ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG

PINTHIP Anon

Nghiên cứu phương pháp phân tích ma trận SVD và một số ứng dụng trong học máy

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

THÁI NGUYÊN - 2019

Số hóa bởi Trung tâm Học liệu và Công nghệ thông tin – ĐHTN http://lrc.tnu.edu.vn

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG

PINTHIP Anon

Nghiên cứu phương pháp phân tích ma trận SVD và một số ứng dụng trong học máy

Chuyên ngành: Khoa học máy tính
Mã số: 8480101

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Người hướng dẫn khoa học: TS. Đàm Thanh Phương

THÁI NGUYÊN - 2019

Số hóa bởi Trung tâm Học liệu và Công nghệ thông tin – ĐHTN http://lrc.tnu.edu.vn

LỜI CAM ĐOAN

Tôi xin cam đoan: Luận văn thạc sỹ chuyên ngành Khoa học máy tính, tên đề tài "Nghiên cứu phương pháp phân tích ma trận SVD và một số ứng dụng trong học máy" là công trình nghiên cứu, tìm hiểu và trình bày do tôi thực hiện dưới sự hướng dẫn khoa học của TS. Đàm Thanh Phương, Trường Đại học Công nghệ Thông tin và Truyền thông - Đại học Thái Nguyên.

Kết quả tìm hiểu, nghiên cứu trong luận văn là hoàn toàn trung thực, không vi phạm bất cứ điều gì trong luật sở hữu trí tuệ và pháp luật Việt Nam. Nếu sai, tôi hoàn toàn chịu trách nhiệm trước pháp luật.

Tất cả các tài liệu, bài báo, khóa luận, công cụ phần mềm của các tác giả khác được sử dụng lại trong luận văn này đều được chỉ dẫn tường minh về tác giả và đều có trong danh mục tài liệu tham khảo.

Thái Nguyên, ngày 18 tháng 7 năm 2019

Tác giả luận văn

Pinthip Anon

LỜI CẢM ƠN

Tác giả xin chân thành cảm ơn TS Đàm Thanh Phương, trường Đại học Công nghệ thông tin và truyền thông - Đại học Thái Nguyên, là giáo viên hướng dẫn khoa học đã hướng dẫn tác giả hoàn thành luận văn này, xin được cảm ơn các thầy, cô giáo trường Đại học công nghệ thông tin và truyền thông nơi tác giả theo học và hoàn thành chương trình cao học đã nhiệt tình giảng dạy và giúp đỡ.

Xin cảm ơn nơi tác giả công tác đã tạo mọi điều kiện thuận lợi để tác giả hoàn thành chương trình học tập.

Và cuối cùng xin cảm ơn gia đình, bạn bè, đồng nghiệp đã động viên, giúp đỡ tác giả trong suốt thời gian học tập, nghiên cứu và hoàn thành luận văn này.

Xin chân thành cảm ơn.

Thái Nguyên, ngày 18 tháng 7 năm 2019

Tác giả luận văn

Pinthip Anon

DANH SÁCH HÌNH VẼ

2.1 Minh họa phân tích SVD 25
2.2 Biểu diễn SVD dạng thu gọn. 28

3.1 Ví dụ về SVD cho nén ảnh 38
3.2 Ví dụ về utility matrix với hệ thống Gợi ý bài hát. Các bài hát được người dùng đánh giá theo mức độ từ 0 đến 5 sao. Các dấu '?' nền màu xám ứng với việc dữ liệu chưa tồn tại trong cơ sở dữ liệu. Recommendation Systems cần phải tự điền các giá trị này. 41
3.3 Giả sử feature vector cho mỗi item được cho trong cột cuối cùng. Với mỗi user, chúng ta cần tìm một mô hình θ_i tương ứng sao cho mô hình thu được là tốt nhất. 43
3.4 Ví dụ về utility matrix dựa trên số sao một user rate cho một item. Một cách trực quan, hành vi của u_0 giống với u_1 hơn là u_2, u_3, u_4, u_5, u_6. Từ đó có thể dự đoán rằng u_0 sẽ quan tâm đến i_2 vì u_1 cũng quan tâm đến item này. 47
3.5 Ví dụ mô tả User-user Collaborative Filtering. a) Utility Matrix ban đầu. b) Utility Matrix đã được chuẩn hoá. c) User similarity matrix. d) Dự đoán các (normalized) ratings còn thiếu. e) Ví dụ về cách dự đoán normalized rating của u_1 cho i_1. f) Dự đoán các (denormalized) ratings còn thiếu. 49
3.6 Matrix Factorization. Utility matrix $\mathbf{Y}$ được phân tích thành tích của hai ma trận low-rank $\mathbf{X}$ và $\mathbf{W}$ 53

DANH MỤC KÝ HIỆU, TỪ VIẾT TẮT

$\mathbb{R}$	Tập hợp số thực.
$\mathbb{Z}$	Tập hợp số nguyên.
$\mathbb{C}$	Tập hợp số phức.
$\mathbb{R}^n$	Không gian Euclide n chiều.
$\|\|.\|\|$	Chuẩn Euclide.
SVD	Singular Value Decomposition.
$rank(A)$	Hạng của ma trận A.
$det(A)$	Định thức của ma trận A.
A^T	Chuyển vị của ma trận A.
A^H	Hermitian của ma trận A.
$\|\mathbf{A}\|_F$	Frobenius norm của ma trận A
$diag(A)$	Đường chéo chính của ma trận A
$trace(A)$	trace của ma trận A
$sng(x)$	Hàm xác định dấu.
$\frac{\partial f}{\partial x}$	Đạo hàm của hàm số f theo biến số $x \in R$
$\nabla_{\mathbf{x}} f$	Gradient (đạo hàm) của hàm số f theo véc tơ x.

MỤC LỤC

Lời cam đoan ..... **i**

Lời cảm ơn ..... **ii**

Danh sách hình vẽ ..... **ii**

Danh mục ký hiệu, từ viết tắt ..... **iv**

Mở đầu **1**

Chương 1. MỘT SỐ KIẾN THỨC CHUẨN BỊ **4**

1.1. Một số kiến thức về đại số tuyến tính **4**

1.2. Giới thiệu về học máy **18**

Chương 2. PHƯƠNG PHÁP PHÂN TÍCH MA TRẬN SVD ... **23**

2.1. Phân tích ma trận SVD **23**

2.2. Một số biến thể của SVD **27**

Chương 3. MỘT SỐ ỨNG DỤNG CỦA SVD TRONG HỌC MÁY . **37**

3.1. Phân tích SVD ứng dụng trong nén ảnh **37**

3.2. Ứng dụng SVD trong hệ gợi ý **39**

Kết luận chung **63**

Tài liệu tham khảo **63**

MỞ ĐẦU

Những năm gần đây, AI - Artificial Intelligence (Trí Tuệ Nhân Tạo), và cụ thể hơn là Machine Learning (Học Máy hoặc Máy Học) nổi lên như một bằng chứng của cuộc cách mạng công nghiệp lần thứ tư (1 - động cơ hơi nước, 2 - năng lượng điện, 3 - công nghệ thông tin). Trí Tuệ Nhân Tạo đang len lỏi vào mọi lĩnh vực trong đời sống mà có thể chúng ta không nhận ra. Xe tự hành của Google và Tesla, hệ thống tự tag khuôn mặt trong ảnh của Facebook, trợ lý ảo Siri của Apple, hệ thống gợi ý sản phẩm của Amazon, hệ thống gợi ý phim của Netflix, máy chơi cờ vây AlphaGo của Google DeepMind, ..., chỉ là một vài trong vô vàn những ứng dụng của AI/Machine Learning. [1].

Machine Learning là một tập con của AI. Theo định nghĩa của Wikipedia, Machine learning is the subfield of computer science that "gives computers the ability to learn without being explicitly programmed". Nói đơn giản, Machine Learning là một lĩnh vực nhỏ của Khoa Học Máy Tính, nó có khả năng tự học hỏi dựa trên dữ liệu đưa vào mà không cần phải được lập trình cụ thể.

Chính vì vậy, việc nghiên cứu phát triển các thuật toán Học máy cũng như xây dựng các ứng dụng đang là một đòi hỏi rất bức thiết của những người làm Khoa học máy tính. Nền tảng toán học của Machine Learning đòi hỏi phải có kiến thức cơ bản về Đại số tuyến tính, tối ưu, xác suất thống kê.

Việc phân tích một ma trận ra thành tích của nhiều ma trận đặc biệt khác (Matrix Factorization hoặc Matrix Decomposition) mang lại nhiều ích lợi quan trọng. Đặc biệt là trong các thuật toán Học máy như giảm số chiều dữ liệu, nén dữ liệu, tìm hiểu các đặc tính của dữ liệu, giải các hệ phương trình tuyến tính, clustering, và nhiều ứng dụng khác. Recommendation System cũng là một trong rất nhiều ứng dụng của Matrix Factorization. SVD (singular value decomposition) là một dạng khai triển của ma trận có rất

nhiều ứng dụng trong những vấn đề liên quan đến nghịch đảo và số hóa các dữ liệu. Hiện nay phân tích SVD của ma trận xuất hiện rất nhiều trong các ứng dụng thực tế như về tín hiệu số, tính các giá trị xấp xỉ trong kĩ thuật, công nghệ thông tin, học máy và được ứng dụng trong các công cụ tìm kiếm trên các websites. Tuy nhiên, nghiên cứu lý thuyết liên quan đến SVD đối với học viên cao học là vấn đề mới, chưa gần gũi và chưa dễ hiểu cho học viên cần nghiên cứu về mảng đề tài thú vị này. Do đó em đã lựa chọn đề tài "Nghiên cứu phương pháp phân tích ma trận SVD và một số ứng dụng trong học máy" thực hiện làm luận văn cao học nhằm mục đích đưa đến cho người đọc cũng như bản thân những kiến thức cơ bản nhất về khai triển SVD và tạo một cái nhìn tổng quan về cách khai triển cũng như một số tính chất, hệ quả quan trọng liên quan đến dạng khai triển này.[2], [8], [11].

Đề tài luận văn này sẽ nghiên cứu một trong những phương pháp Matrix Factorization rất đẹp của Đại số tuyến tính. Phương pháp đó có tên là Singular Value Decomposition (SVD). Ta sẽ thấy, mọi ma trận, không nhất thiết là vuông, đều có thể được phân tích thành tích của ba ma trận đặc biệt. Sau đó các ứng dụng cụ thể của SVD về nén ảnh và hệ thống gợi ý sẽ được nghiên cứu và áp dụng.

Nội dung của luận văn gồm 3 chương:

Chương 1. Các kiến thức cơ sở

Chương này trình bày các kiến thức chuẩn bị cho việc nghiên cứu. Đó là các kiến thức giới thiệu về học máy; Các kiến thức cơ sở về Image Compresstion và Recommendation System cũng như các kiến thức cơ sở về Đại số tuyến tính và giải tích ma trận.

1.1 Giới thiệu về học máy

1.2 Các kiến thức cơ sở về Image Compresstion và Recommendation System.

1.3 Các kiến thức cơ sở về Đại số tuyến tính

Chương 2. Phương pháp phân tích ma trận SVD

Nội dung chương 2 tập trung vào vấn đề phân tích ma trận SVD và các

kiến thức mở rộng về SVD. Cụ thể như sau:

2.1 Phát biểu SVD.

2.2 Các kiến thức mở rộng về SVD

Chương 3. Ứng dụng của SVD

Sau khi nắm rõ các nội dung trong chương 2, chương 3 trình bày kết quả ứng dụng SVD trong một số bài toán học máy. Cụ thể có các nội dung sau:

3.1 Image Compresstion

3.2 Recommendation System.

3.3 Một số ứng dụng khác.

Phần kết luận: Tóm tắt các kết quả đã đạt được và thảo luận hướng phát triển tiếp theo của đề tài.