TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI
VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

———— * ————

ĐỒ ÁN
TỐT NGHIỆP ĐẠI HỌC

NGÀNH CÔNG NGHỆ THÔNG TIN

HỌC TỪ ĐIỂN KHÔNG THƯA CHO BÀI TOÁN PHÂN LOẠI ẢNH

Sinh viên thực hiện: **Nguyễn Đức Tuấn**
Lớp CNTT 3 – K55
Giáo viên hướng dẫn: **PGS.TS. Huỳnh Thị Thanh Bình**

HÀ NỘI 5-2015

PHIẾU GIAO NHIỆM VỤ ĐỒ ÁN TỐT NGHIỆP

1. Thông tin về sinh viên
Họ và tên sinh viên: Nguyễn Đức Tuấn
Điện thoại liên lạc: 01675 254 336 Email: newvalue92@gmail.com
Lớp: Khoa học máy tính K55 - Viện CNTT & TT Hệ đào tạo: Đại học chính quy
Đồ án tốt nghiệp được thực hiện tại: Bộ môn Khoa học máy tính, Viện Công nghệ thông tin và truyền thông, Đại học Bách Khoa Hà Nội
Thời gian làm ĐATN: Từ ngày 1/3/2015 đến 29/5/2015

2. Mục đích nội dung của ĐATN
- Nghiên cứu về bài toán phân loại ảnh
- Nghiên cứu về mô hình thưa và bài toán học từ điển
- Đề xuất mô hình và giải thuật cải tiến học từ điển áp dụng cho bài toán phân loại.

3. Các nhiệm vụ cụ thể của ĐATN
Kế thừa kết quả của giai đoạn thực tập tốt nghiệp, ĐATN thực hiện các công việc sau :
- Nghiên cứu về mô hình thưa, và bài toán học từ điển
- Nghiên cứu bài toán phân loại ảnh và các phương pháp áp dụng học từ điển cho bài toán phân loại
- Đề xuất mô hình học từ điển với ràng buộc l_2
- Cài đặt, tiến hành chạy thực nghiệm và so sánh kết quả.

4. Lời cam đoan của sinh viên:
Tôi – *Nguyễn Đức Tuấn* - cam kết ĐATN là công trình nghiên cứu của bản thân tôi dưới sự hướng dẫn của *PGS. TS. Huỳnh Thị Thanh Bình*.
Các kết quả nêu trong ĐATN là trung thực, không phải là sao chép toàn văn của bất kỳ công trình nào khác.

Hà Nội, ngày 29 tháng 5 năm 2015
Tác giả ĐATN

Nguyễn Đức Tuấn

5. Xác nhận của giáo viên hướng dẫn về mức độ hoàn thành của ĐATN và cho phép bảo vệ:

Hà Nội, ngày tháng năm
Giáo viên hướng dẫn

PSG.TS. Huỳnh Thị Thanh Bình

TÓM TẮT NỘI DUNG ĐỒ ÁN TỐT NGHIỆP

Bài toán học phân loại ảnh là một bài toán có nhiều ứng dụng thực tế hiện nay. Tuy đã có nhiều nghiên cứu khác nhau để giải quyết bài toán này, vẫn còn những thách thức không nhỏ khi muốn tăng độ chính xác phân loại.

Mục đích đề tài nhằm tập trung vào một cách tiếp cận mới trong việc giải quyết bài toán phân loại: cách tiếp cận học từ điển. Dựa trên những nghiên cứu tương tự trước đây, trong đồ án này, tôi đề xuất một mô hình học từ điển với ràng buộc l_2 để giải quyết bài toán, tiến hành thử nghiệm, so sánh và đưa ra đánh giá về hiệu năng của các giải thuật.

Cấu trúc của đồ án gồm sáu chương với nội dung chính sau:

Chương 1 trình bày các kiến thức cơ sở về học máy và không gian vectơ.

Chương 2 trình bày tổng quan về mô hình thưa và bài toán học từ điển.

Chương 3 trình bày về bài toán phân loại ảnh và hướng tiếp cận giải quyết bài toán sử dụng phương pháp học từ điển và biểu diễn thưa.

Chương 4 trình bày mô hình học từ điển với ràng buộc l_2

Chương 5 trình bày kết quả thực nghiệm của đồ án, so sánh và đánh giá hiệu năng của các thuật toán.

Chương 6 trình bày về đánh giá và hướng phát triển của đồ án.

ABSTRACT OF THESIS

LỜI CẢM ƠN

Trước hết, em xin được chân thành gửi lời cảm ơn sâu sắc tới các thầy cô trong trường Đại học Bách Khoa Hà Nội nói chung và trong Viện Công nghệ Thông tin và Truyền thông, bộ môn Khoa học máy tính nói riêng đã tận tình giảng dạy, truyền đạt cho em những kiến thức và kinh nghiệm quý báu trong suốt quá trình học tập và rèn luyện tại trường Đại học Bách Khoa Hà Nội.

Em xin được gửi lời cảm ơn đến PGS. TS. Huỳnh Thị Thanh Bình - Giảng viên bộ môn Khoa học máy tính, Viện Công nghệ Thông tin và Truyền thông, trường Đại học Bách Khoa Hà Nội đã hết lòng giúp đỡ, hướng dẫn và chỉ dạy tận tình trong quá trình em làm đồ án tốt nghiệp. Đồng thời, em xin gửi lời cảm ơn tới TS. Nguyễn Thị Thủy – Giảng viên bộ môn Khoa học máy tính, khoa Công nghệ Thông tin, học viện Nông Nghiệp và TS. Đinh Viết Sang – Giảng viên bộ môn Khoa học máy tính, Viện Công nghệ Thông tin và Truyền thông, trường Đại học Bách Khoa Hà Nội đã cung cấp cho em những kiến thức và thảo luận bổ ích liên quan đến đề tài. Em cũng xin cảm ơn các bạn trong lớp khoa học máy tính K55 và các bạn trong phòng lab đã hỗ trợ em trong quá trình làm đồ án.

Cuối cùng, em xin được gửi lời cảm ơn chân thành tới gia đình, bạn bè đã quan tâm, động viên, đóng góp ý kiến và giúp đỡ trong quá trình học tập, nghiên cứu và hoàn thành đồ án tốt nghiệp.

Hà Nội, ngày 29 tháng 5 năm 2015

LỜI MỞ ĐẦU

Trong những năm gần đây, học máy đang trở thành một trong những lĩnh vực phát triển mạnh mẽ và trở thành một trong các lĩnh vực quan trọng trong trí tuệ nhân tạo. Nhiều thành tựu về học máy được áp dụng trong nhiều lĩnh vực khác nhau như trong tìm kiếm, robot, xử lý ngôn ngữ tự nhiên, khai phá dữ liệu,... đã và đang đem lại những thay đổi lớn trong cách con người tương tác với thế giới.

Đồ án tốt nghiệp này đề cập đến một trong các bài toán quan trọng trong học máy: bài toán phân loại ảnh. Trong bài toán phân loại ảnh, chúng ta phải phân loại ảnh vào một trong các nhãn lớp xác định dựa trên nội dung của chúng. Đây là bài toán cơ bản và có nhiều ứng dụng như trong tìm kiếm ảnh, phân tích nội dung ảnh, quảng cáo dựa trên ngữ nghĩa,… Cùng với đó, việc giải quyết bài toán nhận dạng ảnh là tiền đề cho nhiều bài toán khác như phát hiện đối tượng, phân đoạn ảnh,…

Tuy trên thực tế đã có nhiều nghiên cứu khác nhau để giải quyết bài toán này, song đây vẫn là một trong vấn đề đầy thách thức do nhiều nguyên nhân như tính đa dạng trong đối tượng ảnh, điều kiện, môi trường thu nhận ảnh, tính chất của nhãn đối tượng. Đồ án này đề cập đến một cách tiếp cận trong việc giải quyết bài toán phân loại ảnh: tiếp cận bài toán phân loại ảnh sử dụng mô hình thưa và học từ điển. Trong đồ án này, tôi đã đề xuất một cách tiếp cận mới trong bài toán học từ điển không những đem lại kết quả phân loại tốt mà còn cải thiện đáng kể thời gian phân loại. Qua đó, tôi đưa ra những đánh giá cùng những hướng phát triển trong tương lai cho hướng tiếp cận này.

DANH MỤC HÌNH VẼ

DANH MỤC CÁC BẢNG

MỤC LỤC

DANH MỤC CÁC TỪ VIẾT TẮT VÀ THUẬT NGỮ

Từ viết tắt	Từ đầy đủ	Ý nghĩa
SRC	Sparse representation-based classification	Phân loại dựa trên biểu diễn thưa
MP	Matching pursuit	Giải thuật tìm khớp để tìm biểu diễn thưa
OMP	Orthogonal matching pursuit	Giải thuật tìm khớp trực giao để tìm biểu diễn thưa
CRC_RLS	Collaborative representation based classification with regularized least square	Phân loại sử dụng biểu diễn cộng tác với ràng buộc bình phương tối thiểu
DPL	Projective dictionary pair learning	Tên một giải thuật phân loại học dựa trên học từ điển
D-KSVD	Discriminative KSVD	Giải thuật KSVD phân biệt
LC-KSVD	Label consistent KSVD	Giải thuật KSVD với tính nhất quán nhãn
FDDL	Fisher discrimination dictionary learning	Học từ điển dựa trên điều kiện phân biệt Fisher

CHƯƠNG I: CƠ SỞ LÝ THUYẾT

1. Tổng quan về học máy

1.1. Giới thiệu về học máy

Học là bản năng vô cùng tự nhiên con người. Đó là quá trình mà ta không ngừng tiếp thu, trau dồi kiến thức từ các nguồn thông tin khác nhau. Chúng ta sử dụng các kiến thức được học đó để giải quyết các vấn đề trong cuộc sống.

Cùng với sự phát triển của khoa học công nghệ, chúng ta mong muốn các dịch vụ, thiết bị chúng ta sử dụng cũng có thể học hỏi để giải quyết các vấn đề như cách chúng ta học để giải quyết chúng. Thay vì phải xử toàn bộ thông tin như trước, chúng ta mong muốn, công nghệ có thể thay thế một phần công việc vốn mất thời gian nếu xử lý thủ công để con người có thêm thời gian vào các hoạt động khác. Thay vì bị làm phiền bởi các thư spam, phải tự tay lọc các thư đó, ta mong muốn hệ thống email tự động loại bỏ chúng. Khi cần tra cứu thông tin, chúng ta mong muốn nhanh chóng nhận được thông tin nhanh chóng từ các thiết bị thông minh. Các nhà nghiên cứu thay vì phải mất thời gian kiểm tra tài liệu trên các thư viện truyền thống, họ mong muốn có thể nhanh chóng tìm được những tài liệu trên mạng liên quan đến nội dung họ quan tâm. Với khách hàng mua bán trên mạng, họ mong muốn hệ thống có thể tự động gợi ý các sản phẩm hay, phù hợp dựa trên thông tin họ cung cấp hoặc lịch sử mua bán. Với các công ti quảng cáo, họ mong muốn khách hàng xem những quảng cáo có liên quan thay vì một quảng cáo bất kỳ. Chính vì vậy, lĩnh vực học máy ra đời như một điều tất yếu của lịch sử nhân loại.

Sự phát triển của dữ liệu lớn trong những năm gần đây cũng là một nguyên nhân thúc đẩy sự phát triển của lĩnh vực học máy. Theo ước lượng của IDC, năm 2013, tổng lượng dữ liệu số hóa toàn cầu ước lượng khoảng 4.4 zettabytes (10^21 bytes) và ước lượng năm 2020, con số đó tăng lên gấp 10 lần, chạm mốc 44 zettabytes[1]. Nguồn dữ liệu từ mạng xã hội, dữ liệu trang web, dữ liệu đa phương tiện như video, âm thanh là các nguyên nhân chính dẫn đến sự bùng nổ dữ liệu này. Nhu cầu khai thác dữ liệu tự động từ nguồn dữ liệu khổng lồ đó thúc đẩy việc phát triển trong lĩnh vực học máy và các lĩnh vực có liên quan để có thể khai thác tối đa lợi ích kinh tế từ chúng.

Chúng ta bắt đầu tìm hiểu về học máy từ những khái niệm cơ bản. Khái niệm học máy đã được hình thành từ rất lâu. Từ những năm 1959, Arthur Samuel đưa ra

[1] http://www.emc.com/leadership/digital-universe/2014iview/index.html

định nghĩa đầu tiên về học máy. Theo đó, học máy là "một lĩnh vực mà giúp máy tính có khả năng học mà không cần phải được lập trình một cách tường minh". Tom M. Mitchell sau đó đưa ra một định nghĩa hình thức hơn về học máy. Mitchell định nghĩa một chương trình học máy là chương trình học dựa trên kinh nghiệm:

- Thực hiện tác vụ T
- Học từ kinh nghiệm E
- Được đánh giá thông qua phép đánh giá P

Khái niệm do M. Mitchell đưa ra cũng là khái niệm được sử dụng phổ biến trong nhiều tài liệu giáo trình về học máy hiện nay. Trong mỗi bài toán học máy, ta đều thấy rõ 3 yếu tố này. Ta cùng xem xét một vài ví dụ:

- Bài toán phân loại thư spam trong các hệ thống thư điện tử hiện nay: Hàng ngày, bạn nhận được rất nhiều thư khác nhau: thư cá nhân, thư công việc, thư quảng cáo… Một nguồn không mong muốn là các thư quảng cáo khiến bạn vô cùng khó chịu trong việc quản lý thư từ. Bài toán đặt ra là phân chia thư điện tử thành thư spam và thư không phải spam để nhanh chóng loại bỏ các thư không mong muốn. Trong ví dụ này, T là việc phân loại thư thành thư spam hay không phải thư spam. Độ đánh giá P có thể là độ chính xác phân loại. Đó là tỉ lệ mà giải thuật học máy gán nhãn spam hay không spam chính xác. Kinh nghiệm E trong trường hợp này là tập hợp các email được người dùng phản hồi spam hay không spam.
- Bài toán lái xe tự động: Xây dựng hệ thống thông minh trên xe có thể tự động điều khiển xe, phát hiện và xử lý các tình huống bất thường trên đường để đảm bảo an toàn cho người ngồi trong xe. Trong ví dụ này, tác vụ T ở đây là điều khiển xe trên đường dựa trên các tín hiệu thu được . Đánh giá P lúc này có thể là tỉ lệ thời gian mà vận hành trên đường phố thật mà không gặp tai nạn. Kinh nghiệm E có thể là tập hợp các video liên quan đến việc xử lý các tình huống lái xe được ghi lại.
- Chương trình tự động chơi cờ vua: Trong ví dụ này, ta muốn xây dựng một phần mềm có khả năng chơi cờ với chúng ta. Tác vụ T ở đây là đưa ra nước cờ hợp lý giúp đảm bảo chiến thắng tương ứng với thế trận cờ nhất định. Trong trường hợp này, ta mong muốn các chương trình chơi cờ càng thông minh càng tốt, có thể giải được các thế cờ khó. Khi đó đánh giá P có thể là số lượt thắng người chơi trong tổng số lần họ chơi. Tập E có thể là tập hợp các trận cờ được lưu trữ.

- Bài toán ước lượng giá trị nhà: Thông thường, khi xem xét giá nhà người ta thường dựa trên các thông tin nhất định như kích thước nhà, vị trí, kinh tế xã hội của khu vực đó. Bài toán đặt ra là đưa ra các ước lượng về giá nhà với mỗi trường hợp nhà được cung cấp. Kinh nghiệm E là giá nhà cũng như thông tin các nhà ở trong khu vực đó. Tác vụ T là đưa ra ước lượng giá nhà với một ngôi nhà nhất định. Đánh giá P có thể là độ lệch giữa ước lượng và giá cả thực tế của một vài căn nhà đã thực hiện giao dịch trong khu vực đó.

Thông thường, kinh nghiệm E thường được biểu diễn bởi tập các ví dụ, mẫu cho bài toán mà đã ta đã tiến hành tác vụ trên đó. Tập này gọi là tập huấn luyện (training set). Đồng thời đánh giá hiệu năng P, ta mong muốn có thể kiểm tra các trường hợp mà có thể chưa xuất hiện trong tương lai. Khi đó, ta sẽ sử dụng tập dữ liệu kiểm tra gọi là tập kiểm tra (test set). Về phép đo lường P, tùy bài toán khác nhau ta sẽ sử dụng những độ đo khác nhau.

Trên thực tế, học máy đã và đang được áp dụng rất rộng rãi hiện nay trong nhiều lĩnh vực. Nhiều ứng dụng thực tế đã cho thấy tiềm năng áp dụng của học máy trong tương lai. Có thể kể đến một vài ứng dụng tiêu biểu:

- Các công cụ tìm kiếm như Google, Baidu… không ngừng thu thập và phân tích để đưa ra các văn bản, trang web phù hợp với truy vấn người dùng. Nếu như với cách truyền thống trước đây, ta phải lần lượt tìm các tài liệu từ một nguồn tài liệu hạn chế như tại các thư viện. Thông qua các công cụ tìm kiếm ta có thể tiếp cận các nguồn thông tin trên toàn thế giới với tốc độ nhanh chóng.
- Các trang web thương mại điện tử hiện đại như Amazon, Alibaba… sử dụng học máy để nâng cao trải nghiệm người dùng: Người dùng có thể được gợi ý các sản phẩm liên quan với sản phẩm đang xem xét, có thể đưa ra gợi ý sản phẩm dựa trên sở thích người dùng
- Trợ lý ảo thông minh trên các thiết bị thông minh: Các trợ lý ảo trên các điện thoại thông minh như Siri, Cortana, Google Now cho phép người dùng có thể thông qua giọng nói, đưa ra các câu hỏi và mệnh lệnh. Người dùng có thể sử dụng các trợ lý này có thể hỏi về một thông tin nào đó, yêu cầu thực hiện một tập tác vụ nhất định và đưa ra các gợi ý cho người sử dụng.

1.2. Phân loại bài toán học máy

Các bài toán học máy rất đa dạng và phong phú, tuy nhiên chúng có thể được xếp vào hai nhóm chính:

Nhóm bài toán học có giám sát (supervised learning): trong nhóm bài toán này, tập huấn luyện có thông tin về nhãn cần đoán nhận, ước lượng. Tập huấn luyện có khuôn dạng $\mathbf{D}=\{\{\boldsymbol{x_1},y_1\},\ldots,\{\boldsymbol{x_i},y_i\}\ldots\}$, trong đó các cặp $\{\boldsymbol{x},y\}$ gọi là mẫu huấn luyện tương ứng $\boldsymbol{x}$ là mẫu dặc trưng, y là nhãn của dữ liệu. Nhiệm vụ bài toán có giám sát là nội suy nhãn dựa trên thông tin đầu vào.

Ví dụ: Bài toán lượng giá nhà dựa trên kích thước nhà. Trong trường hợp này, tập huấn luyện là tập các cặp thông tin nhà và giá nhà tương ứng. Theo quy ước trên, $\boldsymbol{x}$ là thông tin kích thước nhà, y là giá cả của nhà.

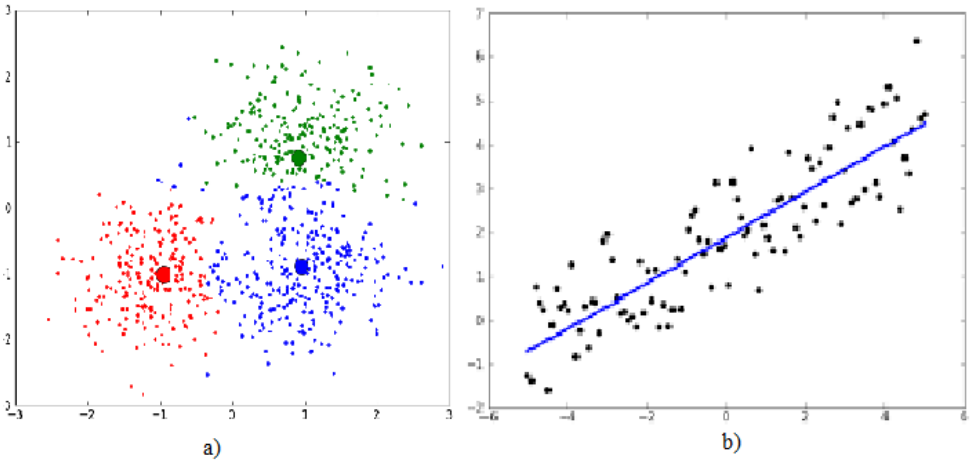
Một số bài toán cơ bản trong nhóm bài toán học có giám sát có thể kể đến:

- *Bài toán phân loại*: Là bài toán trong đó nhãn đầu vào là các giá trị rời rạc. Bài toán phân loại thư spam và không spam là một bài toán phân loại. Trong bài toán phân loại thư, tập nhãn chỉ có một trong hai nhãn spam (tương ứng 0) và không spam (tương ứng 1). Bài toán phân loại ảnh, sẽ được trình bày trong luận văn này cũng là một ví dụ cho bài toán phân loại.
- *Bài toán hồi quy*: là bài toán có giám sát trong đó tập nhãn là giá trị thực thay vì số nguyên như bài toán phân loại (Hình 1.b). Mục đích của bài toán là xây dựng mô hình cho phép đưa ra các dự đoán tương lai dựa trên các thông tin nhất định đã có. Ví dụ: bài toán ước lượng giá nhà dựa trên thông tin kích thước nhà.

Nhóm bài toán học không giám sát (unsupervised learning): là nhóm bài toán trong đó tập huấn luyện chỉ có tập đặc trưng mà không có nhãn thông tin. Mục đích của nhóm bài toán không giám sát nhằm phát hiện các cấu trúc ẩn, các mối liên hệ trong dữ liệu. Một số bài toán không giám sát có thể được kể đến:

- *Bài toán phân cụm dữ liệu*: Phân chia dữ liệu thành các nhóm dữ liệu mà trong các nhóm đó dữ liệu tương đồng với nhau (Hình 1.a). Ví dụ bài toán phân cụm trong thực tế: phân cụm tin tức tự động. Các dịch vụ tin tức tổng hợp như Google News tự động tổng hợp các nguồn tin từ các báo khác nhau và tiến hành phân cụm tự động thành các nhóm chủ đề khác nhau, có thể phát hiện các nhóm nội dung mới, đề tài mới tự động.
- *Bài toán giảm chiều dữ liệu*: Dữ liệu hoặc đặc trưng trong thực tế thường nằm trong không gian nhiều chiều. Ta mong muốn có thể biểu diễn dữ liệu trong gian ban đầu bằng không gian ít chiều hơn mà vẫn dữ được thông tin cơ bản của đối tượng hoặc có thể dùng để biểu diễn trực quan dữ liệu trong không gian 2 chiều hoặc 3 chiều.

- *Phát hiện ngoại lệ (anomaly detection)*: Bài toán phát hiện ngoại lệ là bài toán nhằm nhiệm vụ tìm ra các điểm dữ liệu dị thường, không đúng với phân bố thực của một loại dữ liệu, đối tượng nhất định.



Hình 1: Ví dụ về bài toán học máy

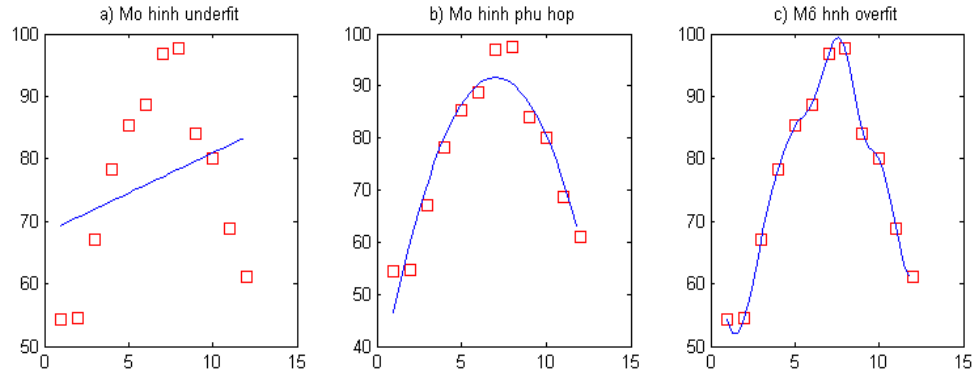
1.3. Vấn đề học chưa đủ và học quá

Trong học máy, ta xây dựng các mô hình để có thể giải quyết một bài toán cụ thể nào đó. Khi xây dựng một mô hình, ta cần tránh hai vấn đề cơ bản: học chưa đủ và học quá.

Học chưa đủ (underfiting) là tình trạng mà trong đó mô hình đạt kết quả đánh giá P thấp trên cả bộ dữ liệu huấn luyện và thử nghiệm. Lý do của vấn đề thường đến từ mô hình. Mô hình được xây dựng "quá đơn giản", không thể biểu diễn tốt được dữ liệu. Hình 2 minh họa các mô hình khác nhau cho một bài toán hồi quy. Trong ví dụ này, ta sử dụng các đa thức để mô hình hóa dữ liệu. Hình 2.a. minh họa cho một mô hình học chưa đủ. Do chỉ sử dụng đa thức bậc 1 xấp xỉ hàm, mô hình học được không thể biểu diễn được tính chất phi tuyến của dữ liệu. Khi giá trị x càng lớn, độ lệch giữa giá trị thực và giá trị tính bởi mô hình càng lớn.

Ngược lại với học đủ, ta có vấn đề học quá (overfiting). Đó là tình trạng mà mô hình biểu tốt cả phần nhiễu của dữ liệu thay vì phản ánh đúng bản chất của dữ liệu. Nguyên nhân của việc học quá có thể đến từ cả mô hình và dữ liệu. Khi mô hình phức tạp đến mức khớp chính xác mọi mẫu dữ liệu huấn luyện, có thể dẫn đến vấn đề học quá. Khi tập dữ liệu huấn luyện nhỏ, không đặc trưng tốt cho dữ liệu, mô hình có thể khớp với dữ liệu nhưng không khớp với tập kiểm tra tổng quát hơn và cũng có thể dẫn đến việc học quá. Do không phản ánh đúng bản chất của dữ liệu, khi tiến hành phép đánh giá P trên bộ dữ liệu huấn luyện, kết quả có thể gần như tuyệt đối song khi tiến hành đo đạt trên bộ dữ liệu kiểm thử, kết quả lại vô cùng tồi.

Trở về ví dụ bài toán hồi quy ở trên. Mô hình biểu diễn trong Hình 2.c là một ví dụ về mô hình học quá. Hàm học được có thể khớp với mọi cặp giá trị {x, y} mẫu ban đầu nhưng đây không thực sự là một mô hình lý tưởng. Trong ví dụ này, ta mong muốn biểu diễn dữ liệu bởi một hàm bậc hai (như trong hình Hình 2.b) thay vì như trong hình Hình 2.c.



Hình 2: Ví dụ về mô hình học chưa đủ và học quá

Vậy làm thế nào để khắc phục tình trạng học không đủ hoặc học quá? Ta sẽ đề cập đến một số giải pháp cho các vấn đề này.

Đối với học chưa đủ, nguyên nhân đến từ bản chất mô hình. Do vậy cần phải thay đổi mô hình để tăng khả năng của mô hình để tăng khả năng biểu diễn của mô hình.

Đối với vấn đề học quá, ta có thể xem xét một số giải pháp liên quan đến mô hình hoặc dữ liệu:

- Thay đổi mô hình: Cần xem xét nhiều mô hình cũng như hiệu chỉnh các tham số thích hợp để hạn chế tình trạng học quá. Ví dụ trong bài toán hồi quy như ở trên, giả sử ta xấp xỉ hàm bằng các đa thức bậc n. Khi đó ta cần phải thay đổi giá trị n đến khi nhận được hàm đa thức biểu diễn dữ liệu đủ tốt.
- Sử dụng các ràng buộc: trong học máy các ràng buộc (regularization term) là các toán tử phạt (penalty) nhằm giới hạn khả năng biểu diễn của mô hình. Các ràng buộc này thường có dạng các nhân tử $\lambda\|F\|_p$ trong đó λ biểu thị mức độ quan trọng của ràng buộc, $\|F\|_p$ là biểu thức phạt, p=0, 1, 2 là giá trị chuẩn (norm) của đại lượng *F*. Các ràng buộc này khiến các tham số mô hình không quá lớn, do vậy hạn chế được tình trạng học quá. Ta sử dụng lại ví dụ bài toán hồi quy ở trên. Khi sử dụng đa thức bậc cao, ta gặp tình trạng học quá. Khi thêm toán tử ràng buộc $\lambda\|p\|_0$ vào mô hình, trong đó *p* là vectơ hệ số

của đa thức, ta có thể ràng buộc khiến phần lớn các hệ số của đa thức học nhận giá trị 0.

- Sử dụng thêm dữ liệu: sử dụng thêm dữ liệu là một cách đơn giản nhưng tỏ ra vô cùng hiểu quả. Đây cũng là một cách được dùng nhiều trong vài năm trở lại đây bởi giải pháp này tận dụng được ưu thế về dữ liệu lớn. Trong trường hợp dữ liệu đầu vào ít, có thể sinh thêm dữ liệu bằng việc sử dụng các phép biến đổi đơn. Ví dụ để tăng tính bất biến của đối tượng ảnh về mặt góc quay, ta có thể sinh mẫu từ ảnh ban đầu thông qua các phép quay ảnh với các góc quay ngẫu nhiên.

2. Không gian vectơ

Trong bài toán học từ điển sẽ được trình bày trong phần tiếp theo, ta sẽ quan tâm đến không gian đặc trưng, biểu diễn vectơ trong hệ cơ sở mới,… Do vậy phần này sẽ cung cấp các kiến thức cơ bản nhất về không gian tuyến tính và hệ cơ sở để giúp người đọc quen với các khái niệm cũng như dễ dàng hơn trong việc tiếp cận các kiến thức được trình bày trong phần sau.

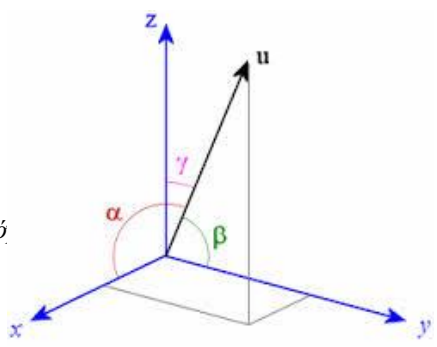
2.1. Một số quy ước

Trước khi đi vào các kiến thức cơ bản, chúng ta có một vài quy ước để cho việc biểu diễn thống nhất:

- Ký hiệu chữ cái in hoa ví dụ *A, B, C,…* biểu thị ma trận.Ta ký hiệu $A \in R^{mxn}$ biểu thị ma trận gồm m dòng, n cột.
- Ký hiệu chữa thường viết nghiêng ***x*** biểu thị một vectơ, chữ thường không viết đậm biểu thị biến số thực hoặc nguyên. Để thuận tiện, ta mặc định vectơ là vecto cột (tức vectơ có kích thước n x 1 trong đó n là chiều của vectơ). Vectơ dòng được biểu diễn bởi x^T.
- Phần tử thứ *i* của vectơ x được biểu diễn bởi: x_i
- Ký hiệu chữ cái với cặp chỉ số dưới a_{ij} (hoặc A_{ij}) biểu diễn phần tử ở *hàng i,* cột *j* của ma trận A.
- a_j là cột thứ *j* trong ma trận, a_j^T là dòng thứ *j* của ma trận *A.*

2.2. Không gian vectơ

Trong môn hình học sơ cấp, chúng ta được làm quen đến không gian 3 chiều.Trong



không gian 3 chiều, mỗi một vectơ v được biểu diễn bởi một bộ ba số (x, y, z). Tập hợp các điểm trên không gian này là một tập hợp vô hạn. Trên tập hợp này, ta cũng định nghĩa phép toán cơ cộng và nhân vectơ:

- Phép cộng vectơ: (x, y, z) + (u, v, w) = (x + u, y + v, z + w)
- Phép nhân với vô hướng: k(x, y, z) = (ku, ky, kz)

Tập hợp các điểm này là một ví dụ về một không gian vectơ. Trong phần này, ta đi vào khái niệm tổng quát của không gian vectơ.

Định nghĩa 1.1. Giả sử ***V*** là một tập hợp mà trên đó định nghĩa hai phép toán hai ngôi:

a) Phép cộng véctơ (ký hiệu là +) : $u,v \in V,\ u+v \in V$

b) Phép nhân véctơ với vô hướng: $u \in V,\ k \in \mathbb{R},\ ku \in V$

V được gọi là không gian vectơ trên trường số thực nếu thỏa mãn các điều kiện sau:

i. Nếu u và $v \in V$ thì $u+v \in V$

ii. Tính giao hoán của phép cộng:

$$\forall u,v \in V,\ u+v=v+u$$

iii. Tính kết hợp của phép cộng:

$$\forall u,v,\mathrm{w} \in V,\ (u+v)+\mathrm{w}=u+(v+\mathrm{w})$$

iv. Tồn tại một phần tử *0*, thỏa mãn:

$$\forall u, \in V, u+0=u$$

v. $\forall u \in V$, tồn tại phần tử đối ký hiệu $-u$, thỏa mãn:

$$u+(-u)=0$$

vi. Nếu $u \in V, \lambda \in \mathbb{R}$, thì $\lambda u \in V$

vii. $\forall u,v \in V, \forall k \in \mathbb{R},\ k(u+v)=ku+kv$

viii. $\forall u \in V, \forall k,h \in \mathbb{R},\ (k+h)u=ku+hu$

ix. $\forall u \in V, \forall k,h \in \mathbb{R},\ \mathrm{h}(\mathrm{k}u)=(hk)\ u$

x. $\forall u \in V,\ 1.u=u$

Trong các tiên đề này, tiên đề (i) và (vi.) nói về tính chất đóng của không gian ***V*** với phép cộng và phép nhân đối với số thực. Tiên đề (iv.) gọi là tiên đề về vecto không.Vectơ $-u$ trong tiên đề (v.) gọi là vectơ đối.

Nếu thay ràng buộc $k \in R$ bởi $k \in C$, ta có khái niệm không gian vectơ trên trường số phức.

Ví dụ: Không gian Euclid R^n một không gian vectơ. Mỗi điểm trong R^n được biểu diễn bởi toạ độ $(x_1, x_2, ..., x_n)$. Không gian 3 chiều trong phần mở đầu là một trường hợp của không gian R^n. Các phép toán trong R^n :

- Phép cộng vectơ : $u + v = (u_1 + v_1, u_2 + v_2, ..., u_n + v_n)$
- Phép nhân vectơ với vô hướng: $ku = (ku_1, ku_2, ..., ku_n)$

Dễ dàng kiểm tra các tính chất của không gian vectơ trên trên R^n. Vectơ không trong không gian R^n là $(0, 0, .., 0)$. Vectơ đối của vecto $u = (u_1, u_2, ..., u_n)$ là vectơ $-u = (-u_1, -u_2, ..., -u_n)$.

Ví dụ: Gọi *P* là tập hợp các đa thức bậc nhỏ hơn hoặc bằng n:

$$P_n = \{p(x) = a_0 + a_1x_1 + a_2x_2 + ... + a_nx_n \mid a_0, a_1, ..., a_n \in \mathbb{R}\}$$

Nếu $p(x) = a_0 + a_1x_1 + a_2x_2 + ... + a_nx_n$ và $q(x) = b_0 + b_1x_1 + b_2x_2 + ... + b_nx_n$. Khi đó, hai phép toán này được định nghĩa như sau:

- $p(x) + q(x) = (a_0 + b_0) + (a_1 + b_1)x_1 + ... + (a_n + b_n)x_n$
- $kp(x) = ka_0 + ka_1x_1 + ... + ka_nx_n$

2.3. Không gian con của không gian vectơ

Định nghĩa 1.2. Không gian con của không gian vectơ ***V*** trên trường số thực ***R*** (gọi tắt là không gian con) là một tập hợp ***W*** nếu tập ***W*** thỏa mãn hai tính chất:

i. $\forall u, v \in W, u + v \in W$

ii. $\forall u \in W, \forall k \in R, ku \in W$

Như vậy, không gian con của không gian vectơ ***V*** là tập con của không gian ***V*** mà trên đó đóng với với hai phép toán nhân với vô hướng và phép cộng. Không gian con cũng là một không gian.

Bất kỳ không gian nào cũng tồn tại hai không gian con: không gian chính nó và không gian chứa duy nhất vectơ *0*.

Ví dụ: tập hợp các điểm (x, y) trên trên đường thẳng ax + by = 0 là không gian con của không gian $\boldsymbol{R^2}$. Dễ dàng kiểm tra hai tính chất này:

- Xét hai điểm $u=(u_1, u_2)$, và $v=(v_1,v_2)$ nằm trên đường thẳng. khi đó: $au_1 + bu_1 = 0$ và $av_1 + bv_1 = 0$. Từ đó dẫn đến $u + v = (au_1+bu_2) + (av_1+bv_2) = a(u_1 + v_1) + b(u_2 + v_2) = 0$ hay $u + v$ cũng thuộc đường thẳng ax + by = 0
- Xét u thuộc đường thẳng. Khi đó với mọi giá trị k thuộc R, ta có: $ku = k(au_1 + bu_2) = a(ku_1) + b(ku_1) = 0$ hay ku cũng thuộc đường thẳng ax + by = 0

Định lý 1.1. Giao của một họ bất kỳ các không gian con của ***V*** là một không gian con của ***V***.

2.4. Không gian sinh bởi một hệ vectơ

2.4.1. Tổ hợp tuyến tính của một hệ vectơ

Định nghĩa 1.3. Gọi V là không gian vectơ trên R. Cho $v_1, v_2, ..., v_m \in V$ .Vectơ $u \in V$ có dạng:

$$u = \alpha_1 v_1 + \alpha_2 v_2 + ... + \alpha_m v_m$$

trong đó $\alpha_i \in R, i = 1...m$ được gọi là tổ hợp tuyến tính của các vectơ $v_1, v_2, ..., v_m$

Ví dụ trong không gian R^2, mọi vectơ đều là tổ hợp tuyến tính của hai vectơ $i=(1, 0)$ và $j=(0, 1)$. Thật vậy: $\forall u \in R^2, u = (x, y) = x * (1, 0) + y * (0, 1) = x * i + y * j$

2.4.2. Không gian sinh bởi hệ vectơ và hệ sinh

Định nghĩa 1.4. Cho tập ***S*** = $\{x_1, x_2, .., x_n\}$ là một họ các vectơ trong không gian vectơ ***V***. Tập hợp tất cả các tổ hợp tuyến tính của ***S*** gọi là bao tuyến tính của họ ***S***, được ký hiệu là span(***S***).

Định lý 1.2. Nếu ***S*** là một họ vectơ trong không gian ***V*** thì W = span(***S***) là một không gian con của ***V***.

Việc chứng minh định lý dựa trên định nghĩa về không gian con. Ta không đề cập chi tiết trong luận văn này.

Định nghĩa 1.5. Nếu span(***S***) = ***V*** tức với mọi vectơ trong ***V*** đều là tổ hợp tuyến tính của hệ ***S*** thì ta nói rằng họ S sinh ra ***V*** hay ***S*** là hệ sinh của ***V***.

c. Họ vectơ độc lập tuyến tính

Định nghĩa 1.6. Ta nói họ vectơ ***S*** = $\{v_1, v_2, \ldots, v_n\}$ của không gian vectơ ***V*** là độc lập tuyến tính nếu biểu thức

$$\alpha_1 v_1 + \alpha_2 v_2 + ... + \alpha_m v_m = 0$$

trong đó $c_1,\ldots,c_n$ là các số thực chỉ xảy ra ki $c_1=c_2=\ldots=c_n=0$

Nếu tồn tại các số $c_1, c_2,.., c_n$ không đồng thời bằng 0 sao cho hệ thức thỏa mãn, khi đó ta nói họ ***S*** là phụ thuộc tuyến tính.

Ví dụ: Họ các vectơ $i=(1, 0)$ và $j=(0, 1)$ là họ vectơ độc lập tuyến tính. Trong khi đó, họ vectơ $u = (1, 0)$ và $v = (2, 0)$ là họ vectơ phụ thuộc tuyến tính vì $2u – v = 0$

2.5. Cơ sở, số chiều và tọa độ của không gian vectơ

Trong các phần trước, ta đã trình bày về hệ vectơ, biểu diễn vectơ thông qua hệ sinh. Trong phần này, ta sẽ quan tâm đến một hệ sinh đặc biệt: hệ cơ sở cũng như những tính chất đặc biệt của nó.

Định nghĩa 1.7. Người ta gọi cơ sở của không gian vectơ ***V*** là họ vectơ $\{e_1, e_2,..,e_n\}$ nếu họ này thỏa mãn hai điều kiện

i) Họ $\{e_1, e_2,..,e_n\}$ độc lập tuyến tính

ii) Họ $\{e_1, e_2,..,e_n\}$ là hệ sinh của ***V***

Ví dụ: họ vectơ $\{i, j\}$ trong đó $i = (1, 0)$ và $j=(0, 1)$ là một cơ sở của không gian vectơ R^2. Lưu ý rằng một không gian có thể có vô số các hệ cơ sở khác nhau. Ví dụ (u,v) trong đó $u=(1, 0)$ và $v=(1, 1)$ là một cơ sở khác của không gian R^2.

Định nghĩa 1.8. Chiều của không gian vectơ: Nếu {e1,e2,..,en} là một cơ sở của không gian vectơ ***V***, ta nói ***V*** là không gian vectơ n chiều, n là số chiều của ***V***, kí hiệu dim(***V***) = n.

Như vậy, trong không gian n chiều, mọi họ vecto n chiều độc lập tuyến tính đều là một hệ cơ sở. Ví dụ (u, v) và (i, j) trong ví dụ trên là các cơ sở khác nhau của không gian R^2.

Định lý 1.3. Nếu $\{e_1, e_2,..,e_n\}$ là một cơ sở của không gian vectơ n chiều ***V*** thì mọi vecto $x \in V$ đều có thể biểu diễn duy nhất:

$$x = \alpha_1 e_1 + \alpha_2 e_2 + ... + \alpha_1 e_n$$

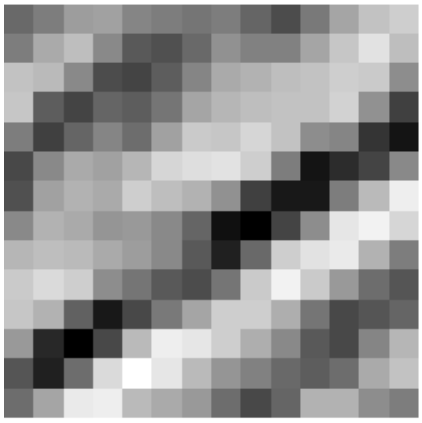
trong đó $\alpha_1, \alpha_2,.., \alpha_n$ là các số thực. Khi đó $(\alpha_1, \alpha_2, ..., \alpha_n)$ gọi là tọa độ của x với cơ sở ***V***. Tùy vào hệ cơ sở, tọa độ tương ứng của vectơ đó với hệ cơ sở có thể thay đổi theo.

CHƯƠNG II: MÔ HÌNH THƯA VÀ HỌC TỪ ĐIỂN

1. Mô hình thưa

1.1. Giới thiệu về mô hình thưa

Ảnh tự nhiên được số hóa trong máy tính dưới dạng ma trận số. Với ảnh màu, ta có 3 ma trận số tương ứng với các kênh màu đỏ (R), màu xanh lục (G) và màu xanh dương (B). Với ảnh một mức xám, ảnh được biểu diễn bởi một ma trận duy nhất. Hình 3 minh họa biểu diễn một mẫu ảnh kích thước 14x14. Cũng có thể xem biểu diễn này dưới dạng vectơ thông qua việc "duỗi" ma trận số để thu được vectơ 14x14 = 156 chiều.

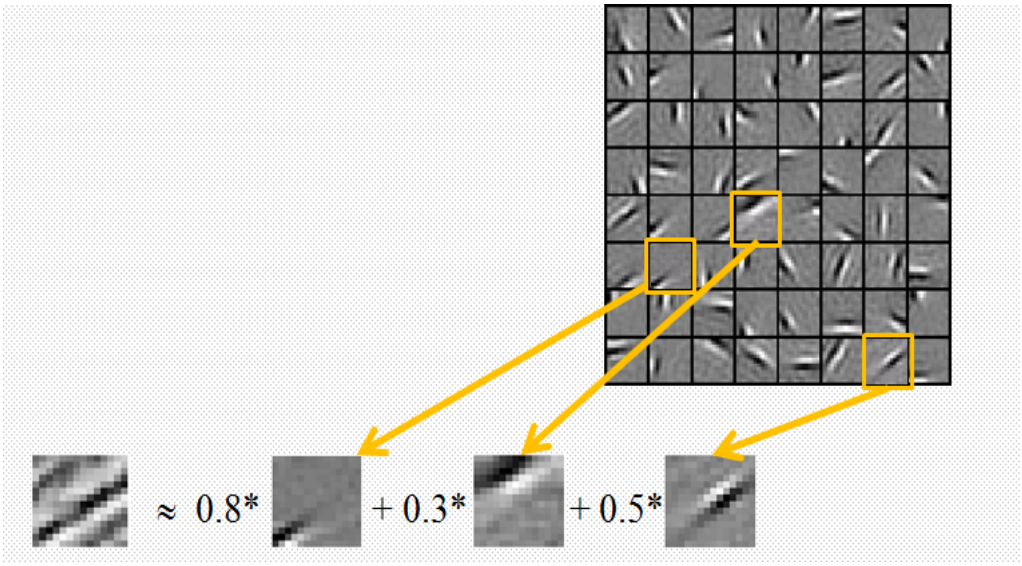


109	125	162	172	143	138	1
112	174	206	139	81	69	
203	201	139	59	60	90	1
218	75	49	104	90	111	1
126	41	99	139	97	163	2
57	148	181	160	189	224	2
67	169	182	168	218	191	1
137	184	169	142	147	133	
185	192	190	173	158	141	
203	228	221	136	113	78	

Hình 3: Mẫu ảnh và biểu diễn dày

Tuy nhiên, cách biểu diễn này còn đơn giản và nhiều hạn chế vì không thể hiện được những tính chất cơ bản về đối tượng. Chỉ một vài thay đổi trong một vài điểm ảnh dẫn đến sự thay đổi của toàn bộ vectơ biểu diễn. Ta mong muốn một cách biểu diễn tốt hơn vậy, có khả năng bất biến hơn với sự biến đổi nhỏ.

Mô hình thưa (sparse-land model) là là một cách tiếp cận nhằm đạt được mục tiêu đó. Mô hình thưa là mô hình trong đó ta biểu diễn tín hiệu ban đầu sang một không gian mới nhiều chiều hơn trong đó chỉ một vài giá trị tọa độ của nó khác không. Ví dụ, mẫu ảnh có thể biểu diễn sử dụng mô hình thưa như sau:



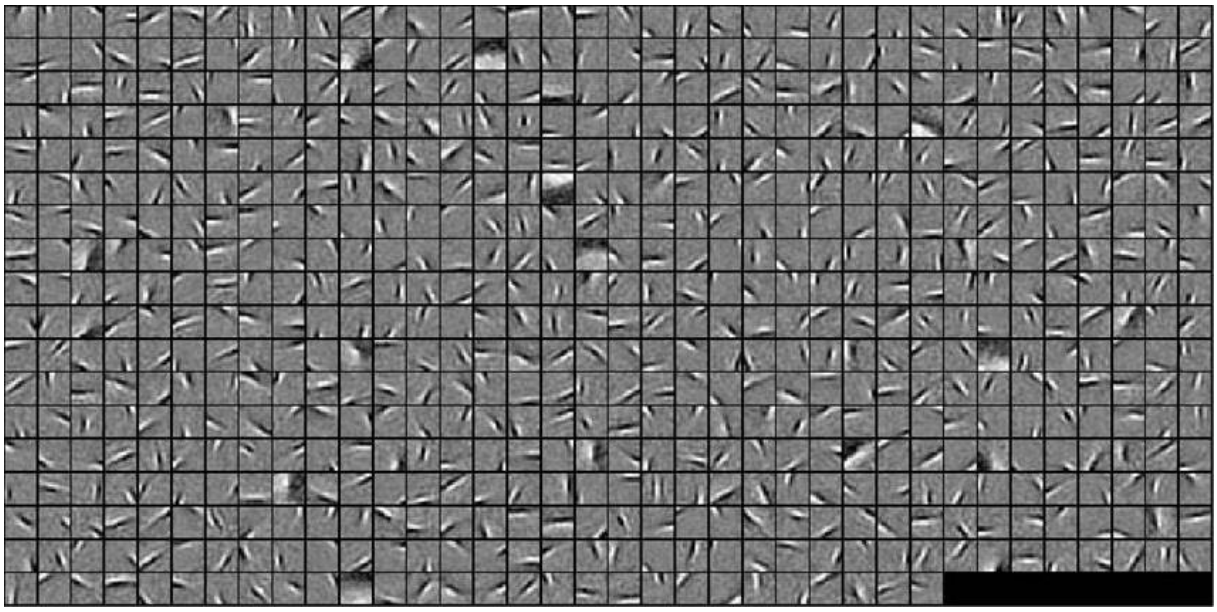
Hình 4: Biểu diễn thưa trên mẫu ảnh

Ta gọi tập các vectơ dùng để biểu diễn không gian mới là từ điển. Ta cũng gọi mỗi vectơ này là một từ trong từ điển. Về bản chất, khái niệm từ điển ở đây tương đương khái niệm hệ sinh trong lý thuyết về không gian vectơ đã đề cập tới trong chương I. Lý do nó không tương đương với khái niệm hệ cơ sở bởi số lượng từ thông thường lớn hơn số chiều của dữ liệu (từ điển có tính overcomplete) do vậy tính độc lập tuyến tính không còn được đảm bảo.

1.2. Mô hình thưa trong sinh học

Mô hình thưa phát triển dựa trên những nghiên cứu về bộ não thú. Bộ não thú có hàng tỉ nơron hoạt động độc lập với nhau. Thông tin trong bộ não được lưu trữ trong não thông qua việc bật/tắt của các nơron [2]. Tương ứng với một tín hiệu đến bộ não, chỉ có một lượng vô cùng nhỏ các nơron ở trạng thái bật. Với mỗi tín hiệu khác nhau, tập các nơron ở trạng thái bật cũng khác nhau. Điều đó gợi ý "tính thưa" trong biểu diễn thông tin hình ảnh trong bộ não. Nghiên cứu của Field [10] còn chỉ ra mối tương quan giữa vùng thị giác cơ bản (v1) của con người và biểu diễn thưa. Tổng quát, vùng thị giác của con người được tổ chức vô cùng phức tạp, gồm nhiều vùng thị giác khác nhau: v1, v2, v3, v4, v5, mỗi vùng giúp tiếp thu những thông tin nhất định về đối tượng hình ảnh [15]. Các vùng thị giác mức cao giúp tiếp thu thông tin mang tính bất biến hơn, trừu tượng hơn về đối tượng. Vùng v1 hay còn gọi là vùng thị giác cơ bản là vùng xử lý thông tin đầu tiên trước khi thông tin đầu ra được chuyển sang các vùng khác. Nghiên cứu về vùng thị giác cơ bản chỉ ra chức năng của vùng v1 hoạt động như các bộ lọc cạnh. Ví dụ khi xử lý một ảnh gồm hai nửa đen và trắng, các nơron chủ yếu lưu trữ thông tin về đường phân cách giữa hai vùng, chỉ có một vài nơron sẽ nắm giữ thông tin về độ tương

phản giữa hai vùng. Thông qua việc học từ điển với mô hình thưa, Field cho thấy mô hình thưa trên từ điển cũng có thể mô phỏng hoạt động của vùng thị giác v1 (Hình 5):



Hình 5: Từ điển học được từ tập mẫu ảnh thô

1.3. Ứng dụng mô hình thưa

Ta đề cập đến một vài ứng dụng của mô hình thưa trong thực tế để thấy được tính ứng dụng của mô hình.

Ứng dụng đầu tiên của biểu diễn thưa là trong việc nén dữ liệu. Trong không gian ban đầu, vectơ biểu diễn ở dạng dầy (dense) khi đa số các thành phần của vectơ đều khác không. Trong biểu diễn thưa, chỉ một tỉ lệ nhỏ các thành phần trong vectơ khác không. Do vậy, ta có thể lưu trữ giá trị mà tại đó hệ số khác không thay vì lưu toàn bộ vectơ, từ đó giảm thiểu dung lượng lưu trữ cần thiết. Mô hình thưa hiện nay đang được sử dụng trong nén ảnh theo chuẩn jpeg.

Bên cạnh việc nén dữ liệu, sử dụng mã hóa thưa còn có thể dùng để khử nhiễu. Thông qua việc học từ điển trên các mẫu ảnh nhiễu, ta nắm được cấu trúc ảnh và sử dụng từ điển học được để loại bỏ nhiễu.

Ứng dụng cuối cùng của biểu diễn thưa là trong việc phân loại ảnh mà sẽ được trình bày trong đồ án này.

2. Bài toán học từ điển

2.1. Mô hình toán học

Trong mô hình thưa trên, ta đề cập đến việc sử dụng từ điển. Vậy từ điển này được sinh ra như nào? Câu trả lời là từ điển được học từ dữ liệu. Trong mục này ta sẽ trình bày bài toán này. Gọi Y là ma trận dữ liệu đầu vào, $Y \in \mathbb{R}^{m \times N}$, mỗi mẫu $y_i \in \mathbb{R}^m$ tương ứng với một cột trong ma trận.

Bài toán đặt ra là xây dựng từ điển $D \in \mathbb{R}^{m \times K}$ (trong đó m << K), và ma trận hệ số $X \in \mathbb{R}^{K \times N}$ thỏa mãn:

$$\underset{D}{\operatorname{argmin}} \|Y - DX\|_F^2 \ \ s.t. \ \|x_j\|_0 \le T \text{ and } \|d_i\|_2^2 = 1,\ 1 \le j \le N,\ 1 \le i \le K \qquad (2.1)$$

Trong đó:

- K: số từ trong từ điển
- d_i là từ thứ i trong từ điển, tương ứng với cột thứ i trong ma trận D
- x_i là hệ số biểu diễn tương ứng với mẫu thứ i, là cột thứ i trong ma trận X
- T: độ thưa trong biểu diễn hệ số
- Ràng buộc $\| \ \|_0$ là chuẩn l_0 của vectơ, nhận giá trị bằng số lượng phần tử của vectơ khác 0
- Ràng buộc $\| \ \|_F^2$ là chuẩn Frobenous của ma trận, $\|X\|_F^2 = tr(X^T X) = tr(XX^T)$ và nhận giá trị chính bằng tổng bình phương của các phần từ trong ma trận.

Trong bài toán học từ điển, K >> m, tức số từ trong từ điển lớn hơn chiều không gian của tín hiệu ban đầu. Như trong phần lý thuyết về hình học không gian, để biểu diễn không gian m chiều, ta chỉ cần một hệ gồm m từ độc lập tuyến tính để biểu diễn chính xác. Do vậy, khi bỏ qua ràng buộc thưa, ta có vô số lời giải cho biểu diễn vectơ. Ràng buộc l_0 giúp bài toán có nghiệm duy nhất. Lý thuyết về tính duy nhất trong biểu diễn thưa được trình bày trong một lĩnh vực toán học là compressed sensing [3].

2.2. Sơ đồ tối ưu hàm mục tiêu

Bài toán tối ưu ở trên là bài toán tối ưu nhiều biến không lồi. Bài toán thường được giải theo cơ chế tối ưu tuần tự, tối ưu một yếu tố khi cố định một yếu tố. Quá trình này được trình bày như trong Giải thuật 2.1.

Việc khởi tạo từ điển ban đầu có thể bằng nhiều cách: lấy ngẫu nhiên từ mẫu, sử dụng các từ điển biết trước (từ điển DCT) hoặc thậm chí khởi tạo ngẫu nhiên. X_{init} có thể được tìm từ D_{init} thông qua các phương pháp tìm mã thưa sẽ được trình bày ở mục sau.

Giải thuật 2.1. Tối ưu bài toán học từ điển

Input : - Tập dữ liệu Y, kích thước từ điển K

- số vòng lặp tối đa: *maxIter*

Output : Từ điển *D*, biểu diễn *X*

Begin

1. Khởi tạo từ điển D gồm K từ
2. iter = 1
3. **while** iter <= *maxIter*:

   Update D while fix X

   Update X while fix D

   iter = iter + 1
4. **End while**

End

Quá trình cập nhật X khi cố định từ điển D còn được gọi là quá trình tìm biểu diễn thưa (sparse coding). Quá trình cập nhật D khi cố định X là quá trình cập nhật từ điển.

2.3. Tìm biểu diễn thưa

Tìm biểu diễn thưa mức T là việc tìm biểu mẫu dữ liệu thông qua từ điển mà trong đó số hệ số khác 0 trong vectơ biểu diễn không quá T. Trên thực tế, việc tìm T từ giúp biển diễn tốt nhất tín liệu là bài toán NP-khó, có thể được quy dẫn về bài toán lựa chọn tập hợp con trong tối ưu hóa tổ hợp [6]. Vì vậy, bài toán được giải bằng các giải thuật xấp xỉ. Trong mục này, ta đề cập tới một số cách tiếp cận giải quyết bài toán tìm biểu diễn thưa.

2.3.1. Tìm biểu diễn thưa sử dụng cách tiếp cận tham lam

Trong hướng tiếp cận này, vectơ thưa khởi tạo bằng vectơ ***0***. Ta lần lượt chọn các từ đùng để biểu diễn tín hiệu đó và cập nhật hệ số biểu diễn theo một tiêu chí tham lam nào đó đến khi đạt tới ngưỡng thưa T mong muốn. Hai giải thuật cơ bản theo hướng tham lam này có thể kể đến MP (matching pursuit) và OMP (Orthogonal matching pursuit).

- **Giải thuật MP (matching pursuit)**: tiêu chí tham lam trong MP là dựa trên tính tương liên giữa từ trong từ điển và phần dư thừa của dữ liệu cần được biểu diễn. Tính tương liên của 2 vectơ *u, v* được định nghĩa là độ lớn của tích vô hướng giữa hai vectơ: $|u^T v|$. Do trong bài toán học từ điển ta ràng buộc đơn vị cho độ dài của các từ trong từ điển, giá trị độ tương liên này thể hiện độ lớn hình chiếu của vectơ tín hiệu lên từ trong từ điển. Hai vectơ có độ tương liên cao, khả năng biểu diễn của nó thông qua vectơ khác tốt. Khi hai vectơ vuông góc với nhau, độ tương liên thấp, cặp vectơ đó không có khả năng biểu diễn lẫn nhau (Hình 6). Giải thuật tìm mã thưa theo phương pháp MP được mô tả trong Giải thuật 2.2:

Giải thuật 2.2. Matching pursuit (MP)

Input: Từ điển D, mẫu cần biểu diễn, độ thưa T

Output: Vectơ hệ số α tương ứng với x

Begin

1. $\alpha \leftarrow \mathbf{0}$
2. $r \leftarrow x$
3. **while** $\|\alpha\|_0 < T$ **do**
4. Lựa chọn từ mà có tương quan lớn nhất với r

$$i \leftarrow \underset{i=1,\ldots,K}{\operatorname{argmax}} \left| d_i^T r \right|$$

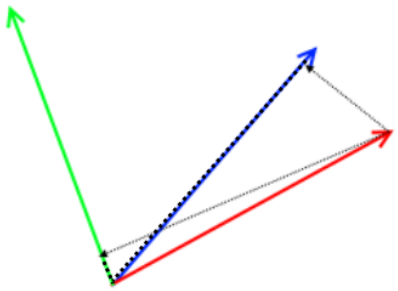
5. Cập nhật hệ số và độ dư thừa

$$\alpha[i] \leftarrow \alpha[i] + d_i^T r$$

$$r \leftarrow r - (d_i^T r) d_i$$

6. **End while**

End



Hình 6: Độ tương liên và khả năng biểu diễn của từ

- **Giải thuật OMP (Othorgonal matching pursuit)**: đây cũng là giải thuật tham lam tương tự với MP. Nhưng khác với MP vốn quan tâm đến khả năng biểu diễn của từng từ độc lập với nhau, OMP quan tâm đến khả năng biểu diễn của tập từ được sử dụng hiện tại. Tại mỗi bước tham lam, giải thuật OMP sẽ chọn từ sao cho khi thêm từ đó vào tập từ sử dụng, tập từ mới sẽ giúp biểu diễn tín hiệu tốt nhất. Gọi D_Γ là từ điển gồm các từ trong đó hệ số khác không. Khi đó, phần hê số tương ứng với từ điển là x_Γ sẽ được tính thông qua :

$$\alpha_\Gamma = \underset{\alpha_\Gamma}{\operatorname{argmin}} \left\| x - D_\Gamma \alpha_\Gamma \right\|_F^2 \tag{2.2}$$

Đạo hàm riêng phần của biểu thức (2.2) ứng với α_Γ là:

$$\frac{\partial f}{\partial \alpha_\Gamma} = -D_\Gamma^T (x - D_\Gamma \alpha_\Gamma) \tag{2.3}$$

Giải $\frac{\partial f}{\partial \alpha_\Gamma} = 0$ ta thu được lời giải:

$$\alpha_\Gamma = \left((D_\Gamma^T D_\Gamma)^{-1} D_\Gamma{}^T \right) x \tag{2.4}$$

Khi đó vectơ dư thừa biểu diễn bởi từ điển D_Γ là:

$$\begin{aligned} r &= x - D * \left((D^T D)^{-1} D^T \right) x = x - D_\Gamma \left((D_\Gamma^T D_\Gamma)^{-1} D_\Gamma{}^T \right) x \\ &= \left(I - D_\Gamma (D_\Gamma^T D_\Gamma)^{-1} D_\Gamma{}^T \right) x \end{aligned} \tag{2.5}$$

Trong bước tham lam ta sẽ chọn từ để làm giảm độ dài vectơ r nhiều nhất. Giải thuật OMP chi tiết được trình bày trong Giải thuật 2.3:

Giải thuật 2.3. Orthogonal Matching pursuit (OMP)

Input: Từ điển D, mẫu cần biểu diễn, độ thưa T

Output: Vectơ hệ số α tương ứng với x

Begin

1. $\Gamma=\emptyset$
2. **for** iter = 1,...,T **do**
3. Chọn từ giúp giảm giá trị hàm mục tiêu nhiều nhất

$$i \leftarrow \text{argmin}\left\{\min_{i\in\Gamma^C}\left\|\left(I - D_{\Gamma\cup\{i\}}(D_{\Gamma\cup\{i\}}^T D_{\Gamma\cup\{i\}})^{-1}D_{\Gamma}{}^T\right)x\right\|_2^2\right\}$$

4. Cập nhật tập từ được sử dụng: $\Gamma \leftarrow \Gamma\cup\{i\}$
5. Cập nhật độ dư thừa:

$$r \leftarrow \left(I - D_{\Gamma}(D_{\Gamma}^T D_{\Gamma})^{-1}D_{\Gamma}{}^T\right)x$$

6. Cập nhật hệ số:

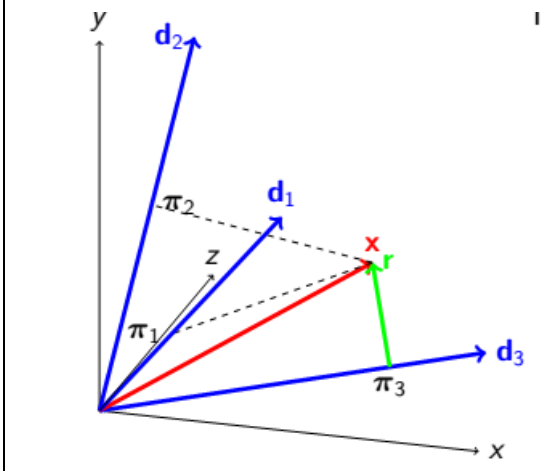
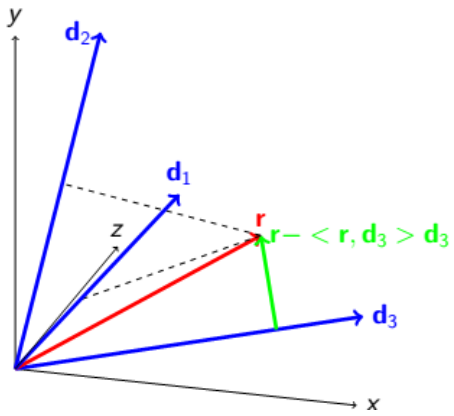
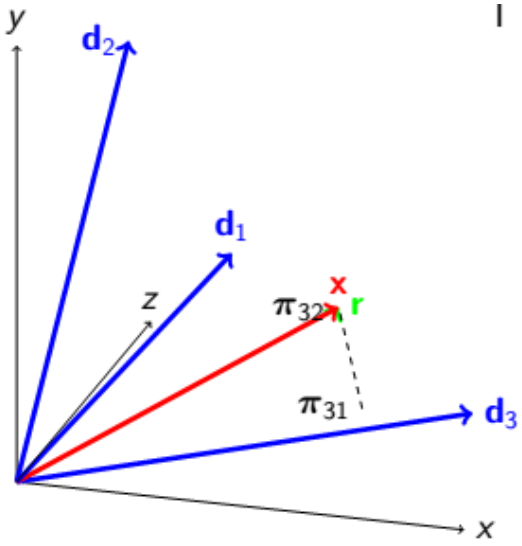
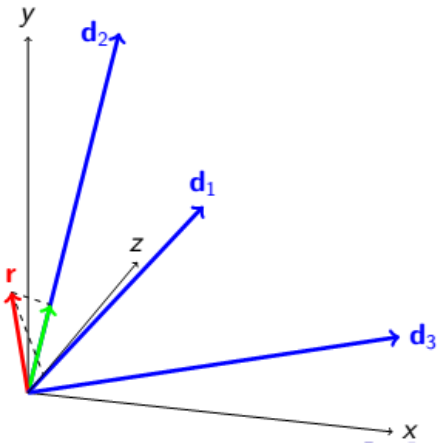
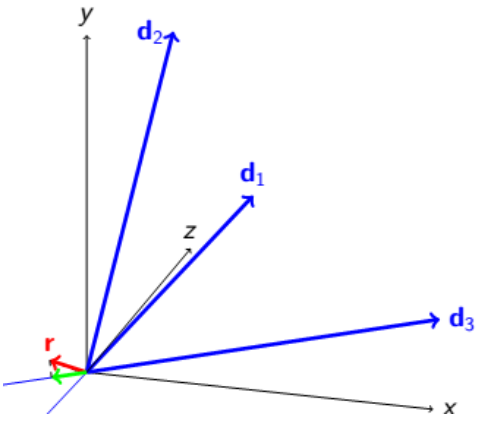
$$\alpha_{\Gamma} \leftarrow (D_{\Gamma}^T D_{\Gamma})^{-1}D_{\Gamma}{}^T x$$

7. **End for**

End

Để rõ hơn, ta sẽ minh họa hai giải thuật tham lam này thông qua ví dụ tìm biểu diễn thưa trong Bảng 1.

Từ Bảng 1, ta nhận thấy không có sự khác biệt giữa OMP và MP trong việc chọn từ đầu tiên. Khi chọn từ thứ 2 cho việc biểu diễn mẫu, MP chọn từ giúp tốt nhất cho biểu diễn phần dư thừa trong khi OMP có thể chọn bất kỳ từ nào trong số từ còn lại bởi mọi cặp từ cùng với d_3 đều cho phép biểu diễn chính xác vectơ r. Giải thuật OMP dừng lại sau bước 2. Tuy nhiên, giải thuật MP vẫn có thể tiếp tục thực hiện. Trong mỗi bước của OMP, ta sẽ cập nhật lại đồng thời hệ số biểu diễn tương ứng với tập các từ tham gia biểu diễn. Trong khi đó, mỗi bước trong MP, ta chỉ cập nhật hệ số tương ứng với từ mới được lựa chọn.

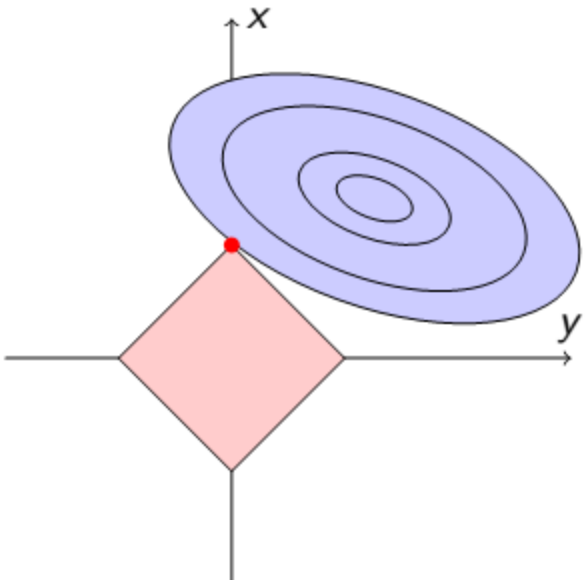
	Giải thuật OMP	Giải thuật MP
Tìm từ 1		
Tìm từ thứ 2		
Tìm từ thứ 3		

Bảng 1: Ví dụ minh họa so sánh giữa MP và OMP

Do chiến lược chọn từ của OMP gần với hàm mục tiêu hơn nên trên thực tế OMP thường đem lại hiệu quả biểu diễn tốt hơn so với MP.

2.3.2. Tìm biểu diễn thưa sử dụng tối ưu với ràng buộc l_1

Việc sử dụng ràng buộc chuẩn l_0 khiến bài toán tìm biểu diễn thưa là bài toán không lồi. Do vậy, một hướng tiếp cận để giải quyết bài toán là sử dụng l_1 thay thế l_0. Chuẩn l_1 biến bài toán thành bài toán lồi, do đó có thể sử dụng các phương pháp tối ưu hàm lồi đã biết để giải quyết. Ở một mức nhất định, chuẩn l_1 cũng thúc đẩy lời giải có dạng vectơ thưa do vậy có thể dùng để thay thế l_0. Hình 7 minh họa hình học cho ý tưởng này. Ta xét vectơ trong không gian 2D. Hình chóp màu hồng biểu diễn norm ball (tập các điểm trên biên hình chóp là tập các điểm mà tại đó giá chuẩn l_1 của chúng bằng nhau). Bài toán mục tiêu có thể xem như việc xấp xỉ hàm mục tiêu bởi điểm trên norm ball. Để xấp xỉ hàm, ta thay đổi tỉ lệ phóng đại của norm ball đến khi norm ball tiếp xúc với giá trị hàm mục tiêu. Tọa độ của đểm tiếp xúc cũng chính là biểu diễn cần tìm. Từ hình vẽ, ta thấy giao điểm có xu hướng cắt tại các điểm trên trục tọa độ hay nói các khác, sử dụng ràng buộc l_1 cũng thúc đẩy yếu tố thưa trong biểu diễn vectơ.



Hình 7: Lời giải với ràng buộc l_1 cho bài toán xấp xỉ hàm

Để giải bài toán tìm biểu diễn thưa sử dụng ràng buộc l_1, tôi sẽ trình bày giải thuật ISTA (Iterative shrinkage-thresholding algorithm).

Hàm mục tiêu của bài toán: $f = \min_X \|Y - DX\|_F^2 + \lambda \|X\|_1$

Đạo hàm của hàm trên : $G = -D^T(Y - DX) + \lambda \text{sign}(X)$

Hàm dấu (sign) tương ứng với một biến số x: $\text{sign}(x) = \begin{cases} 1 & \text{if x} > 0 \\ -1 & \text{if x} < 0 \end{cases}$

Để tiện sử dụng, ta ký hiệu hàm dấu của một ma trận X cũng là sign(X), là một ma trận mà các giá trị phần tử của nó chính là giá trị của hàm dấu của phần tử tương ứng trên ma trận ban đầu.

Giải thuật ISTA hoạt động dựa trên phương pháp ngược đạo hàm (gradient descent). Xuất phát từ một điểm ban đầu, ta dịch chuyển theo hướng ngược đạo hàm để đi đến một điểm mới. Sau một số bước nhất định, quá trình này ngừng lại khi giá trị hàm mục tiêu không tiếp tục giảm. Một điểm lưu ý là hàm mục tiêu với ràng buộc l_1 không tồn tại đạo hàm tại 0. Ta giải quyết vấn đề này thông qua điều chỉnh trong quá trình cập nhật. Trong quá trình cập nhật, khi nhận thấy giá trị của biến đổi dấu, ta đặt nó bằng giá trị 0. Giải thuật ISTA được trình bày trong Giải thuật 2.3:

Giải thuật 2.4: Iterative shrinkage-thresholding algorithm (ISTA)

Input: Từ điển D, mẫu cần biểu diễn, độ thưa T

Output: Vectơ hệ số α tương ứng với x

Begin

1. Khởi tạo X bởi là ma trận **0**
2. **while** X chưa hội tụ :
3. $X^{(t)} \Leftarrow X^{(t)} - \alpha D^T (DX^{(t)} - Y)$
4. $X^{(t)} \Leftarrow \text{shrink}(X^{(t)}, \alpha\lambda)$
5. **End while**
6. Return X

End.

Hàm *shrink* trong công thức trên được được định nghĩa như sau:

$$X_{\text{ij}} = \text{sign}(X_{\text{ij}}) \max(X_{ij} - Y_{ij}, 0) \quad (2.6)$$

2.4. Cập nhật từ điển

Khi cố định ma trận biểu diễn X, ta cần cập nhật lại từ điển D. Bài toán trở thành :

$$D = \underset{D}{\operatorname{argmin}} \|Y - DX\|_F^2 \text{ s.t. } |d_i| = 1 \tag{2.7}$$

Để giải bài toán trên ta có thể sử dụng đạo hàm để tìm lời giải trực tiếp sau đó thực hiện phép chuẩn hóa từ trong từ điển. Tuy nhiên cách cập nhật đó thường không đảm bảo duy trì tính thưa trong biểu diễn của X hiện tại. Đề cập tới vấn đề đó, Elad đề suất giải thuật KSVD [6]. Giải thuật K-SVD cập nhật từ điển theo cơ chế cập nhật từng từ trong khi cố định các từ khác. Cụ thể, khi cập nhật từ thứ *k* trong từ điển, ta có thể biểu diễn lại hàm mục tiêu theo từ thứ *k*:

$$\begin{aligned} E = \|Y - DX\| &= \left\| Y - \sum_{j=1}^{K} d_j x_T^j \right\|_F^2 = \left\| \left(Y - \sum_{j \neq k} d_j x_T^j \right) - d_k x_T^k \right\|_F^2 \\ &= \left\| E_k - d_k x_T^k \right\|_F^2 \end{aligned} \tag{2.8}$$

Đến đây, bài toán đưa về bài toán xấp xỉ ma trận E_k sử dụng ma trận hạng 1. Bài toán này có thể giải quyết sử dụng phân tích SVD [16]. Tuy nhiên lời giải x_T^k có xu hướng không còn đảm bảo tính chất thưa trong biểu diễn thu được ở bước trước. Để khắc phục việc đó, thay vì sử dụng toàn bộ mẫu trong bước cập nhật này, ta chỉ giữ lại các mẫu mà hệ số biểu diễn ứng với từ d_k khác 0, hay tập mẫu mà đang sử dụng d_k để biểu diễn.

Gọi E_k^R là ma trận thu được từ E_k thông qua việc loại bỏ các mẫu mà không sử dụng d_k trong biểu diễn; x_R^k là hệ số tương ứng với E_k^R. Hàm mục tiêu tương đương với việc cực tiểu hóa hàm:

$$f = \left\| E_k^R - d_k x_R^k \right\|_F^2 \tag{2.9}$$

Bài toán trên có thể được giải quyết thông qua phân tích SVD. Sử dụng SVD, ta phân tích ma trận: $E_k^R = \mathrm{U}\Delta\mathrm{V}^T$. Khi đó, ta có lời giải bài toán: d_k là cột đầu tiên của *U*, x_R^k là tích của $\Delta(1,1)$ và vectơ cột thứ nhất của ma trận *V*. Như vậy trong bước cập nhật từ điển theo giải thuật K-SVD, ta đồng thời cập nhật từ điển và vectơ hệ số biểu diễn.

CHƯƠNG III: BÀI TOÁN PHÂN LOẠI ẢNH

1. Bài toán phân loại ảnh

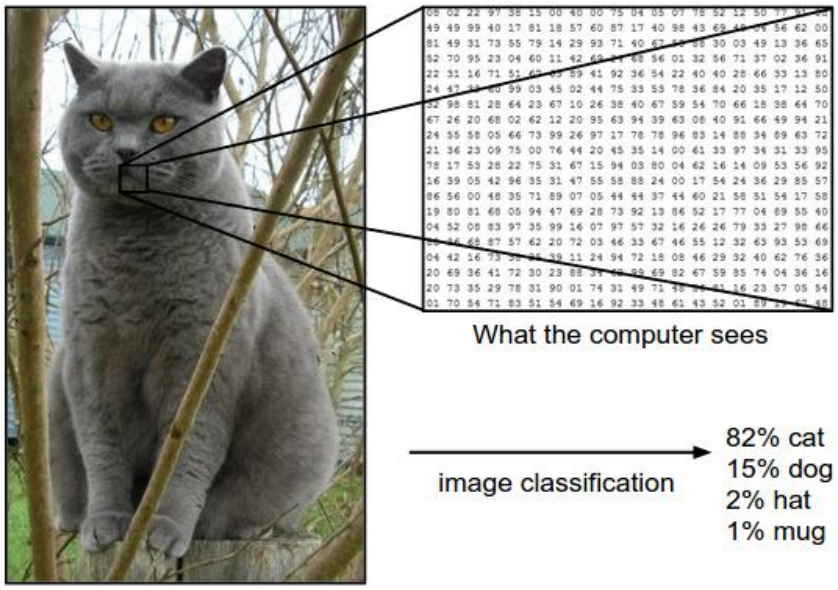
1.1. Giới thiệu bài toán

Phân loại là việc gãn nhãn đối tượng dữ liệu vào một trong các lớp được định nghĩa trước. Bài toán phân loại thuộc lớp bài toán học máy có giám sát.

Ví dụ: phân loại thư điện tử thành thư spam và không phải spam.

Bài toán phân loại ảnh là một lớp bài toán nhỏ hơn của bài toán phân loại khi đối tượng dữ liệu ở đây là ảnh và tập nhãn tương ứng là tập các loại đối tượng. Tuy vậy, bài toán phân loại thực trên thực tế cũng vô cùng đa dạng và phức tạp. Sự đa dạng, phức tạp của bài toán đến từ tập dữ liệu và tập nhãn. Nguồn dữ liệu ảnh có thể là dữ liệu ảnh trên các trang web, dữ liệu ảnh xã hội, ảnh công ty, sách vở,.. Kích thước ảnh có thể biến đổi từ rất nhỏ cho đến rất lớn, số lượng ảnh có thể thay đổi từ hàng trăm cho đến hàng triệu ảnh. Tập nhãn của ảnh cũng có thể thay đổi tùy vào bài toán. Trong một số bài toán cụ thể, tập nhãn có thể liên quan đến những đối tượng cụ thể (những mặt hàng, sản phẩm cụ thể cho đến những bài toán khó khi tập nhãn mang tính trừu tượng hơn.

Để lưu trữ và xử lý, ảnh được số hóa trước khi đưa vào máy tính. Ví dụ, để biểu diễn ảnh trong hệ màu RGB, ảnh được số hóa thông qua 3 ma trận hệ số màu tương ứng với các màu đỏ (R), màu xanh lục (G) và màu xanh lam (B).



Hình 8: Biểu diễn ảnh trên máy tính[2]

[2] Nguồn ảnh: http://cs231n.github.io/classification/

Mặc dù việc nhận dạng đối tượng ảnh với người là việc khá dễ dàng, tuy nhiên để máy có thể hiểu được ảnh là vấn đề không hề đơn giản. Các giải thuật phân loại ảnh tự động thông thường gặp một số thách thức như trong [13]:

- Sự thay đổi góc nhìn: Đối tượng ảnh trong thực tế là các đối tượng ba chiều. Khi chụp ảnh, ta chỉ thể hiện hình chiếu đối tượng trong không gian 2 chiều. Tùy theo góc độ, hình ảnh biểu diễn của đối tượng có thể thay đổi khác nhau.
- Sự thay đổi tỉ lệ:Sự thay đổi kích thước đối tượng trong các ảnh khác nhau
- Biến dạng đối tượng: Trong nhiều trường hợp, đối tượng ảnh không ở hình dạng thường thấy, mà ở dạng đặc biệt, khiến cho việc phân loại đối tượng trở nên khó khăn hơn.
- Đối tượng bị che khuất: Đối tượng không xuất hiện đầy đủ mà chỉ hiển thị một phần.
- Điều kiện ánh sáng khác nhau: điều kiện khác nhau dẫn đến ảnh dữ liệu thu thập về đối tượng có sự biến đổi lớn.
- Nhầm lẫn với môi trường: Đối tượng trong ảnh dễ bị lẫn với môi trường xung quanh.
- Sự đa dạng trong một lớp: Một lớp đối tượng có thể có nhiều biến thể, hình dạng, mẫu mã khác nhau.

Các khó khăn này được mô tả trong Hình 9 :

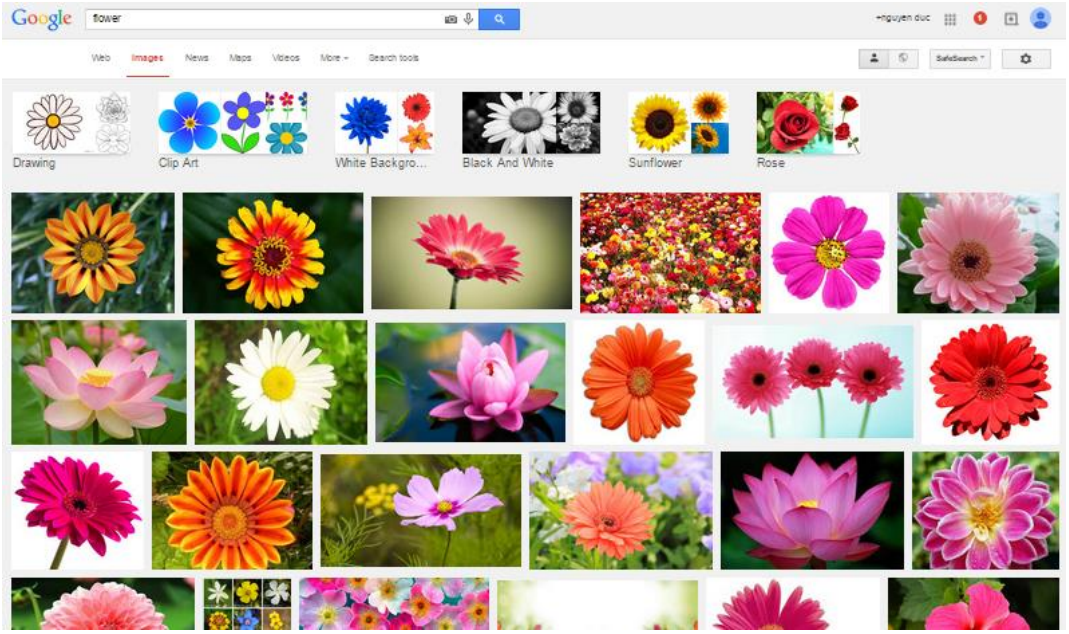
Hình 9: Các yếu tố ảnh hưởng đến việc phân loại ảnh [3]

1.2. Ứng dụng của bài toán phân loại ảnh

Bài toán phân loại đối tượng có ứng dụng rộng rãi trong nhiều lĩnh vực khác nhau:

[3] Nguồn ảnh: http://cs231n.github.io/classification/

- Ảnh được gãn nhãn có thể được sử dụng cho việc tìm kiếm ảnh. Ảnh được phân loại được gán các nhãn phù hợp và việc xử lý tìm kiếm trên hình ảnh thông qua tìm kiếm trên các từ khóa, nhãn phù hợp (Hình 3). Các công cụ tìm kiếm hiện đại như Google, Bing, Baidu,… đều hỗ trợ tính năng tìm kiếm này.



Hình 10: Truy vấn hình ảnh dựa trên từ khóa

- Xây dựng và phát triển thị giác cho robot: phát triển các hệ thống robot có thể nhận biết được đồ vật, khuôn mặt, ứng dụng trong sản xuất, giải trí,…
- Bài toán phân loại ảnh có thể sử dụng để khai thác ảnh ngữ nghĩa.
- Bài toán phân loại ảnh có thể sử dụng làm cơ sở để giải nhiều bài toán khác của thị giác máy tính: xác định vị trí đối tượng, phát hiện đối tượng, phân đoạn đối tượng

1.3. Sơ đồ giải quyết bài toán phân loại ảnh

Bài toán phân loại thường giải quyết theo sơ đồ sau:

Hình 11: Sơ đồ giải quyết bài toán phân loại

Trong sơ đồ này, bước đầu tiên là bước thu thập và tiền xử lý ảnh. Nguồn ảnh đầu vào thường có kích thước, chất lượng khác nhau do điều kiện chiếu sáng. Bước tiền xử lý ảnh này nhằm nâng cao chất lượng ảnh trước khi đưa vào trích chọn đặc trưng. Một số kỹ thuật tiền xử lý có thể sử dụng có thể kể đến như cân bằng histogram, chỉnh kích thước ảnh,…

Bước thứ 2 trong sơ đồ là bước trích chọn đặc trưng. Ảnh đầu vào sẽ qua một bước trích chọn đặc trưng để thu được đặc trưng tiêu biểu cho ảnh. Thông thường, đặc trưng được sử dụng là các đặc trưng được thiết kế trước (hand-designed features). Có thể kể đến một vài đặc trưng tiêu biểu như: SIFT, HoG,… Việc sử dụng các đặc trưng bằng tay thường gặp những hạn chế:

- Cần có hiểu biết tốt về đối tượng: cần hiểu rõ về loại đối tượng, các đặc trưng tiêu biểu cho đối tượng. Mỗi loại đối tượng khác nhau sẽ có thể cần những loại đặc trưng khác nhau.
- Thường tốn thời gian tính toán.
- Không tổng quát cho các bài toán phân loại ảnh nói chung.

Nhằm khắc phục hạn chế này, gần đây cộng đồng học máy giành sự quan tâm lớn với một lĩnh vực học máy mới tên gọi là học biểu diễn (representation learning). Trong lĩnh vực này, ta quan tâm đến các giải thuật (bao gồm các giải thuật học giám sát và không giám sát) nhằm học đặc trưng trực tiếp từ ảnh nguyên gốc nhằm đáp ứng các bài toán học máy cụ thể. Bản thân cách tiếp cận học từ điển được trình bày trong luận văn này cũng là một giải thuật học biểu diễn.

Giai đoạn cuối cùng của sơ đồ này là phân loại. Đặc trưng về ảnh sau khi được trích chọn sẽ được đưa vào các bộ phân loại để học và phân loại để . Một vài bộ học tiêu biểu: SVM, mạng nơron, random forest,…

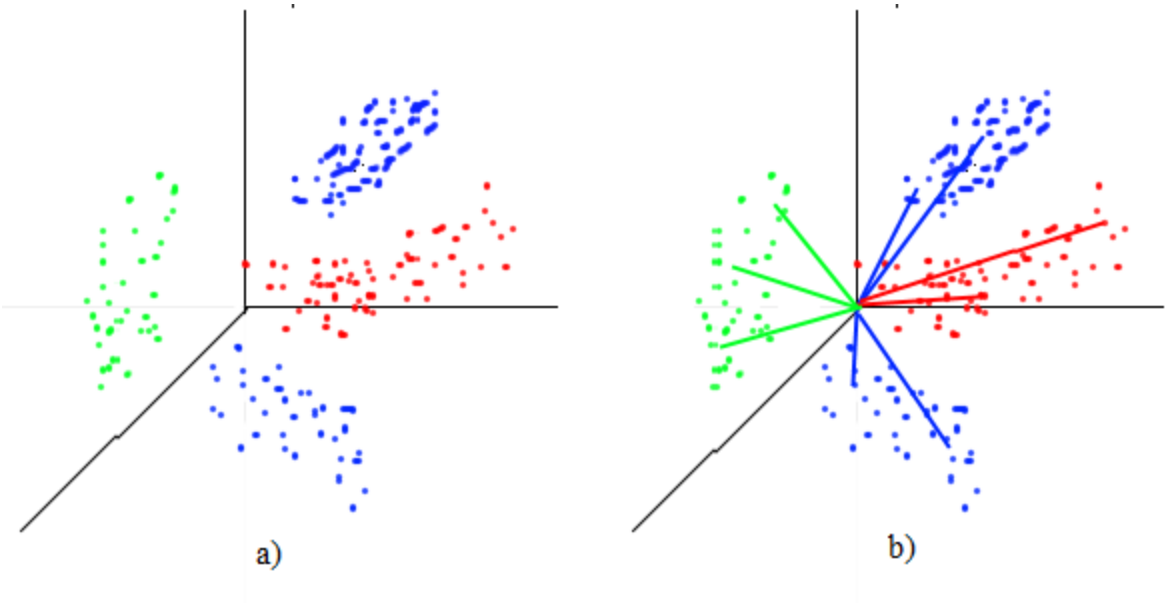
2. Mô hình thưa và học từ điển cho bài toán phân loại ảnh

2.1. Mô hình thưa trong bài toán phân loại

Từ khoảng năm 2006 trở lại đây, học từ điển với biểu diễn thưa trở thành một cách tiếp cận nhận được sự quan tâm lớn của cộng đồng học máy. Mô hình thưa giúp ích cho việc phân loại nói chung và phân loại ảnh nói riêng. Ta sẽ dẫn ra một vài lý do tại sao mô hình này lại phù hợp với bài toán phân loại.

Biểu diễn thưa giúp ánh xạ từ đặc trưng từ không gian ban đầu sang không gian mới cao chiều, phi tuyến tính và dễ phân tách đặc trưng hơn.

Biểu diễn thưa biến đổi dữ liệu từ không gian ít chiều hơn sang không gian nhiều hơn, do đó có khả năng nắm bắt thông tin cấu trúc dữ liệu tốt hơn và phân tách dữ liệu tốt hơn. Ta minh họa vấn đề này thông qua ví dụ dữ liệu như trong Hình 12:



Hình 12: Biểu diễn thưa trong bài toán phân loại

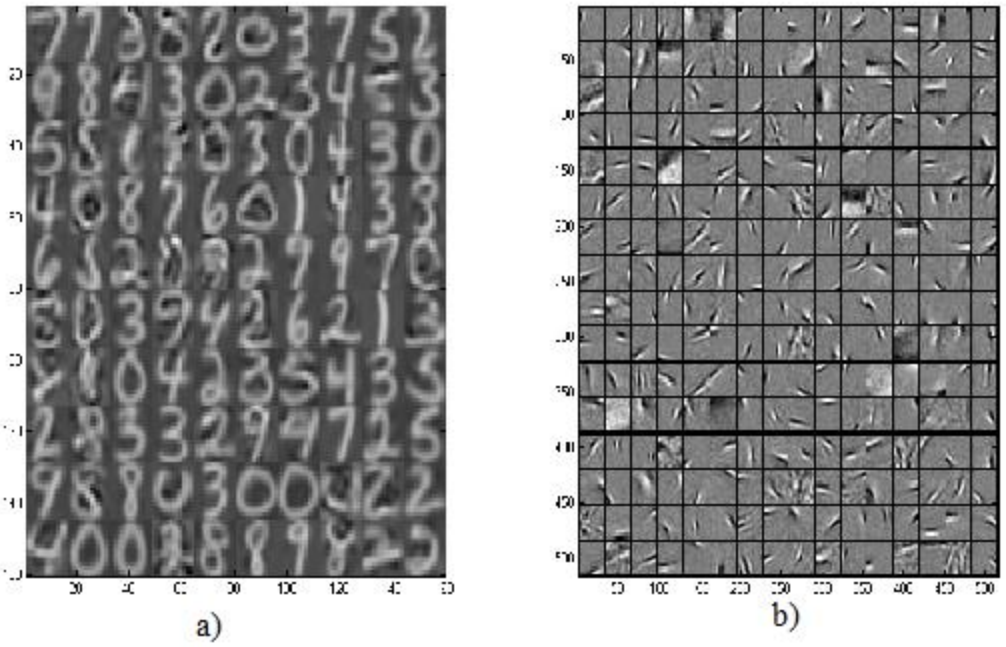
Giả sử tập đặc trưng thu được là tập đặc trưng trong không gian 3 chiều và có phân bố như hình Hình 12.a. Sử dụng các bộ phân loại tuyến tính, ta khó có thể phân tách các mẫu dữ liệu. Tuy nhiên ta có thể thấy tính cấu trúc của dữ liệu: các mẫu dữ liệu của các lớp khác nhau có những phân bố theo những nhóm nhất định. Việc sử dụng biểu diễn thưa với từ điển overcomplete sẽ giúp khai thác được đặc điểm cấu trúc của dữ liệu. Từ điển học được sẽ tương tự như trong Hình 12.b. (mỗi tia trong hình tương ứng với một từ trong từ điển). Trong không gian mới, các tín hiệu có thể được biểu diễn bởi các không gian con tách biệt hoàn toàn nhau (mỗi nhóm tín hiệu được biểu diễn bởi các nhóm từ khác nhau), do vậy dễ dàng trong việc phân tách hơn. Ta cũng minh họa từ điển học được trên dữ liệu thật. Hình 13.a. biểu diễn từ điển học được từ ảnh số trên bộ dữ liệu USPS. Hình 13.b. biểu diễn dữ liệu học được từ các mẫu ảnh thô. Từ điển trên bộ dữ liệu USPS thể hiện rõ khả năng học cấu trúc của từ điển khi các từ học được có dạng như các chữ số thực sự với độ đa dạng cao.

Mô hình thưa tuy có biểu diễn toán học đơn giản nhưng đây là mô hình vô cùng mạnh mẽ bởi tính phi tuyến tính của nó. Tính phi tuyến của mô hình thể hiện ở tính phi tuyến trong tổ hợp tuyến tính của các tín hiệu: tổ hợp tuyến tính của hai tín hiệu trong không gian ban đầu khi sang không gian mới có thể lại được biểu diễn

bởi tập từ hoàn toàn mới, khác với tập từ được sử dụng trong biểu diễn hai tín hiệu ban đầu.

Tính thưa thúc đẩy tính phân biệt của mô hình. Ta lý giải lập luận này thông qua một trường hợp đặc biệt các tín hiệu thuộc các lớp con khác nhau hoàn toàn thuộc các không gian con khác nhau. Khi đó với một từ điển gồm K từ, độ thưa trong vectơ biểu diễn *k*, ta có thể có tới C_K^k không gian con khác nhau, tương ứng với C_K^k loại khuôn mẫu (pattern) đặc trưng khác nhau. Điều đó cho thấy khả năng phân biệt mạnh mẽ của mô hình.

Điểm thứ hai, sử dụng mô hình thưa mềm dẻo hơn và tốt hơn so với cách tiếp cận BoW thông thường như K-means. K-means là một giải thuật không giám sát khá phổ biến dùng để phân cụm dữ liệu. Giải thuật K-means sẽ tìm ra tập K cụm và gán các vectơ tín hiệu vào các cụm gần nhất.Có thể xem K-means như một dạng đặc biệt của mô hình thưa khi chỉ sử dụng một từ cho biểu diễn. Tuy nhiên mô hình thưa không chỉ tổng quát hơn K-means mà giúp nắm bắt cấu trúc dữ liệu tốt hơn. Trở lại ví dụ về việc biểu diễn tín hiệu 3D trong không gian mới cao chiều. Các đặc trưng tương ứng với lớp khác nhau không chỉ có tính cấu trúc nhất định mà ngay cả sự biến đổi của đặc trưng trên từng nhóm tín hiệu cũng có tính cấu trúc nhất định. Khi sử dụng một từ để biểu diễn, ta không thể nắm bắt được sự biến đổi trong từng nhóm tín hiệu này. Mặt khác, các từ trong cùng nhóm vẫn có thể sử dụng các từ khác nhau. Với giá trị độ thưa lớn hơn 1, ta có thể nắm bắt được sự biến đổi trong từng lớp tốt hơn, việc biểu diễn của các mẫu dữ liệu cùng lớp sẽ có xu hướng tương đồng cao.



Hình 13: Tính cấu trúc của từ trong từ điển

2.2. Các nghiên cứu liên quan

Ý tưởng sử dụng biểu diễn thưa trong bài toán phân loại đầu tiên là giải thuật SRC, được đề cập đến trong [1] . Trong cách tiếp cận này, tất cả mẫu dữ liệu được sử dụng làm từ điển biểu diễn. Các mẫu dữ liệu trong cùng một lớp sẽ tạo thành từ điển con cho lớp đó. Mẫu dữ liệu cần kiểm tra sẽ được biểu diễn thưa và phép phân loại được sử dụng độ lỗi biểu diễn.

Mặc dù đem lại kết quả ấn tượng trong bài toán phân loại nhưng SRC gặp một số hạn chế: kích thước từ điển lớn do phải sử dụng toàn bộ mẫu dữ liệu. Một điểm hạn chế khác SRC chỉ hoạt động tốt trong trường hợp số mẫu lớn. Khi số mẫu huấn luyện cho mỗi lớp nhỏ, hiệu quả của giải thuật giảm đáng kể. Do vậy, nhiều giải thuật tập trung vào việc học từ điển giúp thông minh hơn với dữ liệu nhiễu đồng thời đủ mạnh trong việc phân loại. Lấy ý tưởng từ các phương pháp học từ điển không giám sát như K-SVD [6], nhiều đề suất được đưa ra nhằm học từ điển không những có thể xấp xỉ tốt dữ liệu mà còn có tính phân biệt cao.

Dựa trên K-SVD, Q. Zhang đề xuất giải thuật tên là D-KSVD cho bài toán phân loại khuôn mặt [8]. Ý tưởng ở đây là thêm độ lỗi của bộ phân loại tuyến tính vào hàm mục tiêu cần tối ưu. Thông qua đó, từ điển học được không những biểu diễn dữ liệu tốt mà có độ phân biệt cao. D-KSVD sử dụng chính bộ phân loại học được để phân loại.

Zhuolin Jiang đề xuất thêm tính chất nhất quán nhãn (label consistent) [5] trong hàm mục tiêu của bài toán. Từ điển lớn bao gồm các từ điển con cho từng lớp. Tính chất nhất quán nhãn ràng buộc từ điển con có xu hướng biểu diễn dữ liệu lớp đó tốt trong khi biểu diễn dữ liệu lớp khác kém, qua đó nâng cao khả năng phân loại của từ điển.

Các cách tiếp cận học từ điển đề cập ở trên đây tuy ở khía cạnh nhất định giúp ràng buộc tính phân biệt của từ điển tuy nhiên không đề cập đến tính phân biệt trong vectơ biểu diễn. Đề cập tới vấn đề này, Meng Yang [4] đề xuất sử dụng điều kiện Fisher để ràng buộc hệ số biểu diễn, khiến các các vectơ biểu diễn cho các đặc trưng của các đối tượng trong cùng lớp gần nhau trong khi vectơ biểu diễn đặc trưng cho các đối tượng trong các lớp khác nhau sẽ khác xa nhau. Tuy thực nghiệm cho thấy hiệu quả vượt trội của mô hình tuy nhiên trên thực tế thời gian huấn luyện và kiểm tra của giải thuật thường rất lớn, khó đáp ứng yêu cầu thực tế.

Mặc dù học từ điển dựa trên biểu diễn thưa đã chứng tỏ được hiệu quả trong bài toán phân loại tuy nhiên chi phí biểu diễn thưa còn lớn khiến hiệu năng của các giải thuật còn hạn chế. Gần đây việc biểu diễn không thưa bắt đầu được quan tâm và có những kết quả thành công nhất định. L. Zhang [9] phân tích rằng thay vì sử dụng l_0/l_1 cho việc tối ưu, sử dụng l_2 trong những điều kiện nhất định đem lại hiệu quả hơn SRC và có sự tăng tốc về thời gian tính toán đáng kể (từ khoảng 700 đến 1600 lần trong các thực nghiệm của tác giả), từ đó tác giả đề xuất ra mô hình CRC_RLS (Collaborative representation based classification with regularized least square). Ta sẽ đề cập chi tiết về giải thuật trong phần tiếp theo của đồ án.

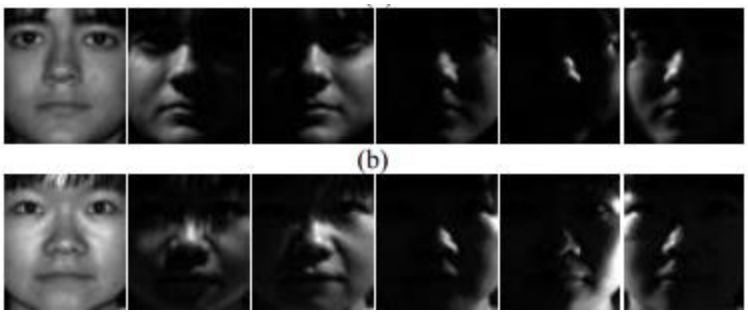
Cũng theo hướng tiếp cận biểu diễn không thưa, trong [4], S. Gu đề xuất mô hình học từ điển DPL. Ý tưởng của tác giả đó là thay vì xây dựng từ điển duy nhất cho cả việc biểu diễn và phân loại, tác giả xây dựng hai từ điển độc lập một từ điển phân tích giúp tăng cường tính phân biệt của mô hình và một từ điển tổng hợp giúp hỗ trợ biểu diễn. DPL đang là mô hình đem lại kết quả tốt nhất theo hướng tiếp cận từ điển cho bài toán phân loại.

CHƯƠNG IV: HỌC TỪ ĐIỂN KHÔNG THƯA CHO BÀI TOÁN PHÂN LOẠI ẢNH

Trong phần này tôi trình bày về mô hình cải tiến cho bài toán học từ điển. Mô hình đề xuất dựa trên ý tưởng của CRC_RLS. Vì vậy trước khi đi vào đề xuất, tôi sẽ trình bày về mô hình CRC_RLS.

1. Giải thuật CRC_RLS

Mô hình CRC_RLS được đề xuất bởi Lei Zhang [9], dựa trên những phân tích về các ràng buộc $l_0/l_1/l_2$ với bài toán phân loại khuôn mặt. Tác giả chỉ ra rằng kể cả việc sử dụng ràng buộc l_2 trong biểu diễn, sử dụng độ lỗi phân loại vẫn có thể đúng. Tiếp đến, tác giả chỉ ra rằng sử dụng biểu diễn thưa như SRC sẽ hiệu quả nếu số lượng mẫu đủ lớn. Trên thực tế việc này khó đạt được do tính đa dạng của góc nhìn khuôn mặt cũng như giới hạn của bộ dữ liệu. Lei Zhang đề xuất ý tưởng biểu diễn cộng tác (collaborative representation) trong biểu diễn mẫu dữ liệu: thay vì chỉ sử dụng từ các mẫu thuộc cùng một lớp (tương ứng từ điển con cho lớp đó), ta có thể sử dụng các mẫu của lớp khác trong việc biểu diễn. Lý do ở đây là các mẫu dữ liệu của lớp khác có thể có nhiều điểm tương đồng với mẫu cần biểu diễn, do vậy có thể sử dụng để hỗ trợ trong biểu diễn mẫu. Ta có thể thấy điều này thông qua một vài ảnh khuôn mặt thuộc hai người khác nhau trong bộ dữ liệu Yaleb (Hình 14).



Hình 14: Mẫu ảnh dữ liệu của hai người khác nhau trên bộ YaleB

Sự tương đồng giữa các mẫu dữ liệu của các lớp khác nhau cho phép các từ điển của các lớp khác nhau hỗ trợ nhau trong biểu diễn mẫu nhưng tính phân biệt của mô hình vẫn được giữ lại do khả năng biểu diễn khác nhau của các từ điển con với mẫu dữ liệu đó: hệ số tương ứng với từ điển con tương ứng với lớp đó cao hơn so với phần hệ số tương ứng với từ điển con thuộc lớp khác. Chính vì vậy, thay vì sử l_0/l_1 cho việc tìm biểu diễn hệ số, tác giả sử dụng l_2 trong ràng buộc hệ số biểu diễn. Đồng thời với đó, việc sử dụng thông tin độ lớn của vectơ hệ số biểu diễn cho việc phân loại, CRC_RLS đem lại cải thiện đáng kể về độ chính xác so với SRC.

Mặt khác, việc sử dụng l_2 trong ràng buộc hệ số khiến bài toán tìm hệ số biểu diễn là bài toán lồi, bài toán có lời giải tường minh. Chính vì vậy, thời gian phân loại của thuật toán được tăng tốc đáng kể. Sơ đồ thuật toán được trình bày trong Giải thuật 4.1:

Giải thuật 4.1: Collaborative Represenation based Classification with Regularized Least square (CRC_RLS)

Input: Tập mẫu huấn luyện, mẫu dữ liệu y, tham số λ

Output: Trả về nhãn cho mẫu y

Begin

1. Chuẩn hóa các mẫu dữ liệu về vectơ đơn vị
2. Tìm hệ số biểu diễn của mẫu y thông qua từ điển X:

$$\alpha = Py \text{ trong đó } P = (X^T X + \lambda I)^{-1} X^T$$

3. Tính toán độ dư thừa biểu diễn trên từ lớp:

$$r_i = \|y - X_i \alpha_i\|_2 / \|\alpha_i\|_2$$

4. Trả về nhãn cho mẫu dữ liệu:

$$\text{identity}(y) = \text{argmin}_i \{r_i\}$$

End.

---

2. Mô hình đề xuất

Tương tự như SRC, CRC_RLS sử dụng toàn bộ mẫu dữ liệu cho việc học từ điển. Do vậy kích thước từ điển rất lớn. Mặt khác, do bỏ qua tính thưa trong biểu diễn, khả năng biểu diễn mẫu dữ liệu của từ điển với một mẫu dữ liệu không phụ thuộc vào một vài từ mà phụ thuộc vào tất cả các từ trong từ điển. Phân bố của mẫu dữ liệu trong tập dữ liệu huấn luyện cũng như việc sử dụng các mẫu dữ liệu tồi có thể làm từ biểu diễn cho lớp có thể làm ảnh hưởng đến độ chính xác. Lưu ý rằng đây chỉ là vấn đề của mô hình biểu diễn với mã "dầy" (dense code). Trong các mô hình phân loại như SRC, điều này không phải vấn đề bởi ta chỉ biểu diễn mẫu qua một vài từ trong từ điển, do vậy việc đa dạng của từ trong từ điển càng giúp biểu diễn chính xác mẫu dữ liệu. Những lý do này dẫn đến nhu cầu học từ điển với ràng

buộc l_2. Tôi gọi tên giải thuật mình đề xuất là NSDL (Non-sparse dictionary learning).

2.1. Hàm mục tiêu

Bài toán đặt ra là xây dựng các từ điển con để biểu diễn dữ liệu lớp đó. Bài toán học từ điển với ràng buộc l_2 trên dữ liệu lớp i có thể được diễn đạt như sau:

$$\underset{D,X}{\operatorname{argmin}} \|Y_i - D_i X_i\|_F^2 + \lambda \|X_i\|_F^2 \text{ s.t. } \forall j, \|d_j\|_2 = 1 \quad (4.1)$$

Do quá trình tối ưu học từ điển là giống nhau giữa các lớp, để đơn giản ký hiệu, ta xem xét bài toán tổng quát, bài toán học từ điển với ràng buộc l_2:

$$\underset{D,X}{\operatorname{argmin}} \|Y - DX\|_F^2 + \lambda \|X\|_F^2 \text{ s.t. } \|d_i\|_2 = 1 \quad (4.2)$$

2.2. Giải thuật tối ưu hóa hàm mục tiêu

Tương tự như bài toán học từ điển đã đề cập trong phần trước, ta cũng áp dụng chiến lược cập nhật từng biến trong khi cố định biến còn lại.

2.2.1. Khởi tạo từ điển

Quá trình khởi tạo từ điển có thể dựa trên khởi tạo ma trận ngẫu nhiên hoặc khởi tạo từ ngẫu nhiên sử dụng tập dữ liệu ban đầu, sau đó chuẩn hóa để đưa về từ điển với tập từ có độ dài đơn vị. Thực nghiệm của tôi cho thấy rằng hai cách khởi tạo cho kết quả tương đương lẫn nhau.

2.2.2. Cập nhật X khi cố định D

Khi cố định D, hàm mục tiêu của bài toán trở thành:

$$X = \underset{X}{\operatorname{argmin}} \|Y - DX\|_F^2 + \lambda \|X\|_F^2 \quad (4.3)$$

Bài toán trên là bài toán lồi. Lời giải của bài toán có thể thu được thông qua việc tìm đạo hàm đối riêng phần với X:

$$X = (D_i^T D_i + \lambda I)^{-1} D_i^T Y_i \quad (4.4)$$

2.2.3. Cập nhật D khi cố định X

Khi cố định X, bài toán trở thành:

$$D = \underset{D}{\operatorname{argmin}} \|Y - DX\|_F^2 \text{ s.t. } \|d_i\|_2^2 = 1 \quad (4.5)$$

Ta áp dụng chiến lược tối ưu từng từ khi cố định các từ còn lại cho việc giải bài toán trên.

Giả sử ta cần cập nhật từ thứ *i* trong từ điển. Khi đó hàm mục tiêu có thể được viết lại:

$$E = \|Y - DX\|_F^2 = \left\|Y - \sum_{j=1}^{K} d_j x_j^T\right\|_F^2 = \left\|Y - \sum_{j<>i}^{K} d_j x_j^T - d_i x_i^T\right\|_F^2 \quad (4.6)$$

Đặt $E_i = Y - \sum_{j<>i}^{K} d_j x_j^T$. E_i biểu diễn độ dư thừa biểu diễn của từ điển khi không sử dụng từ d_i. Khi đó ta có:

$$d_i = \underset{d_i}{\operatorname{argmin}} \left\|E_i - d_i x_i^T\right\|_F^2 \text{ s.t. } \|d_i\| = 1 \quad (4.7)$$

Lời giải của d_i thu được thông qua lấy đạo hàm tương ứng với d_i bỏ qua ràng buộc đơn vị là:

$$d_i = E_i (x_i^T)^T / \left(x_i^T (x_i^T)^T\right) \quad (4.8)$$

Để đảm bảo tính đơn vị của từ trong từ điển, ta cần chuẩn hóa lại từ sau khi thu được bởi bước trên: $d_i = d_i / \|d_i\|$

Cuối cùng, giải thuật học từ điển được trình bày chi tiết trong Giải thuật 4.2:

Giải thuật 4.2: NSDL

Input : - Tập mẫu dữ liệu Y (kích thước m x N)

- kích thước từ điển K, giá trị λ
- số vòng lặp tối đa: *maxIter*

Output : *D, X*

Begin

1. Khởi tạo từ điển:
   - Khởi tạo ma trận ngẫu nhiên kích thước m * K
   - Chuẩn hóa từ điển độ dài đơn vị

---

2. iter = 1
3. **while** iter <= *maxIter*:
    - Update X while fix D: $X = (D_i^T D_i + \lambda I)^{-1} D_i^T Y_i$
    - Update D while fix X:
4. **for** i=1 to K:
5. $E_i = \left\| Y - \sum_{j<>i}^{K} d_j x_j^T \right\|_F^2$
6. $d_i = E_i (x_i^T)^T / \left(x_i^T (x_i^T)^T \right)$
7. $d_i = d_i / \|d_i\|$
8. **End**
9. iter = iter + 1
10. **End while**

End.

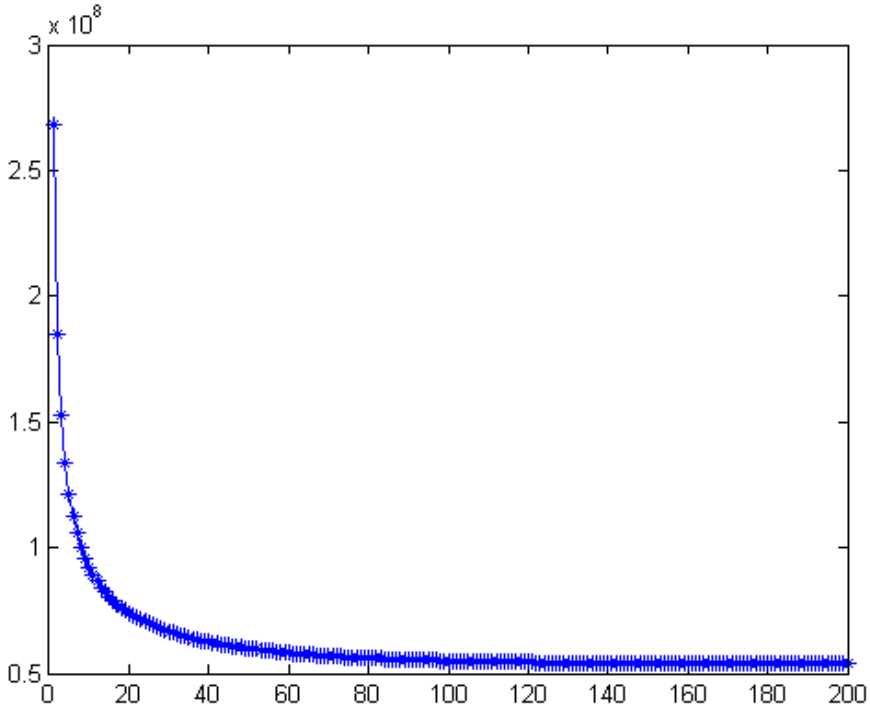
2.3. Phân loại

Ta sử dụng cùng cơ chế phân loại như trong CRC_RLS trong giải thuật đề xuất. Từ điển $D = [D_1, D_2, .., D_L]$ trong đó D_i là từ điển con cho lớp *i*. Ta đặt $P = (D^T D + \lambda I)^{-1} D^T$. Với một mẫu *y*, hệ số biểu diễn cho mẫu *y* là: $\alpha = Py$. Hệ số tương ứng với từ điển con thứ *i* là α_i. Phép gán nhãn được thực hiện như sau:

$$\text{identity}(y) = \arg\min_i \left(\|y - D_i \alpha_i\|_2^2) / \|\alpha_i\|_2^2 \right)$$

2.4. Sự hội tụ của thuật toán

Trong quá trình tối ưu, ta tối ưu hóa đồng thời D và X. Tại mỗi bước cập nhật, ta tối ưu một biến trong khi giữa giá trị của các biến còn loại không đổi, do vậy làm giảm giá trị hàm mục tiêu xuống. Sau một số bước nhất định, quá trình tối ưu dừng lại khi ta gặp một điểm cực tiểu địa phương. Thử nghiệm cho thấy thuật toan hội tụ sau khoảng 50 vòng cập nhật. Hình 15 minh họa giá trị hàm mục tiêu sau từng vòng lặp.



Hình 15: Giá trị hàm mục tiêu sau từng vòng lặp

CHƯƠNG V: KẾT QUẢ THỰC NGHIỆM

Để đánh giá hiệu quả của mô hình, tôi tiến hành một vài thử nghiệm. Thử nghiệm đầu tiên là thử nghiệm về mối quan hệ giữa từ điển kích thước từ điển. Tiếp đến, tôi tiến hành thử nghiệm độ chính xác trên từng bộ dữ liệu khác nhau để kiểm tra độ chính xác giải thuật trên từng bộ dữ liệu khác nhau. Thử nghiệm cuối cùng là thử nghiệm về thời gian phân loại nhằm đánh giá hiệu năng về mặt tốc độ phân loại của giải thuật.

1. Dữ liệu thực nghiệm

Để đánh giá hiệu quả giải thuật trên bài toán phân loại ảnh, tôi tiến hành thử nghiệm trên hai bộ dữ liệu khuôn mặt: YaleB và AR và một bộ dữ liệu đối tượng đồ vật: Caltech-101.

Để có thể so sánh với các giải thuật khác, tôi sử dụng đặc trưng và các thông số như trong các bài [4] [5] để thiết lập cho thử nghiệm của mình.

Với các mẫu dữ liệu về khuôn mặt, tôi sử dụng đặc trưng random face cho các thử nghiệm. Từ vectơ ảnh thô ban đầu thu được thông qua việc "duỗi" ma trận ảnh, ta sẽ ánh xạ chúng sang không gian mới thông qua phép nhân với một ma trận ngẫu nhiên theo phân phối gauss có trung bình 0 và độ lệch chuẩn 1. Với bộ dữ liệu YaleB, chiều của không gian là 504, với bộ dữ liệu AR, chiều của không gian này 540.

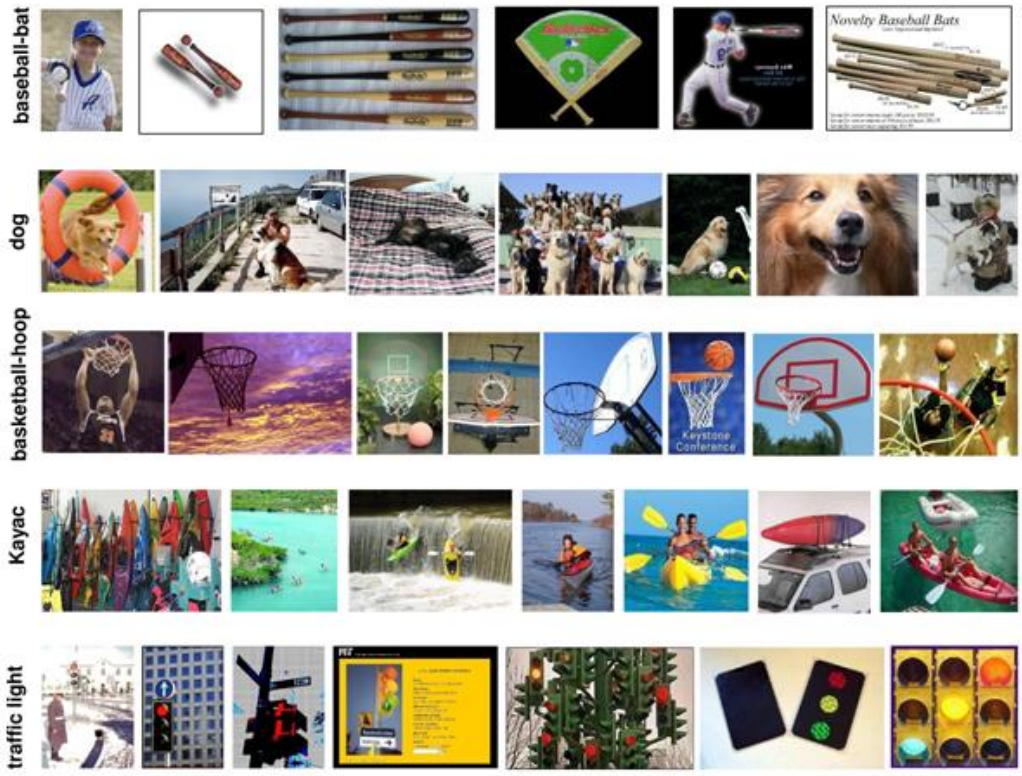
Với bộ dữ liệu Caltech-101, đặc trưng pyramid feature [21] được sử dụng nhằm biểu diễn được tốt hơn đối tượng so với đặc trưng random face ở trên. Cụ thể, từ mỗi ảnh, đặc trưng SIFT sẽ được trích chọn từ các mẫu (patch) ảnh 16 x 16 từ một lưới ảnh có bước nhảy 6; vectơ thu được sẽ được gom cụm thông qua K-means và tiến hành pooling dựa trên các các lưới kích thước 1x1, 2x2 và 4x4. Với mỗi lưới ảnh, ta thu được một vectơ riêng. Vectơ cuối cùng thu được thông qua việc ghép các vectơ đặc trưng lại với nhau. Với giải thuật K-means, số cụm được thiết lập là k=1024. Sau tất cả các bước trên, ta tiến hành giảm chiều dữ liệu còn 3000 sử dụng giải thuật PCA.

Tất cả các đặc trưng này có thể được tìm thấy trên trang:
http://www.umiacs.umd.edu/~zhuolin/projectlcksvd.html .

- **Bộ dữ liệu YaleB face:** Bộ dữ liệu extended YaleB [17] gồm 2414 ảnh khuôn mặt của 38 người, trung bình khoảng 64 ảnh cho mỗi đối tượng. Ảnh gốc ban đầu được cắt lấy phần ảnh chứa mặt ở trung tâm với kích thước

192x168. Ta sử dụng đặc trưng random face với 504 chiều cho bài toán phân loại. Bộ dữ liệu được phân chia tương tự như trong các bài [5], [4], [7] 30 ảnh cho việc huấn luyện, còn lại cho việc kiểm tra.

- **Bộ dữu liệu AR :** Bộ dữ liệu AR [18] gồm 4000 ảnh khuôn mặt của 126 người. Mỗi người có 26 ảnh khuôn mặt. Trong bộ AR, đối tượng khuôn mặt có sự biến về góc chụp, cảm xúc, chiếu sáng và đeo kính. Tôi sử dụng bộ dữ liệu con để làm tập thử nghiệm. Tập thử nghiệm gồm 2600 ảnh của 50 nam và 50 nữ. Từ tập thử nghiệm được chọn ra, tôi sử dụng 20 ảnh mỗi lớp đối tượng để huấn luyện và còn lại cho việc kiểm tra.
- **Bộ dữ liệu Caltech-101:** Caltech-101 [20] là bộ dữ liệu lớn về đối tượng, gồm 101 lớp đối tượng và một lớp nền. Số lượng ảnh trong một lớp dao động từ 31 đến 800. Kích thước ảnh trong tập dữ liệu khoảng 300x200. Tương tự trong cài đặt của các bài báo [5], tôi sử dụng 30 mẫu mỗi lớp cho việc huấn luyện, còn lại cho việc kiểm tra. Ảnh minh họa cho bộ dữ liệu này được thể hiện trong Hình 16.



Hình 16: Bộ dữ liệu Caltech-101

2. Môi trường thực nghiệm

Chương trình được cài đặt bằng ngôn ngữ matlab trên môi trường Ubuntu 14.04. Cấu hình máy tính chạy thử nghiệm như sau:

STT	Phần cứng	Loại
1	CPU	Intel® Xeon(R) CPU E5-2650 v2 @ 2.60GHz (16 core)
2	RAM	32G DDR3 @1333Mhz
3	Ổ cứng	HDD 640G

3. Độ đo

Để đánh giá kết quả của bài toán phân loại, ta sử dụng độ đo độ chính xác. Độ đo chính xác được định nghĩa là tỉ lệ phần trăm số ảnh được gán nhãn đúng trong tổng số ảnh:

$$p = \frac{\text{Tổng số ảnh được gán nhãn đúng}}{\text{Tổng số ảnh}} \quad (\%)$$

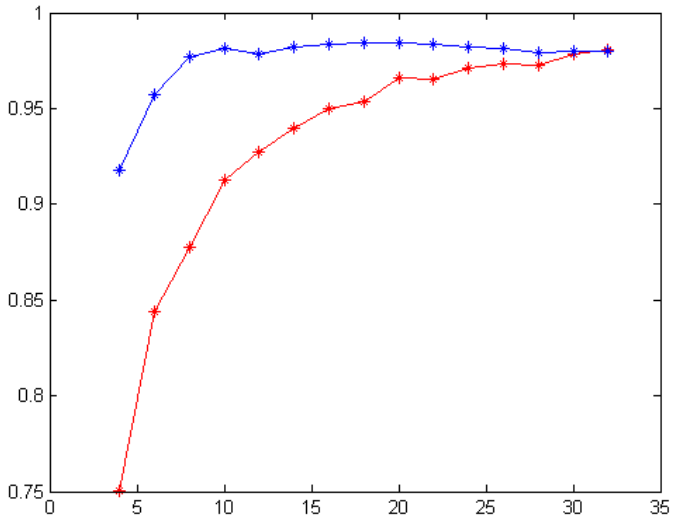
Đơn vị đo của độ chính xác là %.

4. Kết quả thực nghiệm

4.1. Thực nghiệm học từ điển với các kích thước từ điển khác nhau

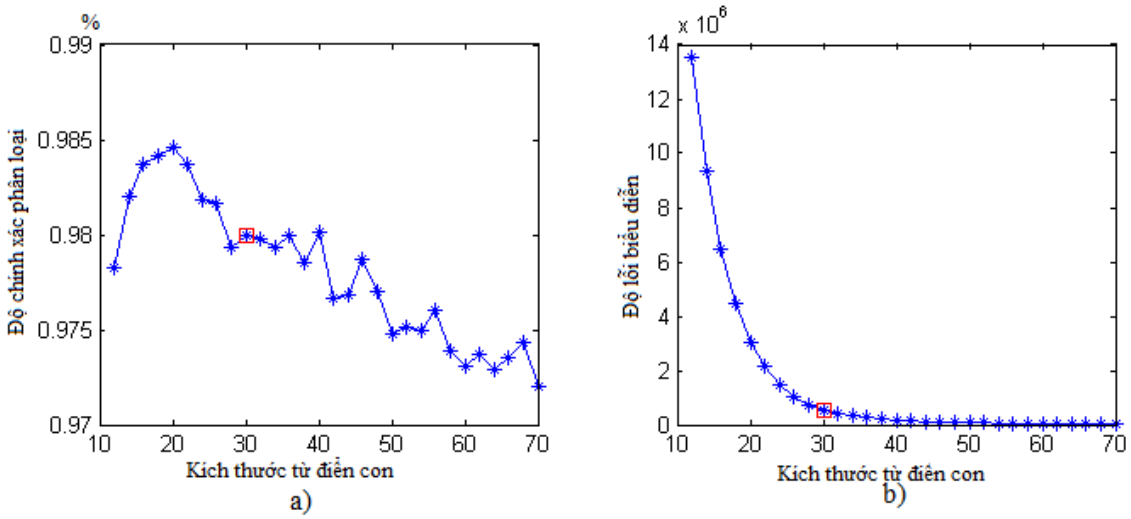
Trong thử nghiệm này, tôi tiến hành thử nghiệm giải thuật trên bộ từ điển với các kích thước từ điển khác nhau để thấy được ảnh hưởng của kích thước từ điển đến khả năng phân biệt của từ điển. Thử nghiệm được tiến hành trên bộ dữ liệu YaleB. Để thấy được ý nghĩa của việc học từ điển, ta so sánh giải thuật với CRC_RLS. Với CRC_RLS, tập mẫu được sử dụng làm từ điển thông qua việc chọn tập ngẫu nhiên một tập con từ tập huấn luyện. Các tham số trong thử nghiệm được thiết lập như sau: kích thước từ điển trên mỗi lớp thay đổi trong phạm vi từ 4 đến đến 32 (số mẫu tối đa trên từng lớp trong tập huấn luyện) với bước nhảy 2. Tham số λ=0.012 được sử dụng cho cả CRC_RLS và giải thuật đề xuất. Kết quả này được thể hiện trong hình Hình 17. Như có thể thấy, khi số từ nhỏ, độ chính xác của CRC_RLS sử dụng khởi tạo ngẫu nhiên giảm mạnh. Trong khi đó đối chiếu với kết quả tương ứng chạy bởi giải thuật NSDL, độ chính xác giảm ít hơn khi kích thước từ điển giảm dần. Lý do ở đây là do việc sử dụng mẫu trực tiếp từ tập mẫu không đại diện tốt cho dữ liệu. Thông qua việc học từ điển, từ điển học được có khả năng

đại diện tốt hơn cho dữ liệu, khả năng biểu diễn mẫu khi số lượng từ trong từ điển giảm nhỏ ít bị ảnh hưởng hơn.



Hình 17: Độ chính xác phân loại với kích thước khác nhau

Đồng thời với thử nghiệm này, tôi tiến hành thêm thử nghiệm đo độ lỗi biểu diễn với của từ điển với các kích thước khác nhau để có những đánh giá sâu hơn. Trong thử nghiệm này, kích thước từ điển được thay đổi từ 4 đến 70 với bước nhảy 2. Ta cũng đo đạt độ chính xác biểu diễn cũng như độ lỗi khi sử dụng CRC_RLS với toàn bộ mẫu. Kết quả độ chính xác cũng như độ lỗi biểu diễn bởi từ điển được thể hiện trong Hình 18. Ô vuông màu đỏ thể hiện kết quả chạy bởi CRC_RLS, trong khi đường màu xanh thể hiện kết quả chạy sử dụng giải thuật đề xuất. Như trong hình vẽ thể hiện, ta có hai nhận xét. Thứ nhất, độ chính xác cũng như độ lỗi biểu diễn khi sử dụng toàn bộ mẫu khớp với độ chính xác và độ lỗi khi sử dụng giải thuật đề xuất với kích thước bằng kích thước tập mẫu. Thứ hai, độ chính xác tăng khi kích thước từ điển tăng đến mức nhất định (khoảng 20) và giá trị này nhỏ hơn kích thước tập mẫu và bắt đầu giảm khi kích thước từ điển tiếp tục tăng. Nhận xét thứ nhất chỉ ra rằng, việc học từ điển giúp tổng quát hóa cho dữ liệu. Nhận xét thứ hai lý giải tác dụng của việc học từ điển trong việc hạn chế các yếu điểm của việc sử dụng dữ liệu như từ điển trong biểu diễn không thưa đã đề cập ở trên. Khi quan sát đồ thị về độ lỗi của biểu diễn giảm khi kích thước từ điển tăng gợi ý rằng tính phân biệt của từ điển giảm khi kích thước từ điển đủ lớn và tăng dần. Bằng việc cân bằng giữa khả năng biểu diễn và độ phân biệt của từ điển, ta có thể đạt được kết quả phân loại tốt nhất cho bài toán phân loại.



Hình 18: Độ chính xác và độ lỗi biểu diễn với các kích thước từ điển khác nhau.

4.2. Thực nghiệm độ chính xác trên các bộ dữ liệu khác nhau

Trong thử nghiệm về độ chính xác, tôi thử nghiệm mô hình trong hai điều kiện. Trường hợp thứ nhất, tôi giữ kích thước từ điển cần học bằng với kích thước từ điển trong các bài báo liên quan dùng để so sánh và và tối ưu tham số λ. Tôi gọi mô hình này là NSDL-1. Trường hợp hai, cả tham số kích thước từ điển λ cùng được tối ưu để đạt được kết quả tốt nhất của mô hình đề xuất. Tôi ký hiệu mô hình với cài đặt này bởi ký hiệu NSDL-2. Các tham số trong học từ điển được chọn thông qua kiểm chứng chéo (cross-validation). Kết quả thu được là kết quả trug bình sau 5 lần chạy trên mỗi bộ dữ liệu với các cách chia dữ liệu ngẫu nhiên khác nhau.

Với NSDL-1, tham số thử nghiệm như sau:

Bộ dữ liệu	Kích thước từ điển con	Kích thước từ điển	Giá trị λ
YaleB	15	570	0.01
AR	20	2000	0.012
Caltech-101	30	3060	6.5

Với NSDL-2, các tham số được thiết lập như sau như trong bảng:

Bộ dữ liệu	Kích thước từ điển con	Kích thước từ điển	Giá trị λ
YaleB	15	570	0.01
AR	20	2000	0.012

Caltech-101	26	2652	6.5

Kết quả thử nghiệm với các bộ dữ liệu được trình bày trong các bảng 2, 3, 4.

Bảng 2: Bảng so sánh kết quả độ chính xác của các giải thuật trên bộ dữ liệu YaleB

Giải thuật	Độ chính xác (%)
NSC	94.7
SVM	95.6
SRC	96.5
DLSI	97.0
FDDL	96.7
LCKSVD	96.7
MMDL	97.3
DPL	97.5
CRC	96.5
CRC (toàn bộ mẫu)	98.1
NSDL-1	**98.5**
NSDL-2	**98.5**

Bảng 3: Bảng so sánh kết quả độ chính xác của giải thuật khác nhau trên bộ AR

Giải thuật	Độ chính xác (%)
NSC	92.0
SVM	96.5
SRC	97.5
DLSI	97.5
FDDL	97.5
LCKSVD	97.8
MMDL	97.3
DPL	**98.3**
CRC	98.0
NSDL-1	97.6

NSDL-2	97.6

Bảng 4: Bảng so sánh kết quả độ chính xác của giải thuật khác nhau trên bộ Caltech-101

Giải thuật	Độ chính xác (%)
NSC	70.1
SVM	64.6
DLSI	73.1
FDDL	73.2
LC-KSVD	73.6
DPL	73.9
CRC (toàn bộ dữ liệu)	75.5
NSDL-1	76.4
NSDL-2	**76.5**

Như có thể thấy, trong 2 thử nghiệm với bộ dữ liệu YaleB, Caltech-101 giải thuật NSDL cho kết quả cao vượt trội so với các giải thuật còn lại. Với bộ dữ liệu AR, kết quả thấy hơn so với giải thuật DPL và LC-KSVD. Để ý rằng với bộ dữ liệu AR, kích thước từ điển tối ưu cũng chính bằng kích thước tập huấn luyện. Điều đó dẫn đến một lý giải cho kết quả này có thể vì dữ liệu cho từng lớp không đủ biểu diễn tốt được dữ liệu.

4.3. Thực nghiệm về thời gian phân loại

Trong thử nghiệm này, tôi tiến hành thử nghiệm thời gian tính toán của NSDL so với các phương pháp SRC, CRC_RLS, DPL. Thời gian đo đạt ở đây là thời gian trung bình để phân loại một đối tượng, tính bằng (s/ảnh). Kết quả được thống kê lại trong bảng Bảng 5.

Bảng 5: Bảng so sánh thời gian phân loại (s/ảnh) trên các bộ dữ liệu

	SRC	CRC_RLS	DPL	NSDL
YaleB	0.005	0.000096	0.000088	**0.000063**
AR dataset	0.0077	0.000221	**0.000220**	0.000221
Caltech-101	0.012	0.001034	0.001003	**0.000944**

Như có thể thấy trong bảng, khi kích thước từ điển nhỏ hơn kích thước tập mẫu (với bộ từ điển YaleB, Caltech-101), thời gian tính toán giảm đáng kể so với sử dụng toàn bộ mẫu, tỉ lệ thuận với kích thước từ điển học được. Với hai bộ AR và

Caltech-101, không có sự khác biệt lớn giữa 3 giải thuật CRC_RLS, DPL và NSDL khi toàn bộ mẫu được sử dụng. Kết quả này khác với trong báo cáo [4] khi trong thử nghiệm của mình mà tác giả cho thấy thời gian tính toán giữa CRC_RLS và DPL chênh nhau 5 đến 10 lần. Nguyên nhân ở đây có thể đến từ việc tối ưu mã trong quá trình cài đặt. Từ Bảng 5, ta cũng có thể nhận thấy có sự cải thiện đáng kể khi chuyển điều kiện ràng buộc từ ràng buộc thưa sang ràng buộc không thưa khi thời gian tính toán so với SRC cho mỗi ảnh tăng tốc từ 10 cho đến 80 lần, tùy theo kích thước từ điển học được.

CHƯƠNG VI: KẾT LUẬN

1. Đánh giá

1.1. Các kết quả đạt được

Về mặt lý thuyết, đồ án đã trình bày được những nội dung cơ bản sau:

- Các khái niệm cơ bản về học máy, lý thuyết về không gian vectơ và bài toán học từ điển, bài toán phân loại ảnh sử dụng cách tiếp cận học từ điển
- Đề xuất mô hình học từ điển sử dụng ràng buộc l_2 giúp cải thiện cả kết quả nhận dạng và thời gian tính toán.

Về mặt thực nghiệm, đồ án đã thu được một số kết quả sau:

- Cài đặt thành công mô hình đề xuất
- Thử nghiệm mô hình trên các bộ dữ liệu khác nhau gồm có AR, yaleB, Caltech-101. Thử nghiệm cho thấy kết quả vượt trội của mô hình so với các cách tiếp cận hướng học từ điển về độ chính xác cũng như hiệu quả về mặt thời gian của thuật toán.

1.2. Hạn chế

Do thời gian hạn chế, tôi chưa thử nghiệm được với các bộ dữ liệu khác, có thêm các thử nghiệm để đánh giá và phân tích sâu hơn về giải thuật đề xuất.

2. Phương hướng phát triển

Thành công trong việc học từ điển với chuẩn l_2 gợi mở cho những hướng mới về học từ điển thay thế học từ điển sử dụng ràng buộc biểu diễn thưa. Trong tương lai, tôi hướng đến phát triển giải thuật theo một số hướng:

- Thêm các ràng buộc mạnh vào bài toán: Trong mô hình đề xuất học từ điển, không có yếu tố được thêm vào để thúc đẩy tính phân biệt của mô hình. Trong tương lai, tôi tập trung nghiên cứu các yếu tố phù hợp cho phương pháp học từ điển với ràng buộc l_2.
- Cải thiện kiến trúc từ điển: nhiều kiến trúc từ điển như cây được đề xuất với ràng buộc l_2. Tôi hi vọng có thể áp dụng trong tiếp cận ràng buộc l_2.

TÀI LIỆU THAM KHẢO

1. Wright, John, et al. "Robust face recognition via sparse representation." *Pattern Analysis and Machine Intelligence, IEEE Transactions* 31.2:210-227, 2009.
2. Foldiak, P. Sparse coding, http://www.scholarpedia.org/article/Sparse_coding , last visited May 2015.
3. Davenport, Mark A., et al. "Introduction to compressed sensing." *Preprint* 93, 2011.
4. S. Gu, L. Zhang, W. Zuo, and X. Feng. "Projective dictionary pair learning for pattern classification". *In Advances in Neural Information Processing Systems, pages 793–801, 2014*.
5. Jiang, Zhuolin, Zhe Lin, and Larry S. Davis. "Label consistent k-svd: learning a discriminative dictionary for recognition." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 35.11: 2651-2664, 2013.
6. Aharon, Michal, Michael Elad, and Alfred Bruckstein. K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation." *Signal Processing, IEEE Transactions on* 54.11: 4311-4322, 2006.
7. Yang, Meng, David Zhang, and Xiangchu Feng. "Fisher discrimination dictionary learning for sparse representation." *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011.
8. Zhang, Qiang, and Baoxin Li. "Discriminative K-SVD for dictionary learning in face recognition." *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010.
9. Zhang, Lei, et al. "Collaborative representation based classification for face recognition." *arXiv preprint arXiv:1204.2358*, 2012.
10. Olshausen, Bruno A., and David J. Field. "Sparse coding with an overcomplete basis set: A strategy employed by V1?." *Vision research* 37.23 (1997): 3311-3325.
11. Nguyễn Quốc Việt, Nguyễn Cảnh Lương. Đại số tuyến tính, Nhà xuất bản khoa học và kỹ thuật Hà Nội, 2005.
12. Bishop, Christopher M. *Pattern recognition and machine learning*. Vol. 4. No. 4. New York: springer, 2006.
13. Karpathy, Andrej. Image classification, http://cs231n.github.io/classification/, last visited May 2015.

14. Feature learning, http://en.wikipedia.org/wiki/Feature_learning, last visited May 2015.

15. Visual cortex, http://en.wikipedia.org/wiki/Visual_cortex, last visited May 2015.

16. Singular value decomposition, http://en.wikipedia.org/wiki/Singular_value_decomposition, last visited May 2015.

17. A. Georghiades, P. Belhumeur, and D. Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643-660, 2001.

18. Martinez, A., Benavente., R.: The ar face database. *CVC Technical Report* 1998.

19. S. Lazebnik, C. Schmid, and J. Ponce, "Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories", *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2007.

20. L. FeiFei, R. Fergus, and P. Perona, "Learning Generative Visual Models from Few Training Samples: An Incremental Bayesian Approach Tested on 101 Object Categories", *Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshop Generative Model Based Vision*, 2004.

21. S. Lazebnik, C. Schmid, and J. Ponce, "Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Cate- gories," *Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2007*.