

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN

Đào Đức Anh - Nguyễn Thành Nhân

Xây dựng hệ thống gợi ý sản phẩm
dựa trên mô hình Auto-Encoder

KHÓA LUẬN TỐT NGHIỆP CỬ NHÂN
CHƯƠNG TRÌNH CHÍNH QUY

Tp. Hồ Chí Minh, tháng MM/YYYY

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN

Đào Đức Anh - 1712270

Nguyễn Thành Nhân - 1712631

**Xây dựng hệ thống gợi ý sản phẩm
dựa trên mô hình Auto-Encoder**

KHÓA LUẬN TỐT NGHIỆP CỬ NHÂN
CHƯƠNG TRÌNH CHÍNH QUY

GIÁO VIÊN HƯỚNG DẪN

ThS. Trần Trung Kiên

Tp. Hồ Chí Minh, tháng MM/YYYY

Lời cảm ơn

Tôi xin chân thành cảm ơn ...

Mục lục

Lời cảm ơn	i
Đề cương chi tiết	ii
Mục lục	ii
Tóm tắt	v
1 Giới thiệu	1
2 Kiến thức nền tảng	2
2.1 Mô hình rút trích đặc trưng “Auto-Encoder”	3
2.1.1 “Undercomplete Auto-Encoder”	3
2.1.2 “Denoising Auto-Encoder”	3
2.2 “Variational Auto-Encoder”	3
2.2.1 Mô hình xác suất	3
2.2.2 “Variational Inference”	3
2.2.3 “Kullback-Leiber Divergence”	3
3 Mô hình “Variational Auto-Encoder” cho bài toán xây dựng hệ thống gợi ý	4
3.1 “Explicit feedback” và “Implicit feedback”	5
3.2 “Multinomial log-likelihood” cho bài toán xây dựng hệ thống gợi ý	5
3.3 “Mul-VAEs”	5

3.3.1	Quá trình huấn luyện mô hình	5
3.3.2	Quá trình phát sinh gợi ý	5
3.4	“KL-Vanishing”	5
3.4.1	“KL-Annealing”	5
4	Thí nghiệm	6
4.1	Tập dữ liệu sử dụng	6
4.2	Các thiết lập thí nghiệm	6
4.3	Các kết quả thí nghiệm	6
4.3.1	Kết quả mô hình cài đặt so với bài báo	6
4.3.2	Tại sao “Multinomial log-likelihood” phù hợp với bài toán xây dựng hệ thống gợi ý	6
4.3.3	So sánh với DAE	6
4.3.4	Vấn đề KL-Vanishing	6
4.3.5	Cải tiến...	6
5	Kết luận và hướng phát triển	7
5.1	Kết luận	7
5.2	Hướng phát triển	7
	Danh mục công trình của tác giả	8
	Tài liệu tham khảo	9

Danh sách hình

Danh sách bảng

Chương 1

Giới thiệu

Với việc bùng nổ dữ liệu trên mạng Internet hiện nay, con người khó có thể nắm bắt được hết tất cả các thông tin. Do đó việc tìm kiếm dữ liệu, thông tin phù hợp với nhu cầu của bản thân dần trở nên khó khăn. Việc có một hệ thống gợi ý hỗ trợ chúng ta trong việc tìm kiếm thông tin là cực kỳ hữu ích. Một hệ thống gợi ý sản phẩm là một bài toán trong lĩnh vực khai thác dữ liệu và học máy. Hệ thống gợi ý được xây dựng để dự đoán những sản phẩm phù hợp với người dùng, đặc biệt hiện nay việc đưa ra quyết định khi mà có quá nhiều lựa chọn dành cho người dùng là không hề dễ dàng. Điều này dẫn đến vai trò của một hệ thống gợi ý ngày càng quan trọng hơn, không chỉ hỗ trợ người dùng đưa ra quyết định mà còn đóng góp trong việc phát triển doanh nghiệp khi mà việc thu hút khách hàng và nâng cao trải nghiệm người dùng sẽ phụ thuộc vào một hệ thống gợi ý sản phẩm hiệu quả. Có rất nhiều lĩnh vực cần xây dựng hệ thống gợi ý có thể kể đến như thương mại điện tử, các nền tảng cung cấp các dịch vụ đa phương tiện (âm thanh, hình ảnh, video, ...), mạng xã hội (Facebook, tweeter, linkedin, ...). Theo số liệu tổng hợp được thì 75% phim được thuê trên Netflix - một nền tảng cung cấp video nổi tiếng hiện nay đến từ hệ thống gợi ý; 38% lượt click từ người dùng GOOGLE cũng đến từ hệ thống gợi ý và Amazon - một nền tảng mua bán trực tuyến mà 35% sản phẩm được bán thông qua hệ thống gợi ý sản phẩm.

Chương 2

Kiến thức nền tảng

Tong chương này, đầu tiên chúng tôi sẽ trình bày về mô hình “Auto-Encoders”, một mạng nơ-ron được dùng để học đặc trưng ẩn dựa trên phương pháp học không giám sát. Sau đó, chúng tôi giới thiệu và trình bày về nền tảng xác suất của “Variational Auto-encoders” (VAEs) và lợi ích mang lại của mô hình này so với “Auto-Encoder” trong tác vụ học đặc trưng ẩn; Những điểm lợi này chính là lý do mà chúng tôi tập trung nghiên cứu VAEs. Bên cạnh đó, chúng tôi sẽ trình bày về “Maximum Likelihood Estimation”, một phương pháp dùng để đánh giá các tham số của mô hình, đại diện cho các tham số của các phân phối xác suất dựa trên dữ liệu huấn luyện. Chương này đặc biệt là phần về “Variational Auto-Encoders” cung cấp những kiến thức nền tảng để có thể hiểu rõ về những đề xuất của chúng tôi ở chương kế tiếp.

2.1 Mô hình rút trích đặc trưng “Auto-Encoder”

2.1.1 “Undercomplete Auto-Encoder”

2.1.2 “Denoising Auto-Encoder”

2.2 “Variational Auto-Encoder”

2.2.1 Mô hình xác suất

“Maximum Likelihood Estimation”

2.2.2 “Variational Inference”

2.2.3 “Kullback-Leiber Divergence”

Chương 3

Mô hình “Variational Auto-Encoder” cho bài toán xây dựng hệ thống gợi ý”

Chương này trình bày về những đóng góp của luận văn. Ở đây, Chúng tôi phân tích hai loại dữ liệu phản hồi chính từ người dùng là: “explicit feedback” và “implicit feedback”. Đặc biệt, chúng tôi tập trung nghiên cứu mở rộng mô hình “Variational Auto-Encoders” cho implicit feedback với hàm loss là “Multinomial Log-likelihood” ở hàm mục tiêu. Chúng tôi gọi “Variational Auto-Encoders” với hàm loss như vậy là “Mul-VAE”. Đóng góp của chúng tôi là làm rõ Mul-VAE ở hai điểm:

- *Tính xếp hạng: Chúng tôi chỉ ra điểm phù hợp của Multinomial Log-likelihood cho bài toán xây dựng hệ thống gợi ý sản phẩm so với các hàm Log-likelihood thông dụng khác.*
- *KL-Annealing: chúng tôi cũng đưa ra một cách “heuristic” nhằm lựa chọn siêu tham số của mô hình Mul-VAEs.*

- 3.1 “Explicit feedback” và “Implicit feedback”
- 3.2 “Multinomial log-likelihood” cho bài toán xây dựng hệ thống gợi ý
- 3.3 “Mul-VAEs”
 - 3.3.1 Quá trình huấn luyện mô hình
 - 3.3.2 Quá trình phát sinh gợi ý
- 3.4 “KL-Vanishing”
 - 3.4.1 “KL-Annealing”

Chương 4

Thí nghiệm

4.1 Tập dữ liệu sử dụng

4.2 Các thiết lập thí nghiệm

4.3 Các kết quả thí nghiệm

4.3.1 Kết quả mô hình cài đặt so với bài báo

4.3.2 Tại sao “Multinomial log-likelihood” phù hợp với bài toán xây dựng hệ thống gợi ý

4.3.3 So sánh với DAE

4.3.4 Vấn đề KL-Vanishing

4.3.5 Cải tiến...

Chương 5

Kết luận và hướng phát triển

5.1 Kết luận

5.2 Hướng phát triển

Danh mục công trình của tác giả

1. Tạp chí ABC
2. Tạp chí XYZ

Tài liệu tham khảo