

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

УДК 004.932

АЛГОРИТМЫ КЛАССИФИКАЦИИ ИЗОБРАЖЕНИЙ С БОЛЬШИМ ЧИСЛОМ КАТЕГОРИЙ ОБЪЕКТОВ

© 2013 г.

А.Н. Половинкин

Нижегородский госуниверситет им. Н.И. Лобачевского

alexey.polovinkin@gmail.com

Поступила в редакцию 14.05.2013

Рассматривается задача классификации изображений с большим числом категорий. Предлагаются основанные на деревьях решений и их ансамблях модификации алгоритмов классификации, учитывающие семантическую иерархию объектов. Приводятся результаты вычислительных экспериментов, демонстрирующие лучшую точность предсказания на ряде наборов данных.

Ключевые слова: компьютерное зрение, машинное обучение, классификация изображений, дерево решений, случайный лес, семантическая иерархия.

Введение

Важное прикладное значение имеют задачи классификации изображений с большим числом категорий объектов. Проблемы подобного рода возникают в различных областях: например, поиск изображений по ключевым словам, корректное распознавание объектов для робототехники и т.п. До последнего времени алгоритмы для решения данной задачи мало исследовались, в том числе и ввиду вычислительной сложности. Одни из первых проведенных вычислительных экспериментов подобного рода на базах изображений Caltech-101 [1] и ImageNet [2] показали, что применение общепризнанных на настоящее время алгоритмов не дает удовлетворительной точности распознавания. Лучшие результаты для задач данного типа продемонстрировали два класса алгоритмов. Первый из них основан на использовании векторов Фишера [3] для описания изображений и линейной машины опорных векторов [4] в качестве классификатора. Второй класс алгоритмов использует для описания изображений непосредственно интенсивности пикселей, в качестве классификатора используются конволюционные нейронные сети [5]. Альтернативным подходом к улучшению точности распознавания является использование информации о структуре связей между категориями объектов. В работах [6, 7] выдвинута и экспериментально апробирована идея о зависимости качества алгоритмов классификации от того, насколько эти алгоритмы принимают во внимание семантиче-

ские зависимости между рассматриваемыми категориями.

В данной работе предлагаются основанные на деревьях решений и их ансамблях модификации алгоритмов классификации изображений, учитывающие семантическую иерархию объектов в базе данных WordNet [8].

Работа построена следующим образом. В разделе 1 приводится описание используемых алгоритмов обучения с учителем «случайный лес» и «машина опорных векторов». В разделе 2 описываются разработанные методы классификации изображений с большим числом категорий с учетом семантической иерархии категорий объектов. Результаты вычислительного эксперимента и их анализ приведены в разделе 2.4.2.

1. Теоретические основы

1.1. Алгоритм обучения с учителем «случайный лес»

Случайный лес (random forest) [9] является одним из наиболее универсальных и эффективных алгоритмов обучения с учителем, применимым как для задач классификации, так и для задач восстановления регрессии. Идея метода заключается в построении ансамбля параллельно обучаемых независимых деревьев решений [10]. Итоговая классификация объекта выполняется путем голосования всех деревьев, входящих в состав ансамбля.

Среди достоинств алгоритма можно выделить высокое качество предсказания, способность эффективно обрабатывать данные с большим числом категорий и признаков, внут-

ренную оценку обобщающей способности модели, высокую распараллеливаемость и масштабируемость. Также метод обладает всеми преимуществами деревьев решений, в том числе отсутствием необходимости предобработки входных данных и обработкой как вещественных, так и категориальных признаков, поддерживающей работу с отсутствующими значениями.

Алгоритм CART построения дерева решений [10] строит разбиение пространства признаков на непересекающиеся области путем его рекурсивного разбиения. Фактически каждому узлу дерева соответствуют некоторая область пространства признаков и правило, по которому осуществляется ее разделение на две, приписанные к дочерним вершинам. Так как задача построения оптимального дерева, минимизирующего ошибку на обучающей выборке при заданном ограничении на его размеры, является NP-трудной [11], для обучения данной модели используется жадная стратегия максимального уменьшения функции, описывающей неоднородность данных $I(D)$, где D есть область в пространстве признаков, соответствующая некоторому узлу.

Пусть $y \in \{1, \dots, K\}$, где K – число категорий объектов, и область D содержит M объектов из обучающей выборки; обозначим относительную частоту категории k в данном узле дерева $\hat{p}_k = \frac{1}{M} \sum_{x_i \in D} I(y_i = k)$, где I – индикаторная функция. Категория \tilde{k} с наибольшей относительной частотой соответствует ответу, предсказываемому для всех объектов, относящихся к данному узлу. В качестве функции, описывающей неоднородность данных в узле $I(D)$, обычно рассматриваются следующие [9]:

- ошибка классификации:

$$\frac{1}{M} \sum_{x_i \in D} I(y_i \neq \tilde{k}) = 1 - \hat{p}_{\tilde{k}};$$

- индекс Джини:

$$\sum_{k \neq \tilde{k}} \hat{p}_k \hat{p}_{\tilde{k}} = \sum_{k=1}^K \hat{p}_k (1 - \hat{p}_k);$$

- кросс-энтропия:

$$-\sum_{k=1}^K \hat{p}_k \log \hat{p}_k.$$

Обучение модели случайного леса состоит из независимого построения определенного количества деревьев по следующей процедуре: генерация случайной подвыборки из обучающей выборки (размер данной подвыборки обычно составляет 50–70% от размера всей обучающей выборки) и построение дерева решений по данной подвыборке (причем в каждом новом узле дерева переменная для разбиения выбирается не из всех признаков, а из случайно выбранного их подмножества небольшой мощности).

1.2. Алгоритм обучения с учителем «машина опорных векторов»

Метод опорных векторов [4] относится к наиболее известным и широко используемым на практике алгоритмам обучения с учителем. Изначально данный метод был предназначен для задач бинарной классификации, но легко обобщается на задачи классификации с большим числом классов, а также на задачи восстановления регрессии. В ходе обучения алгоритма строится линейный пороговый классификатор $\text{sgn}(\sum_{j=1}^n w_j x_j - w_0)$, где уравнение $\langle w, x \rangle = w_0$ описывает гиперплоскость, разделяющую классы. Если предположить, что выборка линейно разделима, тогда очевидно, что разделяющая гиперплоскость не единственна, поскольку существуют и другие её положения, реализующие то же самое разбиение выборки. Идея метода заключается в том, чтобы выбрать w и w_0 оптимальным с точки зрения некоторого критерия способом. Можно потребовать, чтобы разделяющая гиперплоскость максимально далеко отстояла от ближайших к ней объектов обоих классов. Первоначально данный принцип классификации возник из эвристических соображений: вполне естественно полагать, что максимизация зазора (margin) между классами должна способствовать более уверенной классификации. В дальнейшем этот принцип получил мощное теоретическое обоснование.

Чтобы разделяющая гиперплоскость как можно дальше отстояла от точек выборки, ширина полосы должна быть максимальной. Нетрудно показать, что ширина полосы максимальна, когда норма вектора w минимальна. Итак, в случае, когда выборка линейно разделима, достаточно простые геометрические соображения приводят к следующей задаче: требуется найти такие значения параметров w и w_0 , при которых норма вектора w минимальна при некоторых ограничениях. Таким образом, исходная задача бинарной классификации свелась к задаче квадратичного программирования.

Алгоритм легко обобщается на случай линейно не разделимых данных (что обычно и происходит на практике) путем добавления к оптимизируемой функции штрафа за ошибки на объектах обучающей выборки. Существует и другой подход к решению данной проблемы: это переход от исходного пространства признаков описаний объектов X к новому пространству H (которое называется спрямляющим) с помощью некоторого преобразования $\psi: X \rightarrow H$. Если пространство H имеет достаточно высокую размерность, то можно надеять-

ся, что в нём выборка окажется линейно разделимой (легко показать, что если выборка не противоречива, то всегда найдётся пространство конечной размерности, в котором она будет линейно разделима).

Для того чтобы использовать метод опорных векторов для задачи классификации с числом классов $K > 2$, возможны две стратегии:

1. «Каждый против каждого»: обучить $K(K-1)/2$ различных моделей на всех возможных подзадачах бинарной классификации. Новый объект классифицируется всеми построенными моделями, затем выбирается преобладающий класс.

2. «Один против всех»: обучить K моделей на задачах бинарной классификации вида «один класс против всех остальных». Класс нового объекта выбирается по максимальному значению margin.

2. Алгоритмы классификации изображений, учитывающие семантическую иерархию категорий

В настоящем разделе приводятся описания двух предлагаемых алгоритмов построения случайного леса, учитывающих семантическую иерархию категорий объектов. Первый из них использует модифицированную матрицу штрафа. Второй сводит задачу к иерархической классификации.

2.1. Описание базы WordNet

WordNet [8] представляет собой большую базу данных существительных, глаголов, прилагательных и наречий английского языка, сгруппированных в наборы синонимов, каждый из которых соответствует определенной сущности. Между данными наборами определены связи, которые отражают их семантические и лексические отношения. Среди типов этих связей можно выделить отношение включения, которое связывает более общие понятия (например, «мебель») с более детальными (например, «шкаф»). Последний, в свою очередь, может быть связан с понятием «книжный шкаф» и т.д. (очевидно, что данное отношение обладает свойством транзитивности). Общую структуру можно описать в виде направленного ациклического графа, вершинам которого соответствуют понятия, а ребрам – отношения между данными понятиями. В построенном графе можно выделить корневую вершину, которой соответствует понятие «сущность», являющуюся общим предком для всех остальных вершин.

2.2. Построение случайного леса с использованием модифицированной матрицы штрафа

Одним из возможных использований семантической иерархии объектов при обучении ансамбля деревьев решений является её применение при построении каждого отдельного дерева в ансамбле путем модификации функции неоднородности (индекса Джини). Основная цель данной модификации заключается в том, чтобы увеличить «штраф» за нахождение в узле объектов, помеченных категориями, находящимися далеко друг от друга в семантической иерархии. Вычисляемый таким образом индекс Джини можно записать следующим образом:

$\sum_{k \neq \bar{k}} w_{k\bar{k}} \hat{p}_k \hat{p}_{\bar{k}}$, где веса w_{ij} вычисляются с помощью следующего алгоритма.

Вход: H – граф, описывающий структуру понятий в базе WordNet, h – высота дерева; $\{C_1, \dots, C_K\}$ – категории объектов, для которых необходимо вычислить веса.

Для всех $i = 1, \dots, K$

a) $w_{ii} = 0$;

b) для всех $j = i + 1, \dots, K$

i. Вычислить $h(i, j)$ – максимальную глубину узла, являющегося общим предком для категорий C_i и C_j в графе H .

ii. $w_{ij} = h - h(i, j)$, $w_{ji} = w_{ij}$.

Для всех $i = 1, \dots, K$

нормировать w_{ij} таким образом, чтобы

$$\sum_{j=1}^K w_{ij} = 1.$$

После построения модифицированной функции штрафа применяется обычный алгоритм построения случайного леса [9].

2.3. Иерархическая классификация

Один из наиболее известных подходов к решению задачи классификации объектов с большим числом категорий заключается в использовании иерархической классификации [12]. Основная идея данного подхода заключается в построении иерархии категорий объектов, организованной в виде дерева таким образом, что категории, относящиеся к каждому узлу дерева, делятся на некоторое число кластеров, каждый из которых соответствует одному узлу-потомку. Процесс продолжается до тех пор, пока в каждом терминальном узле дерева не будет содержаться ровно по одной категории. В каждом узле дерева строится классификатор (обычно бинарный), который обучает модель классификации кластеров категорий, относящихся к узлам-потомкам. Для классификации нового объекта применяется цепочка из классификаторов,

ведущая из корня дерева к терминальному узлу, соответствующему тому или иному классу.

Для построения дерева, описывающего иерархию объектов в базе WordNet, разработан следующий алгоритм:

Вход: H – граф, описывающий структуру понятий в базе WordNet; $\{C_{i_1}, \dots, C_{i_m}\}$ – категории объектов, для которых необходимо построить разбиение.

1. Если $m = 1$, то СТОП.

2. Вычислить $h(i_1, \dots, i_m)$ – минимальную глубину узла, являющегося общим предком для всех категорий C_{i_1}, \dots, C_{i_m} в графе H .

3. Для всех $i = i_1, \dots, i_m$

Вычислить $C(i)$ – предка узла i глубины $h(i_1, \dots, i_m) + 1$.

4. Выполнить разбиение множества $\{C_{i_1}, \dots, C_{i_m}\}$ на подмножества $CG_1 \cup \dots \cup CG_K$, где $CG_k = \{C_{j_1}, \dots, C_{j_k}\}$ $C(j_1) = C(j_2) = \dots = C(j_k)$.

5. Выполнить шаги (1)–(4) для всех CG_1, \dots, CG_K .

Таким образом, выходом данного алгоритма является дерево T с числом терминальных узлов, равным числу возможных категорий, описывающее иерархию категорий. В каждом узле данного дерева строится отдельный классификатор (случайный лес), объединение данных классификаторов составляет иерархический классификатор. Алгоритм построения может быть описан следующим образом:

Вход: $D = \{(x^i, y^i), i = 1, \dots, n\}$ – обучающая выборка; T – дерево, описывающее иерархию категорий объектов

1. Для каждого нетерминального узла $t \in T$

a) получить список категорий $\{C_{i_1}, \dots, C_{i_m}\}$, соответствующих данному узлу;

b) выбрать подмножество из обучающей выборки $\tilde{D} \subseteq D$, такое, что метки категорий для всех объектов из \tilde{D} принадлежат множеству $\{C_{i_1}, \dots, C_{i_m}\}$;

c) получить список потомков узла $t: \{t_1, \dots, t_K\}$ и соответствующих им наборов категорий CG_1, \dots, CG_K ;

d) обучить классификатор на наборе данных \tilde{D} , для которого векторы признаков \bar{x}^i соответствуют векторам признаков из \tilde{D} , а категории $\bar{y}^i \in \{1, \dots, K\}$ (в зависимости от того, какому набору категорий из CG_1, \dots, CG_K принадлежит y^i).

2.4. Вычислительный эксперимент

2.4.1. Описание используемых данных и алгоритмов

Для проведения вычислительного эксперимента использовались наборы данных Caltech-101 [1], Caltech-256 [1] и ILSVRC2010 [13]. База данных Caltech-101 содержит 9144 изображения (размером 300 на 200 пикселей), относящиеся к 101 различной категории, число изображений на одну категорию изменяется от 31 до 800 (в среднем на одну категорию приходится около 50 изображений). Caltech-256 является расширением базы Caltech-101 (число категорий равно 256; общее число изображений – 30608); минимальное число изображений, относящихся к одной категории, увеличилось до 80, максимальное – до 827; среднее число изображений на одну категорию – до 100). База данных ILSVRC2010 представляет собой подмножество базы данных ImageNet, содержащее 1000 категорий (общее число изображений равно 1200000, максимальная размерность изображения равна 500 пикселям).

Для построения признакового описания изображений использовался bag-of-words подход [14]. Сначала каждое изображение масштабировалось таким образом, чтобы максимальная размерность не превосходила 300 пикселей. Из масштабированного таким образом изображения извлекались SIFT признаки (для ускорения вычислений использовалась dense SIFT модификация [15], дескрипторы извлекались из перекрывающихся патчей размером 20×20 пикселей, взятых с шагом в 10 пикселей). Извлеченных дескрипторов случайным образом формировалась выборка (объем выборки: для набора данных Caltech-101 – 10^5 объектов, для набора данных Caltech-256 – 10^6 объектов, для набора данных ILSVRC2010 – 10^7 объектов). Полученная выборка подавалась на вход алгоритма кластеризации «K means», центроиды построенных кластеров соответствовали «визуальным словам» (число кластеров для наборов данных Caltech-101 и Caltech-256 принималось равным 1000, для ILSVRC2010 – 2000). Итоговое признаковое описание изображения представляет собой гистограмму, описывающую частоту вхождения «визуальных слов» в набор дескрипторов ключевых точек для данного изображения.

Проводилось сравнение алгоритмов, использующих предлагаемую здесь методику: случайный лес с модифицированной матрицей штрафа (RF_COST_MATRIX), иерархический алгоритм классификации изображений на основе случайного леса (RF_HIERARCHICAL), иерархический алгоритм классификации изображений на основе машины опорных векторов (SVM_HEIRARCHICAL), – с «классическими» алгоритмами «случайный лес» (RF) и «машина опорных

Таблица 1

	Caltech-101	Caltech-256	ILSVRC2010
RF	0.181	0.325	0.497
RF_COST_MATRIX	0.172	0.317	0.523
RF_HIERARCHICAL	0.154	0.265	0.445
SVM_HIERARCHICAL	0.171	0.281	0.482
SVM_ONE_VS_ALL	0.191	0.331	0.423

Таблица 2

	ILSVRC_100_RANDOM	ILSVRC_100_HIERARCHICAL
RF	0.312	0.271
RF_COST_MATRIX	0.287	0.225
RF_HIERARCHICAL	0.275	0.151
SVM_HIERARCHICAL	0.271	0.163
SVM_ONE_VS_ALL	0.151	0.149

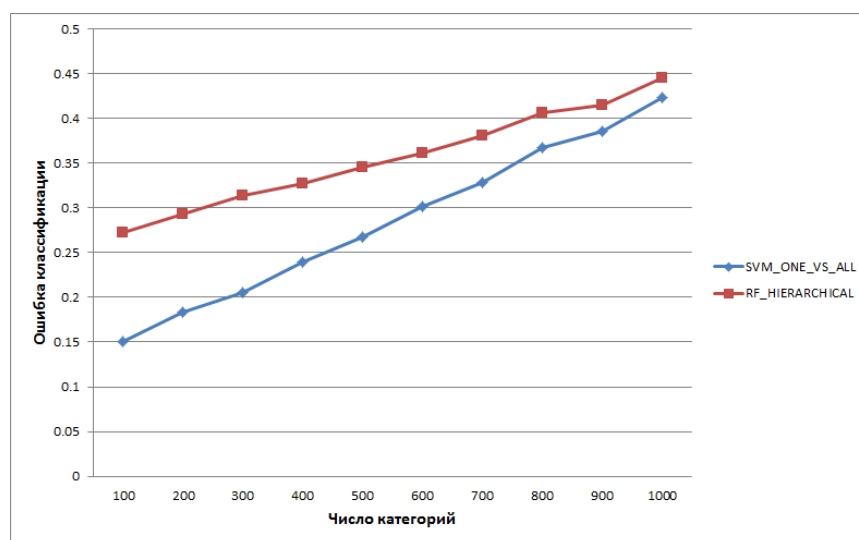


Рис. 1. Зависимость тестовой ошибки от числа категорий объектов в наборе данных

векторов» со схемой принятия решений «один против всех» (SVM_ONE_VS_ALL).

Вследствие высокой вычислительной сложности для алгоритмов, использующих машину опорных векторов, применялись только линейные ядра (использовалась оптимизированная библиотека LIBLINEAR [16]).

2.4.2. Результаты вычислительного эксперимента

В таблице 1 приведены результаты работы алгоритмов (ошибка отнесения к чужой категории) для описанных выше алгоритмов.

Видно, что иерархический алгоритм классификации изображений на основе случайного леса дает лучшую точность предсказания на наборах данных Caltech-101 и Caltech-256. Машина опорных векторов демонстрирует небольшое преимущество на наборе данных ILSVRC-2010.

Рассмотрим, как меняется ошибка классификации в зависимости от числа категорий в ILSVRC-2010, для этого случайным образом выберем подмножество категорий в данном наборе данных (изменяя мощность данного подмножества от 100 до 1000 с шагом, равным 100) и построим график (см. рис. 1) зависимости ошибки классификации на тестовой выборке от числа категорий.

Из графика следует, что с ростом числа категорий разность в точности предсказания между RF_HIERARCHICAL и SVM_ONE_VS_ALL уменьшается, при дальнейшем увеличении числа категорий (например, если взять полную версию набора данных ImageNet) следует ожидать одинаковой точности предсказания данных алгоритмов.

В работе также исследовалась зависимость ошибки предсказания от структуры рассматриваемых данных. Для проведения эксперимента

было выбрано подмножество ILSVRC2010, содержащее 100 категорий двумя различными способами: 1) случайным образом (ILSVRC_100_RANDOM); 2) выбирались категории, относящиеся к четырем высокоуровневым понятиям базы WordNet (ILSVRC_100_HIERARCHICAL). Как видно из приведенных в таблице 2 результатов эксперимента, использование иерархических подходов дает существенное улучшение на категориях, организованных в соответствии с семантической структурой WordNet.

Результаты вычислительного эксперимента свидетельствуют, что предложенные алгоритмы демонстрируют лучшую точность предсказания на некоторых наборах данных. В дальнейшем планируется разработка модифицированной реализации алгоритмов, основанной на деревьях решений с распределенными данными, что позволит работать с большей размерностью пространства признаков и, возможно, улучшить качество предсказания.

Автор благодарит В.Л. Ерухимова, Н.Ю. Золотых и И.Б. Меерова за полезные обсуждения.

Работа выполнена в рамках программы «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007–2013 годы», государственный контракт № 11.519.11.4015.

Список литературы

1. CaltechDatasets. URL: <http://www.vision.caltech.edu/html-files/archive.html> (дата обращения 01.03.2013).
2. Deng J., Dong W., Socher R., et al. ImageNet: A large-scale hierarchical image database // Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2009. P. 248–255.
3. Perronnin F., Liu Y., Sanchez J., Poirier H. Large-Scale Image Retrieval with Compressed Fisher Vectors // Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2010. P. 3384–3391.
4. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. Springer, 2008.
5. Krizhevsky A., Sutskever I., Hinton G. ImageNet Classification with Deep Convolutional Neural Networks // Advances in Neural Information Processing Systems (NIPS). 2012. P. 1106–1114.
6. Deng J., Berg A., Fei-Fei L. Hierarchical Semantic Indexing for Large Scale Image Retrieval // Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2011.
7. Deng J., Berg A., Li K., Fei-Fei L. What does classifying more than 10,000 image categories tell us? // Proc. of the 12th European Conference of Computer Vision. 2010. P. 71–84.
8. Fellbaum C. WordNet: An Electronic Lexical Database. MIT Press, 1998.
9. Breiman L. Random Forests // Machine Learning. 2001. V. 45. № 1. P. 5–32.
10. Breiman L., Friedman J.H., Olshen R.A., Stone C.J. Classification and Regression Trees. Wadsworth & Brooks, 1984.
11. Гэри М., Джонсон Д. Вычислительные машины и труднорешаемые задачи. М.: Мир, 1982.
12. Griffin G., Perona P. Learning and using taxonomies for fast visual categorization // Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2008. P. 1–8.
13. Large Scale Visual Recognition Challenge 2010 (ILSVRC2010). URL: <http://www.image-net.org/challenges/LSVRC/2010/index> (дата обращения 01.03.2013).
14. Csurka G., Bray C., Dance C., Fan L. Visual categorization with bags of keypoints // Workshop on Statistical Learning in Computer Vision, ECCV. 2004. P. 1–22.
15. Bosch A., Zisserman A., Munoz X. Image classification using random forests and ferns // Proc. of the IEEE 11th International Conference on Computer Vision (ICCV). 2007. P. 1–8.

ALGORITHMS FOR IMAGE CLASSIFICATION WITH A LARGE NUMBER OF OBJECT CATEGORIES

A.N. Polovinkin

An image classification problem with a large number of object categories is considered. Classification algorithm modifications are proposed which are based on decision trees and their ensembles with the account of semantic hierarchies of objects. The results of numerical experiments presented show better prediction accuracy on a number of data sets.

Keywords: computer vision, machine learning, image classification, decision tree, random forest, semantic hierarchy.