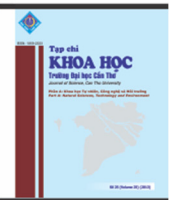




Tạp chí Khoa học Trường Đại học Cần Thơ
website: sj.ctu.edu.vn



PHÂN LỚP ẢNH VỚI GIẢI THUẬT GIẢM GRADIENT NGẪU NHIÊN ĐA LỚP

Đỗ Thanh Nghị¹ và Phạm Nguyên Khang¹

¹ Khoa Công nghệ Thông tin & Truyền thông, Trường Đại học Cần Thơ

Thông tin chung:

Ngày nhận: 17/04/2013

Ngày chấp nhận: 24/12/2013

Title:

Classifying images with
multiclass stochastic gradient
descent algorithm

Từ khóa:

Biểu diễn đặc trưng không
đổi SIFT, Mô hình túi từ
BoW, Máy học véc tơ hỗ trợ
SVM, Phương pháp giảm
gradient ngẫu nhiên SGD

Keywords:

Scale-Invariant Feature
Transform - SIFT, Bag-of-
Words - BoW, Support Vector
Machines - SVM, Stochastic
Gradient Descent - SGD

ABSTRACT

In this paper, we present a new algorithm, MC-SGD (Multiclass Stochastic Gradient Descent), to effectively classify multiclass images. The representation of the images is based on the bag-of-words (BoW), which is constructed from the local descriptors (the Scale-Invariant Feature Transform method – SIFT). The pre-processing step brings out datasets with a very large number of dimensions. We propose a new algorithm called MC-SGD that is suited for classifying very-high-dimensional datasets. The numerical test results on a real dataset showed that our algorithm MC-SGD outperforms Support Vector Machines (SVM) using non-linear kernel functions (Radial Basis Function - RBF).

TÓM TẮT

Trong bài này, chúng tôi trình bày giải thuật mới, giảm gradient ngẫu nhiên (Multiclass Stochastic Gradient Descent - MC-SGD), cho phân lớp hiệu quả dữ liệu ảnh đa lớp. Tập dữ liệu ảnh biểu diễn ảnh bằng mô hình túi từ (Bag-of-Words - BoW) sử dụng các nét đặc trưng không đổi với những biến đổi tỉ lệ (Scale-Invariant Feature Transform - SIFT), dựa trên đặc trưng cục bộ, không bị thay đổi trước những biến đổi tỉ lệ ảnh, tịnh tiến, phép quay, không bị thay đổi một phần đối với phép biến đổi hình học affine (thay đổi góc nhìn) và mạnh với những thay đổi về độ sáng, sự nhiễu và che khuất. Chúng tôi đề nghị một giải thuật phân lớp đa lớp mới, giảm gradient ngẫu nhiên MC-SGD, cho phép phân lớp hiệu quả dữ liệu có số chiều lớn thu được từ bước biểu diễn ảnh. Kết quả thực nghiệm trên tập dữ liệu thực cho thấy giải thuật MC-SGD phân lớp nhanh, chính xác hơn khi so sánh với giải thuật máy học véc tơ hỗ trợ (Support Vector Machines - SVM) sử dụng hàm nhân phi tuyến (Radial Basis Function - RBF).

1 GIỚI THIỆU

Phân lớp ảnh là gán nhãn tự động cho từng ảnh theo chủ đề đã được định nghĩa trước dựa vào nội dung của ảnh. Phân lớp ảnh có nhiều ứng dụng trong thực tế như nhận dạng chữ số trên chi phiếu ngân hàng, mã số trên bì thư của dịch vụ bưu chính, hay các chữ số trên các biểu mẫu nói chung, đặc biệt là tổ chức nội dung trang web một cách tự động bằng cách đánh nhãn tự động ảnh.

Hệ thống phân lớp ảnh thường bao gồm hai bước: rút trích đặc trưng từ nội dung ảnh và huấn luyện mô hình máy học để gán nhãn tự động từ các đặc trưng này. Hiệu quả của hệ thống phân lớp phụ thuộc vào các phương pháp sử dụng ở hai giai đoạn trên.

Các nghiên cứu trước đây (LeCun *et al.*, 98), (Viola & Jones, 01), (Zheng & Daoudi, 04) sử dụng tiếp cận rút trích đặc trưng dựa trên phát hiện của các điểm, màu sắc, kết cấu (texture), tổ chức

đồ (histogram). Mạng nơ-ron (neural networks), máy học véc tơ hỗ trợ (support vector machines), giải thuật boosting được huấn luyện để phân lớp ảnh.

Gần đây, một hướng tiếp cận của (Bosch *et al.*, 06) dựa vào phương pháp biểu diễn ảnh bằng các nét đặc trưng không đổi với những biến đổi tỉ lệ SIFT (Lowe, 04) và mô hình túi từ BoW. Đặc trưng cục bộ SIFT không bị thay đổi trước những biến đổi tỉ lệ ảnh, tịnh tiến, phép quay, không bị thay đổi một phần đối với phép biến đổi hình học affine (thay đổi góc nhìn) và mạnh với những thay đổi về độ sáng, sự nhiễu và che khuất. Một ảnh được biểu diễn bởi tập hợp túi từ được xây dựng bằng cách áp dụng một giải thuật gom nhóm lên các véc tơ mô tả cục bộ SIFT. Giai đoạn tiền xử lý cho ra một tập dữ liệu với số chiều rất lớn. Tác giả đề xuất giảm chiều dữ liệu và sử dụng k láng giềng để phân lớp hiệu quả ảnh.

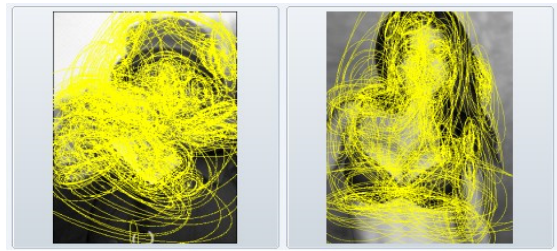
Chúng tôi đề xuất sử dụng ý tưởng của phương pháp biểu diễn ảnh bằng đặc trưng không đổi SIFT và mô hình túi từ. Tuy nhiên thay vì giảm chiều thì chúng tôi đề xuất một giải thuật học mới, giảm gradient ngẫu nhiên MC-SGD, cho phép phân lớp hiệu quả dữ liệu có số chiều lớn thu được từ bước biểu diễn ảnh. Kết quả thực nghiệm trên tập dữ liệu ảnh thực từ ImageNet (Deng *et al.*, 12) chỉ rằng giải thuật mới MC-SGD phân lớp nhanh, chính xác khi so sánh với giải thuật máy học véc tơ hỗ trợ SVM (Vapnik, 95) sử dụng hàm nhân phi tuyến (Radial Basis Function - RBF).

Phần tiếp theo của bài viết được trình bày như sau: phần 2 trình bày ngắn gọn về biểu diễn ảnh bằng mô hình túi từ của đặc trưng cục bộ không đổi, phần 3 trình bày giải thuật phân lớp đa lớp MC-SGD. Phần 4 trình bày các kết quả thực nghiệm tiếp theo sau đó là kết luận và hướng phát triển.

2 BIỂU DIỄN ĐẶC TRƯNG KHÔNG ĐỔI VÀ MÔ HÌNH TÚI TỪ

Biểu diễn ảnh là một bước quan trọng trong phân loại ảnh. Bước này có ảnh hưởng rất lớn đến kết quả phân loại cuối cùng. Trong lãnh vực phân lớp và tìm kiếm ảnh, đặc trưng cục bộ SIFT (Lowe, 04) là những điểm đặc trưng, vùng đặc trưng để biểu diễn ảnh rất hiệu quả, ngày càng trở nên phổ biến. Nghiên cứu tiên phong của (Bosch *et al.*, 06) đề xuất hệ thống phân lớp ảnh dựa trên đặc trưng SIFT và mô hình túi từ (xuất phát từ ý tưởng phân lớp văn bản). Giai đoạn biểu diễn ảnh theo mô hình túi từ và đặc trưng không đổi SIFT bao gồm 3 bước chính: (i) phát hiện và biểu diễn các nét đặc trưng

cục bộ, (ii) xây dựng từ điển các từ trực quan và (iii) biểu diễn ảnh dưới dạng véc tơ tần số xuất hiện các từ trực quan trong ảnh.

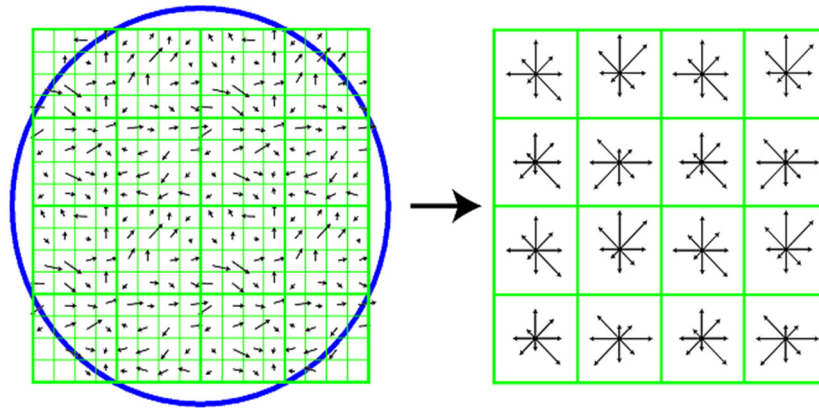


Hình 1: Các điểm đặc trưng được phát hiện bởi thuật toán Hessian-Affine

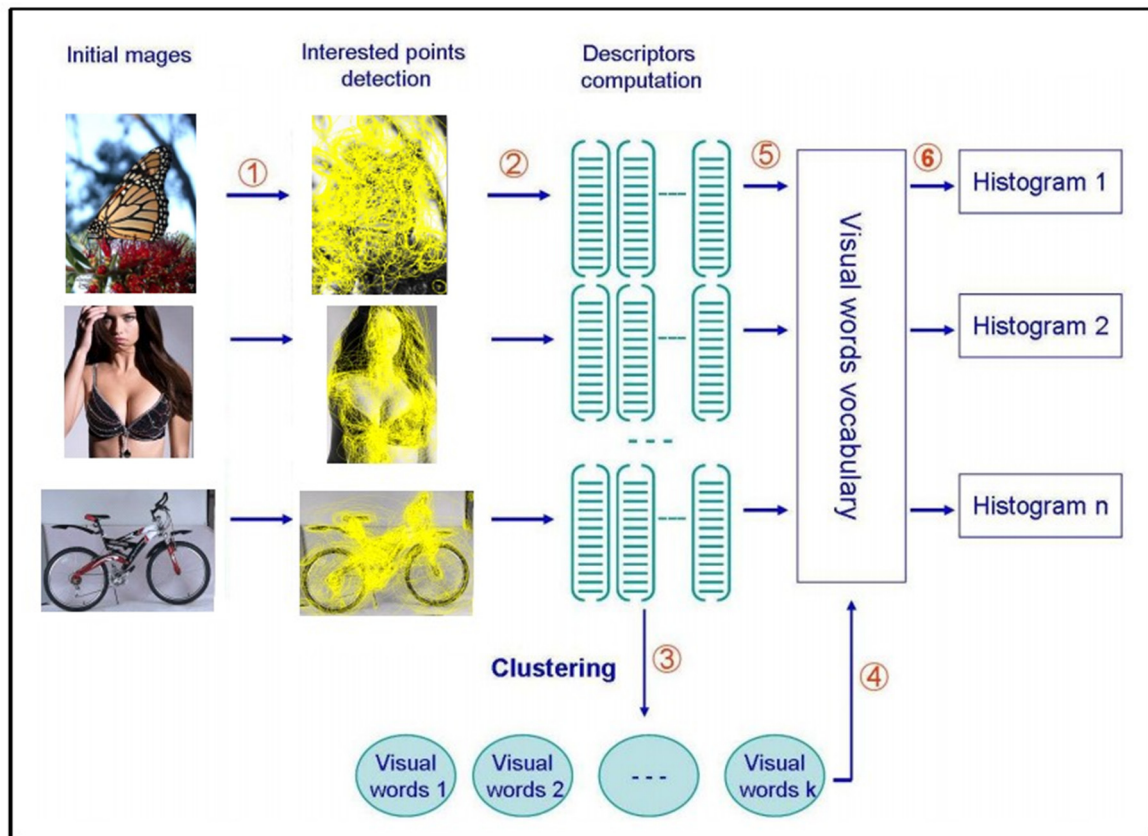
Ở bước đầu tiên, ảnh được đưa về dạng mức xám. Các điểm đặc trưng (Hình 1) được tính trên những ảnh này bằng cách sử dụng các giải thuật phát hiện điểm đặc trưng cục bộ (local feature detector) như là Harris-Affine, Hessian-Affine (Mikolajczyk & Schmid, 04). Những điểm đặc trưng này có thể là cực trị cục bộ của phép toán DoG (Difference of Gaussian) hoặc là cực đại của phép toán LoG (Laplace of Gaussian). Sau đó, vùng xung quanh các điểm đặc trưng được xác định và mô tả bằng các véc tơ mô tả cục bộ. Véc tơ mô tả SIFT được đánh giá rất cao bởi giới chuyên môn trong việc biểu diễn các vùng xung quanh điểm đặc trưng bởi vì nó không đổi đối với những biến đổi tỉ lệ, tịnh tiến, phép quay, và không đổi một phần với đối với những thay đổi về góc nhìn, đồng thời nó cũng rất mạnh với những thay đổi về độ sáng, sự che khuất, nhiễu.

Hình 2 minh họa một ví dụ của véc tơ mô tả SIFT được xây dựng từ vùng cục bộ xung quanh một điểm đặc trưng. Mỗi véc tơ mô tả là một ma trận 4×4 các tổ chức đồ. Mỗi tổ chức đồ có 8 khoảng tương ứng với 8 hướng. Do đó, mỗi véc tơ mô tả SIFT là một véc tơ $4 \times 4 \times 8 = 128$ chiều. Lúc này, mỗi ảnh được biểu diễn bởi một tập các véc tơ mô tả SIFT.

Bước kế tiếp là thiết lập các từ trực quan từ các mô tả cục bộ đã được tính ở bước trước. Thuật giải *k*-means (MacQueen, 67) được thực hiện trên các véc tơ mô tả để phân các véc tơ SIFT thành vào các nhóm (cluster) và mỗi cluster tương ứng với một từ trực quan. Tập các cluster này tạo thành một từ điển. Sau cùng, mỗi véc tơ mô tả trong ảnh sẽ được gán vào cluster gần nhất (dựa vào khoảng cách mỗi véc tơ đến các tâm của các cluster đại diện đã được định nghĩa trước đó). Tiếp theo, một ảnh sẽ được biểu diễn bằng tần số của các từ trực quan trong ảnh. Hình 3 mô tả các bước tạo mô hình BoW biểu diễn các ảnh.



Hình 2: Đặc trưng cục bộ SIFT được tính toán từ vùng xung quanh điểm đặc biệt (vòng tròn): gradient của ảnh (trái), véc-tơ mô tả (phải)



Hình 3: Tạo mô hình BoVW để biểu diễn ảnh

3 THUẬT GIẢI GIẢM GRADIENT NGẪU NHIÊN (MC-SGD)

Giai đoạn tiền xử lý cho một tập dữ liệu với số chiều rất lớn (ví dụ, 50000 từ trực quan với nhiều đặc trưng đầu vào với mỗi đặc trưng chỉ chứa ít thông tin cho phân lớp). Giải thuật máy học véc tơ

hỗ trợ SVM (Vapnik, 95) là mô hình hiệu quả và phổ biến cho vấn đề phân lớp những tập dữ liệu có số chiều lớn. Xuất phát từ cài đặt hiệu quả giải thuật SVM bằng phương pháp giảm gradient ngẫu nhiên SGD (Bottou & Boussquet, 08), chúng tôi phát triển giải thuật MC-SGD cho phân lớp đa lớp tập dữ liệu có số chiều lớn này.

3.1 Giải thuật máy học véc tơ hỗ trợ SVM

Xét ví dụ phân lớp nhị phân tuyến tính như Hình 4. Cho m phần tử x_1, x_2, \dots, x_m trong không gian n chiều, có nhãn (lớp) của các phần tử là y_1, y_2, \dots, y_m có giá trị 1 hoặc -1 . $y_i = 1$, nếu x_i thuộc lớp $+1$ (lớp dương, lớp chúng ta quan tâm), $y_i = -1$, nếu x_i thuộc lớp -1 (lớp âm hay các lớp còn lại). SVM tìm siêu phẳng tối ưu (xác định bởi véc tơ pháp tuyến w và độ lệch của siêu phẳng b) dựa trên 2 siêu phẳng hỗ trợ của 2 lớp. Các phần tử lớp $+1$ nằm bên phải của siêu phẳng hỗ trợ cho lớp $+1$, các phần tử lớp -1 nằm phía bên trái của siêu phẳng hỗ trợ cho lớp -1 . Những phần tử nằm ngược phía với siêu phẳng hỗ trợ được coi như lỗi. Khoảng cách lỗi được biểu diễn bởi $z_i \geq 0$ (với x_i nằm đúng phía của siêu phẳng hỗ trợ của nó thì khoảng cách lỗi tương ứng $z_i = 0$, còn ngược lại thì $z_i > 0$ là khoảng cách từ điểm x_i đến siêu phẳng hỗ trợ tương ứng của nó). Khoảng cách giữa 2 siêu phẳng hỗ trợ được gọi là lề. Siêu phẳng tối ưu (nằm giữa 2 siêu phẳng hỗ trợ) tìm được từ 2 tiêu chí là cực đại hóa lề (lề càng lớn, mô hình phân lớp càng an toàn) và cực tiểu hóa lỗi. Vấn đề dẫn đến việc giải bài toán quy hoạch toàn phương (1):

$$\min \Psi(w, b, z) = (1/2) \|w\|^2 + c \sum_{i=1}^m z_i \quad (1)$$

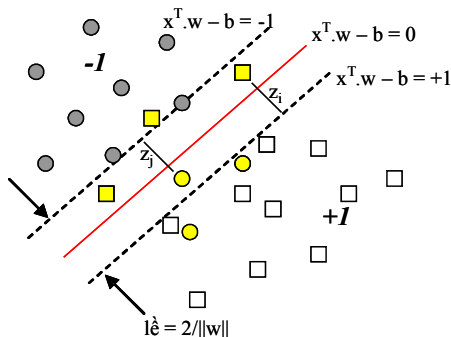
s.t.

$$y_i(w \cdot x_i - b) + z_i \geq 1$$

$$z_i \geq 0 \quad (i=1, 2, \dots, m)$$

hằng $c > 0$ sử dụng để chỉnh độ rộng lề và lỗi

Giải bài toán quy hoạch toàn phương (1), thu được (w, b) . Phân lớp phần tử x dựa vào dấu của $(w \cdot x - b)$.



Hình 4: Phân lớp tuyến tính với máy học véc tơ hỗ trợ

Bài toán quy hoạch toàn phương (1) được nghiên cứu phổ biến trong toán tối ưu. Hiện có nhiều giải thuật tiêu biểu như LibSVM (Chang & Lin, 01), SMO (Platt, 98), Newton (Mangasarian, 01) đều có độ phức tạp bậc 2 với số phần tử dữ liệu.

3.2 Giải thuật giảm gradient ngẫu nhiên (SGD)

Một cài đặt cho giải thuật SVM của (Bottou & Bousquet, 08) dựa trên phương pháp giảm gradient ngẫu nhiên, có độ phức tạp tuyến tính với số phần tử dữ liệu. Bằng cách thay thế z_i bởi $w \cdot x_i$, y_i (không xét độ lệch b) từ các ràng buộc vào hàm mục tiêu của (1), việc tìm siêu phẳng tối ưu của SVM có thể được thực hiện bởi (2):

$$\min \Psi(w, x, y) = (\lambda/2) \|w\|^2 + (1/m) \sum_{i=1}^m \max\{0, 1 - y_i(w \cdot x_i)\} \quad (2)$$

Phương pháp giảm gradient (GD) thực hiện tối ưu vấn đề (2) bằng cách cập nhật w tại lần lặp thứ $(t+1)$ với tốc độ học η_t , như trong (3):

$$w_{t+1} = w_t - (\eta_t / m) \sum_{i=1}^m \nabla_w \Psi(w_t, x_i, y_i) \quad (3)$$

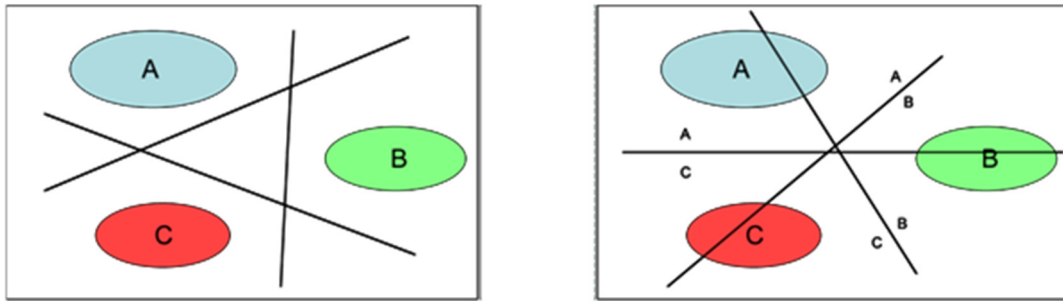
Phương pháp giảm gradient ngẫu nhiên (SGD) thực hiện đơn giản bước cập nhật w_{t+1} chỉ sử dụng một phần tử ngẫu nhiên (x_t, y_t) tại mỗi lần lặp:

$$w_{t+1} = w_t - \eta_t \nabla_w \Psi(w_t, x_t, y_t) \quad (4)$$

Có thể thấy rằng giải thuật SGD đơn giản, thực hiện các bước lặp, mỗi bước lặp chỉ lấy 1 phần tử ngẫu nhiên từ tập dữ liệu, thực hiện cập nhật w thay vì phải giải bài toán quy hoạch toàn phương (1). Giải thuật SGD có độ phức tạp tuyến tính với số phần tử của tập dữ liệu học, phân lớp dữ liệu có số phần tử và số chiều lớn rất hiệu quả (Bottou & Bousquet, 08).

3.3 Giải thuật giảm gradient ngẫu nhiên cho phân lớp đa lớp (MC-SGD)

SGD cũng như hầu hết các giải thuật SVM đều xuất phát từ vấn đề phân lớp nhị phân (2 lớp: dương và âm). Chúng tôi mở rộng giải thuật SGD để có thể giải quyết vấn đề phân lớp tập dữ liệu có c lớp ($c \geq 3$) hay còn gọi là đa lớp.



Hình 5: Phân lớp đa lớp, 1-vs-all (trái), 1-vs-1 (phải)

Để giải quyết phân lớp tập dữ liệu đa lớp, giải thuật SVM thường dựa trên 2 phương pháp đơn giản là 1-vs-all (Vapnik, 95) và 1-vs-1 (Krebel, 99).

Phương pháp 1-vs-all xây dựng c mô hình SVM nhị phân, mô hình thứ t tách lớp t (lớp dương) ra khỏi các lớp khác (âm).

Phương pháp 1-vs-1 xây dựng $c(c-1)/2$ mô hình SVM nhị phân, mỗi mô hình tách một cặp 2 lớp.

Việc phân lớp dựa vào bình chọn khoảng cách đến các siêu phẳng thu được từ SVM nhị phân.

Như đã chỉ ra trong thực tế của phân lớp kho dữ liệu ảnh rất lớn, hàng chục ngàn lớp (Sanchez & Perronnin, 11), (Deng et al., 12), 1-vs-all thì đơn giản cho kết quả tốt cho vấn đề phân lớp ảnh.

Tuy nhiên, khi áp dụng 1-vs-all vào trong SGD để giải quyết vấn đề đa lớp, chúng ta lại gặp một khó khăn lớn, huấn luyện các SGD nhị phân trên tập dữ liệu mất cân bằng. Giả sử tập dữ liệu chúng ta có 100 lớp, thì khi sử dụng 1-vs-all, mô hình SGD thứ t tách lớp t (lớp dương chỉ chiếm khoảng 1%) ra khỏi các lớp khác (lớp âm chiếm khoảng 99%). SGD gặp khó khăn do sự mất cân bằng. SGD chỉ thường cập nhật w trong (4) khi lỗi xuất hiện thường là phần tử thuộc lớp âm mà ít khi làm điều đó với lớp dương vì xác suất lấy mẫu một phần tử lớp dương khoảng 1%, trong khi lớp âm là 99%. Mặc dù điều khó khăn này xảy ra nhưng độ chính xác tổng thể vẫn là 99% trong khi không thể tách được lớp t ra khỏi các lớp khác.

Để giải quyết vấn đề này, chúng tôi đề xuất chiến lược xây dựng balanced bagging cho từng mô hình SGD nhị phân dựa trên lấy mẫu giảm và cập nhật trọng số bất đối xứng. Huấn luyện mô hình SGD nhị phân tách lớp t (dương, thiểu số)

khỏi các lớp khác (âm, đa số), cần xây dựng k mô hình cơ sở SGD như sau:

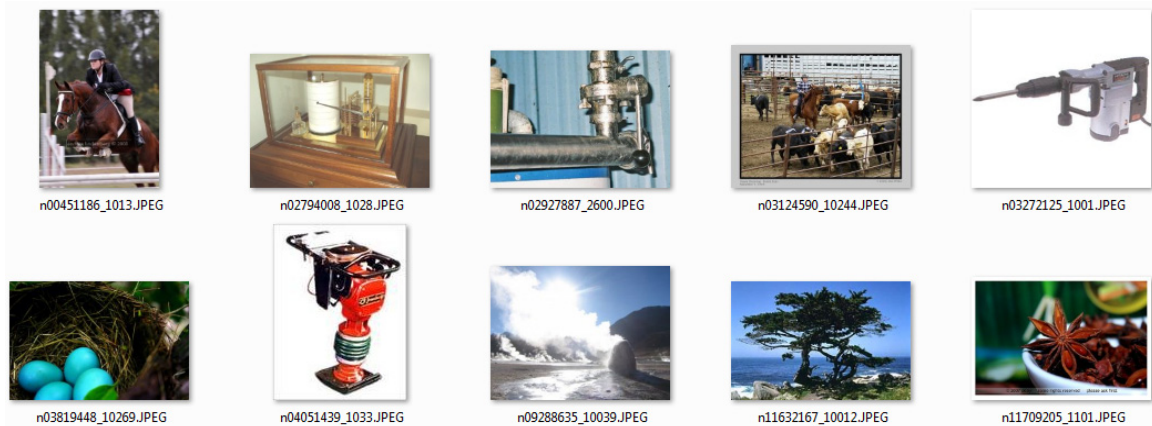
- Lấy mẫu giảm lớp âm sao cho số lượng phần tử lớp âm bằng với số phần tử của lớp dương.
- Sử dụng tập mẫu giảm lớp âm và dữ liệu của lớp dương làm tập huấn luyện mô hình cơ sở SGD.
- Huấn luyện SGD nhị phân, chú ý sử dụng công thức (4) với cập nhật trọng số lớn hơn khi phân lớp sai dữ liệu thuộc lớp dương (thiểu số), và trọng số nhỏ hơn khi phân lớp sai dữ liệu thuộc lớp âm (đa số).

Kết thúc, chúng ta tổng hợp k mô hình cơ sở SGD thu được mô hình SGD nhị phân tách lớp t từ các lớp khác.

Cần chú ý rằng, balanced bagging sử dụng tập mẫu giảm của lớp âm, giúp cân bằng phân bố dữ liệu giữa 2 lớp, khả năng lấy mẫu một phần tử của lớp dương và lớp âm là gần như nhau khi tiến hành cập nhật w trong (4). Hơn nữa, lấy mẫu giảm của lớp âm làm tăng khoảng cách tách lớp (dương, âm). Điều này tạo điều kiện cho SGD hội tụ với tốc độ nhanh hơn so với sử dụng tập đầy đủ. Chính vì lý do đó mà giải thuật MC-SGD sử dụng các balanced bagging có thể phân lớp hiệu quả tập dữ liệu đa lớp.

4 KẾT QUẢ THỰC NGHIỆM

Để tiến hành đánh giá hiệu quả của giải thuật MC-SGD cho phân lớp ảnh đa lớp, chúng tôi đã cài đặt giải thuật MC-SGD bằng ngôn ngữ lập trình C/C++. Ngoài ra, chúng tôi cũng cần so sánh MC-SGD với một giải thuật SVM chuẩn, được sử dụng phổ biến trong cộng đồng máy học là LibSVM (Chang & Lin, 01). Tất cả các giải thuật đều được thực hiện trên một máy tính cá nhân (Intel 3GHz, 2GB RAM) chạy hệ điều hành Linux.



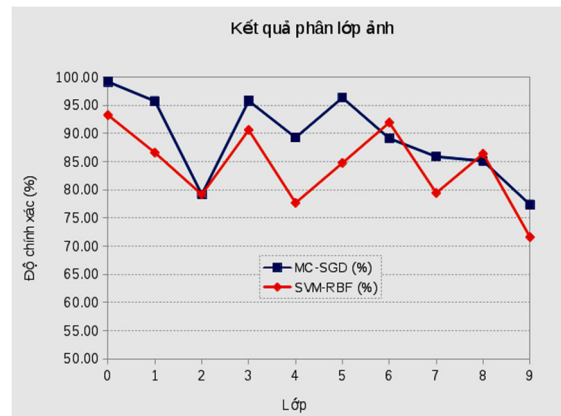
Hình 6: Ảnh mẫu trong tập ImageNet 10 lớp

Tập dữ liệu thực nghiệm được lấy về từ ImageNet (Deng *et al.*, 12). Chúng tôi chọn tập gồm 6675 ảnh của 10 lớp (xem Hình 6). Chúng tôi tách tập dữ liệu ra thành tập huấn luyện có 4450 ảnh và tập kiểm thử có chứa 2225 ảnh. Để biểu diễn ảnh bằng mô hình túi từ trực quan, chúng tôi sử dụng giải thuật phát hiện đặc trưng cục bộ Hessian Affine của (Mikolajczyk & Schmid, 04) để rút trích các véc tơ mô tả SIFT. Sau đó, thực thi giải thuật k -means (MacQueen, 67) để gom nhóm các véc tơ mô tả SIFT vào 50000 clusters tương ứng với 50000 từ trực quan. Giai đoạn tiền xử lý tạo ra hai tập (bảng) dữ liệu huấn luyện, kiểm thử, tương ứng với 4450 và 2225 phần tử, 50000 chiều và 10 lớp. Chúng tôi cố gắng thay đổi số clusters (từ trực quan từ 1000 đến 100000) để tìm các kết quả thực nghiệm tốt nhất. Cuối cùng, chúng tôi thu được độ chính xác ổn định với 50000 từ trực quan. Bước tiền xử lý này là duy nhất cho hai giải thuật máy học mà chúng tôi kiểm thử trong bài viết.

Chúng tôi sử dụng tập huấn luyện để xây dựng mô hình MC-SGD và SVM sử dụng hàm nhân phi tuyến RBF (SVM-RBF). Điều chỉnh tham số dựa vào nghi thức kiểm tra chéo (hold-out) được áp dụng trên tập huấn luyện. MC-SGD sử dụng tham số $\lambda = 0.1$ (hằng số quy tắc, dùng để điều chỉnh độ rộng lề phân hoạch), lặp 7 chu kỳ (epoch) là hội tụ đến kết quả tốt nhất. Chúng tôi cố gắng sử dụng hàm nhân RBF của SVM (hàm nhân RBF của hai điểm dữ liệu x_i, x_j là $K[i, j] = \exp(-\gamma \|x_i - x_j\|^2)$). Giải thuật SVM sử dụng hàm nhân RBF (với $\gamma = 0.00001$) và hằng số quy tắc $c = 1000$ (điều chỉnh độ rộng lề phân hoạch và lỗi) cho kết quả tốt nhất. Kết quả thu được trên tập kiểm thử như trình bày trong Bảng 1, Hình 7.

Bảng 1: Kết quả phân lớp ảnh

Lớp	MC-SGD (%)	SVM-RBF (%)
0	99.16	93.25
1	95.73	86.59
2	79.17	79.17
3	95.77	90.61
4	89.27	77.68
5	96.34	84.76
6	89.11	91.94
7	85.89	79.44
8	85.11	86.38
9	77.37	71.60
Overall	88.72	83.96



Hình 7: Kết quả phân lớp ảnh

So sánh kết quả cho thấy được MC-SGD phân lớp chính xác hơn SVM-RBF. MC-SGD cho kết quả tốt nhất 8 trong 10 lớp dữ liệu ảnh và cho độ chính xác tổng thể cao hơn gần 5% so với SVM-RBF. Hơn nữa, MC-SGD chỉ mất thời gian huấn luyện là 2.20 giây trong khi SVM-RBF cần đến 111.67 giây, hay nói cách khác MC-SGD nhanh hơn SVM-RBF 50 lần.

Với các kết quả phân lớp này, chúng tôi tin rằng giải thuật MC-SGD cho phép phân lớp hiệu quả dữ liệu có số chiều lớn thu được từ bước biểu diễn ảnh bằng mô hình túi từ và đặc trưng không đối xứng SIFT.

5 KẾT LUẬN VÀ ĐỀ XUẤT

Ý tưởng của việc tăng chiều trong biểu diễn ảnh bằng mô hình túi từ để có thể phân lớp đơn giản bằng mô hình tuyến tính mà không cần đến mô hình phi tuyến là tiếp cận rất hiệu quả. Khi số chiều dữ liệu là nhỏ, chúng ta cần đến mô hình phi tuyến để giải quyết tốt vấn đề phân lớp. Tuy nhiên, huấn luyện mô hình phi tuyến có độ phức tạp rất cao so với mô hình tuyến tính. Nhưng ngược lại, mô hình tuyến tính thì chỉ làm việc tốt trên tập dữ liệu có số chiều lớn. Chính vì lý do đó, tăng số chiều biểu diễn ảnh bằng mô hình túi từ để có thể sử dụng mô hình phân lớp tuyến tính là ý tưởng tuyệt vời cho cả độ chính xác và thời gian huấn luyện mô hình. Chúng tôi đã đề xuất giải thuật phân lớp tuyến tính đa lớp, giảm gradient ngẫu nhiên MC-SGD, cho phép phân lớp hiệu quả dữ liệu có số chiều lớn thu được từ bước biểu diễn ảnh. Kết quả thực nghiệm trên tập dữ liệu thực cho thấy giải thuật MC-SGD phân lớp nhanh, chính xác hơn khi so sánh với giải thuật máy học véc tơ hỗ trợ (Support Vector Machines - SVM) sử dụng hàm nhân phi tuyến (Radial Basis Function - RBF).

Chúng tôi cũng vừa phát triển giải thuật MC-SGD song song cho phép tăng tốc quá trình thực thi trên máy tính có nhiều bộ xử lý, nhóm hay lưới máy tính. Trong tương lai gần, chúng tôi sử dụng giải thuật MC-SGD song song để thực hiện phân lớp tập dữ liệu thử thách ImageNet-2012 có hơn 1 triệu ảnh, 1000 lớp khác nhau. Bên cạnh đó, chúng tôi cũng muốn chúng tôi dự định ứng dụng phương pháp đề xuất vào vấn đề phân lớp ảnh, văn bản, video.

TÀI LIỆU THAM KHẢO

1. Bosch, A., Zisserman, A., Munoz, X.: Scene classification via pLSA. In: Proceedings of the European Conference on Computer Vision, pp. 517–530 (2006).
2. Bottou, L., Bousquet, O.: The tradeoffs of large scale learning. In Advances in Neural Information Processing Systems (20):161–168 (2008).
3. Chang, C.C., Lin, C.J.: LIBSVM – a library for support vector machines (2001). <http://www.csie.ntu.edu.tw/~cjlin/libsvm>

4. Deng, J., Berg, A., Satheesh, S., Su, H., Khosla, A., Li, F.-F.: Large Scale Visual Recognition Challenge. (2012).
5. Guermeur, Y.: Svm multiclass, théorie et applications. Thèse HDR (2007)
6. Krebel, U.: Pairwise classification and support vector machines. *Advances in Kernel Methods: Support Vector Learning* pp. 255–268 (1999).
7. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. In Proceedings of the IEEE, 1998, pp. 2278–2324.
8. Lowe, D.: Distinctive image features from scale invariant keypoints. *International Journal of Computer Vision* pp. 91–110 (2004).
9. MacQueen, J.: Some methods for classification and analysis of multivariate observations. Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, *University of California Press* Vol.1, pp. 281–297 (1967).
10. Mangasarian O.L.: Mathematical Programming for Support Vector Machines. INRIA Rocquencourt, France July 17 (2001).
11. Mikolajczyk, K., Schmid, C.: Scale and affine invariant interest point detectors. *International Journal of Computer Vision* 60(1), 63–86 (2004).
12. Platt J.: Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines. Microsoft Research Technical Report MSR-TR-98-14 (1998)
13. Quinlan, J.R.: *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Mateo, CA (1993).
14. Sanchez, J., Perronnin, F.: High-dimensional signature compression for large scale image classification. In IEEE Conference on Computer Vision and Pattern Recognition pp. 1665–1672 (2011).
15. Vapnik, V.: *The Nature of Statistical Learning Theory*. Springer-Verlag (1995).
16. Viola, P.A., Jones, M.J.: Rapid object detection using a boosted cascade of simple features. In IEEE Conference on Computer Vision and Pattern Recognition pp. 511–518 (2001).
17. Zheng, H., Daoudi, M.: Blocking adult images based on statistical skin detection. *Electronic Letters on Computer Vision and Image Analysis* 4(2):1–14 (2004).