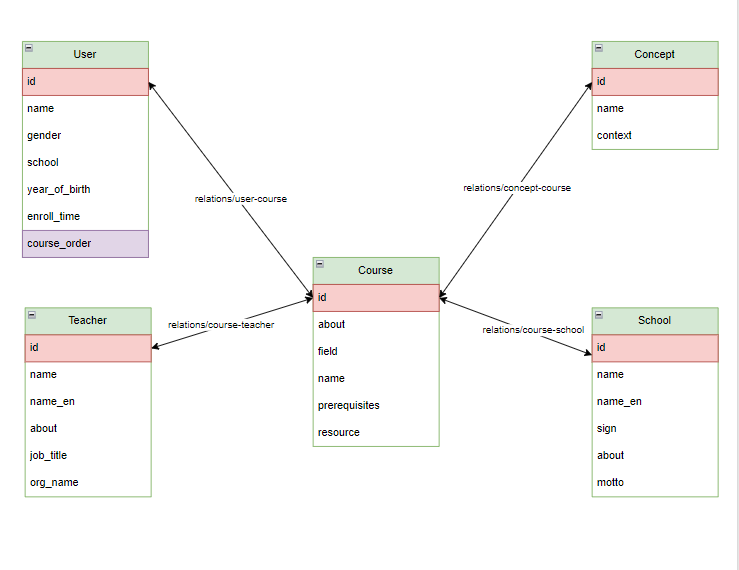
1. **KGAT: Knowledge Graph Attention Network for Recommendation**[1]: một phương pháp dựa trên GNN (Mạng Nơ-ron Đồ thị) sử dụng đồ thị tri thức nền (Knowledge Base) để cải thiện khuyến nghị. Phương pháp này được tái tạo dựa trên mạng co-occurrence network.
2. **Truelearn: A family of bayesian algorithms to match lifelong learners to open educational resources** [2]: một phương pháp khuyến nghị giáo dục sử dụng một họ của Mạng Bayesian.
3. **BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer** [3]: một mô hình khuyến nghị tiền huấn luyện điển hình dựa trên kiến trúc Transformer song hướng như BERT.
4. **KB4Rec: A Data Set for Linking Knowledge Bases with Recommender Systems** [4]: Đề xuất phương pháp tổ chức dữ liệu dưới dạng một Đồ thị tri thức (Knowledge Base - KB). Sau đó liên kết tập dữ liệu giữa hệ thống khuyến nghị (RS) với KB.

# **4. Tổng quan: ý tưởng, tính cấp thiết, tính mới.**

* **Ý tưởng**: Ý tưởng chính của nhóm là khuyến nghị các nguồn tài nguyên học tập (khóa học) cho người dùng. Chúng tôi định nghĩa nhiệm vụ này là khuyến nghị các khóa học tiếp theo cho học viên dựa trên chuỗi các khóa học lịch sử của học viên. Nhiệm vụ này không chỉ yêu cầu mô hình hóa hành vi của học viên một cách tốt nhất mà còn cần xem xét vai trò của kiến thức được kèm theo trong các khóa học, cấu trúc của các khóa học, v.v.
* **Tính cấp thiết**: Với sự phổ biến ngày càng nhiều của các nền tảng học trực tuyến, việc xây dựng một hệ thống khuyến nghị khóa học hiệu quả dựa trên một bộ dữ liệu lớn dần trở thành một nhiệm vụ quan trọng và cấp thiết. Đồng thời, MOOCCubeX [5] chứa đựng một lượng lớn thông tin về hành vi học tập, sở thích và nhu cầu của người dùng, cung cấp một nguồn dữ liệu quý giá để phân tích và hiểu biết sâu rộng về các yếu tố ảnh hưởng đến việc học trực tuyến. Vì vậy, việc sử dụng một mô hình học sâu và khai phá dữ liệu để xây dựng một hệ thống khuyến nghị cho MOOCCubeX [5] để đề xuất các khóa học phù hợp là cần thiết, giúp tăng cường trải nghiệm học tập và giúp người dùng đạt được kết quả học tập tốt hơn.
* **Tính mới:** Trong quá trình tìm hiểu, chúng tôi nhận thấy chưa có một dự án nghiên cứu nào thực hiện việc khuyến nghị các khóa học dựa trên nguồn tài nguyên MOOCCubeX [5]. Ngoài ra, đề xuất một phương pháp tiền xử lý dữ liệu, một mô hình học sâu để phục vụ nhiệm vụ khuyến nghị trên. Ngoài ra, xây dựng Website phục vụ mục đích sử dụng của người dùng.

# **5. Input-ouput:**

* *Input*: Chúng tôi sẽ sử dụng tập dữ liệu **entities/user.json,**  **entities/course.json, entities/concept.json,**  **entities/teacher.json,**  **entities/school.json** để xây dựng một đồ thị chung nhất cho KGAT. Cụ thể, giữa hai thực thể (hay còn gọi là entities, hai thực thể này phải có quan hệ khóa chính & khóa ngoại với nhau) bất kì ta sẽ có một quan hệ (ví dụ: *user -> học -> khóa học* và *khóa học -> được học bởi -> user*, vì đây là quan hệ hai chiều). Trong tập MOOCubeX mà nhóm sử dụng có quan hệ như sơ đồ:



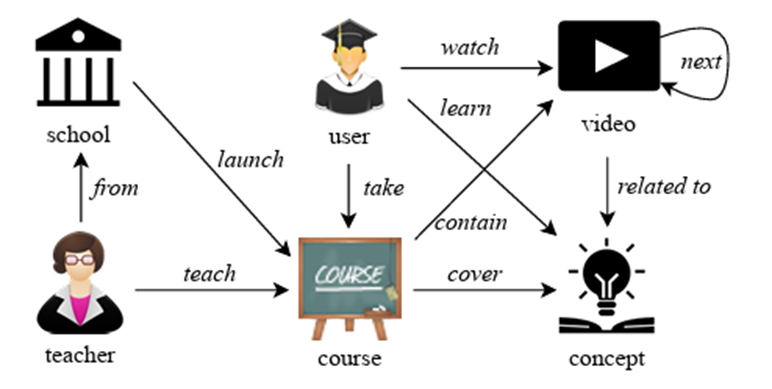
Các file **relations** là các file được cung cấp sẵn trong MOOCubeX để liên kết khóa chính & khóa ngoại của hai thực thể lại với nhau (có dạng id\_entity\_1 id\_entity\_2). Dễ thấy ta có 4 cặp relations, **suy ra đồ thị KGAT sẽ có tổng cộng 2\*4=8 relation khác nhau.**

* *Output*: Với mỗi user i có embedding **u\_i (lấy từ mô hình KGAT sau khi huấn luyện)**, ta sẽ có embedding của mỗi khóa học j là **c\_j (lấy từ mô hình KGAT sau khi huấn luyện)** thì ta sẽ tính **x = argmax(softmax(u\_i^T@c\_j))** (xác suất lớn nhất mà khóa học j được đề xuất cho user i) và đề xuất khóa học x cho user i (có thể lấy **top\_k** nếu cần, tức là lấy **top\_k** khóa học có xác suất lớn nhất).

# **6. Mục tiêu đề tài.**

## **6.1. Nội dung 1**

1. **Mục tiêu 1:** Xây dựng mô hình học sâu với bộ dữ liệu MOOCCubeX [5] với nghiệm vụ Khuyến nghị khóa học cho người dùng.
2. **Phương pháp 1**: Tập trung vào việc sử dụng mô hình KGAT[1] để dự đoán và khuyến nghị các khóa học có thể thu hút sự quan tâm của người dùng trên các nền tảng MOOCs. Điều này sẽ dựa trên lịch sử tương tác giữa người dùng và các khóa học (tức là, một người dùng đã tương tác với một khóa học nếu người dùng đã đăng ký khóa học đó). Kết quả trả về của mô hình sẽ là top-k khóa học được đề xuất cho người dùng.



Hình vẽ biểu diễn sự đa dạng các mối quan hệ trong bộ dữ liệu MOOCCubeX [5]. Trong đó, một user được coi là quan tâm đến một khóa học nếu user đã đăng ký hoặc có khả năng sẽ đăng ký khóa học đó trong tương lai.

1. **Sản phẩm 1:** pretrained mô hình KGAT được huấn luyện trên bộ dữ liệu MOOCCubeX [5] với nhiệm vụ Khuyến nghị khóa học.

## **6.2. Nội dung 2.**

1. **Mục tiêu 2:** Tìm hiểu các hệ thống lưu trữ dữ liệu phù hợp (Microsoft Azure), sau đó tiến hành triển khai lưu trữ bộ dữ liệu của nhóm trên nền tảng này.
2. **Phương pháp 2:** Tận dụng các dịch vụ mà Microsoft Azure cung cấp, cụ thể: **Storage services** (dịch vụ lưu trữ) và **Data management** (Quản lý dữ liệu).
3. **Sản phẩm 2:** Thu được một **Azure Tables** dùng để lưu trữ các dữ liệu dưới dạng SQL.

## **6.3. Nội dung 3.**

1. **Mục tiêu 3:** Xây dựng ứng dụng Website để phục vụ việc tương tác giữa người dùng và Hệ thống khuyến nghị.
2. **Phương pháp 3:** Sau khi thu được pretrained của mô hình KGAT [1] được huấn luyện trên bộ dữ liệu MOOCCubeX [5] trên nhiệm vụ Khuyến nghị các khóa học cho học viên. Sau đó sử dụng FastAPI (backend) và ReactJS (front-end) để triển khai website.
3. **Sản phẩm 3:** Thu được một ứng dụng Website trực quan, dễ dàng sử dụng, cho phép người dùng nhập vào các khóa học đã học, trả về tập gồm top-k các khóa học được hệ thống khuyến nghị.

# **7. Kết quả dự kiến, sản phẩm đề tài.**

* Phương pháp sử dụng một mạng GNN (Mạng Nơ-ron Đồ thị) có sử dụng đồ thị tri thức nền (Knowledge Base) để cải thiện hiệu suất khuyến nghị cho các khóa học được huấn luyện trên bộ dữ liệu MOOCCubeX [5].
* Lưu trữ thành công dữ liệu trên Microsoft Azure bằng việc sử dụng dịch vụ Azure Storage.
* Xây dựng thành công một Website khuyến nghị các khóa học cho người dùng với cơ sở là các khóa học từ bộ dữ liệu MOOCCubeX [5].

# **8. Tài liệu tham khảo.**

**[1]**: Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. 2019. Kgat: Knowledge graph attention network for recommendation. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 950–958.

**[2]**: Sahan Bulathwela, Maria Perez-Ortiz, Emine Yilmaz, and John Shawe-Taylor. 2020. Truelearn: A family of Bayesian algorithms to match lifelong learners to open educational resources. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34. 565–573.

**[3]**: Fei Sun, Jun Liu, Jian Wu, Changhua Pei, Xiao Lin, Wenwu Ou, and Peng Jiang. 2019. BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer. In Proceedings of the 28th ACM international conference on information and knowledge management. 1441–1450.

**[4]**: WayneXinZhao, Gaole He, Hong-Jian Dou, Jin Huang, Siqi Ouyang, and Ji-Rong Wen. 2018. KB4Rec: A Dataset for Linking Knowledge Bases with Recommender Systems. CoRR abs/1807.11141 (2018).

**[5]**: Jifan Yu, Yuquan Wang, Qingyang Zhong, Gan Luo, Yiming Mao, Kai Sun, Wen zheng Feng, Wei Xu, Shulin Cao, Kaisheng Zeng, et al. 2021. MOOCCubeX: a large knowledge-centered repository for adaptive learning in MOOCs. In Proceedings of the 30th ACM International Conference on Information & Knowledge Management. 4643–4652.