

МЕТОДЫ РЕШЕНИЯ ЗАДАЧИ КЛАССИФИКАЦИИ ИЗОБРАЖЕНИЙ С БОЛЬШИМ ЧИСЛОМ КАТЕГОРИЙ

А.Н. Половинкин

Нижегородский госуниверситет им. Н.И. Лобачевского

Рассматривается задача классификации изображений с большим числом категорий объектов. Предлагаются основанные на деревьях решений и их ансамблях модификации алгоритмов классификации, учитывающие семантическую иерархию объектов в базе данных WordNet. Приводятся результаты вычислительного эксперимента.

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. Впервые проведенные вычислительные эксперименты подобного рода на базах изображений Caltech [3] и ImageNet [6] показали, что применение общепризнанных на настоящее время алгоритмов не дает удовлетворительной точности распознавания. В работах [4, 5] выдвинута и экспериментально апробирована идея о влиянии на качество алгоритмов классификации того, насколько эти алгоритмы принимают во внимание семантические зависимости между рассматриваемыми категориями. В данной работе рассматриваются основанные на деревьях решений и их ансамблях модификации алгоритмов классификации, учитывающие семантическую иерархию объектов в базе данных WordNet.

1. Алгоритм обучения с учителем «случайный лес»

Случайный лес [1] 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 [2]. 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 классов и признаков, внутреннюю оценку обобщающей способности модели, высокую распараллеливаемость и масштабируемость. Также метод обладает всеми преимуществами деревьев решений, в том числе отсутствием необходимости предобработки входных данных, обработкой как вещественных, так и категориальных признаков, поддержкой работы с отсутствующими значениями.

Напомним, что дерево решений [2] строит разбиение пространства признаков на непересекающиеся области путем его рекурсивного разбиения. Фактически каждому узлу дерева соответствует некоторая область пространства признаков и правило, по которому осуществляется ее разделение на две, приписанных к дочерним вершинам. Так как задача построения оптимального дерева, минимизирующего ошибку на обучающей выборке при заданном ограничении на его размеры, является NP-трудной, для обуче-

ния 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 класса k в данном узле дерева $\hat{p}_k = \frac{1}{M} \sum_{x_i \in D} I(y_i = k)$. Класс \tilde{k} с наибольшей относительной частотой соответствует ответу, предсказываемому для всех объектов, относящихся к данному узлу. В качестве функции, описывающей неоднородность данных в узле $I(D)$, возможны следующие варианты [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).

2. Описание базы WordNet

WordNet [7] 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 связывают более общие понятия (например, «мебель») с более детальными (например, «шкаф»), который, 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.

3. Алгоритм построения 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 (индекс 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: $\sum_{k \neq \tilde{k}} w_{k\tilde{k}} \hat{p}_k \hat{p}_{\tilde{k}}$, где веса w_{ij} вычисляются следующим образом:

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

1. 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 классов C_i и C_j в графе H .
 - ii. $w_{ij} = h - h(i, j)$, $w_{ji} = w_{ij}$.
2. Для всех $i = 1, \dots, K$
- а. Нормировать w_{ij} таким образом, чтобы $\sum_{j=1}^K w_{ij} = 1$.

4. Иерархический алгоритм классификации изображений с использованием случайного леса

Один из наиболее известных подходов к решению задачи классификации объектов с большим числом классов заключается в использовании иерархической классификации [8]. Основная идея данного подхода заключается в построении иерархии классов объектов, организованной в виде дерева таким образом, что классы, 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 одному классу. В каждом узле дерева строится классификатор (обычно бинарный), который обучает модель классификации кластеров классов, относящимся 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; $\{C_{i_1}, \dots, C_{i_m}\}$ – классы 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 классов C_{i_1}, \dots, C_{i_m} в графе H .
3. Для всех $i = i_1, \dots, i_m$
 - а. Вычислить $C(i)$ – предка узла i глубины $h(i_1, \dots, i_m) + 1$.
5. Выполнить разбиение множества $\{C_{i_1}, \dots, C_{i_m}\}$ на подмножества $CG_1 \cup \dots \cup CG_K$, где $CG_k = \{C_{j_1}, \dots, C_{j_k}\}$, $C(j_1) = C(j_2) = \dots = C(j_k)$.
6. Выполнить шаги (1)-(5) 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 $t \in T$
 - а. получить список категорий $\{C_{i_1}, \dots, C_{i_m}\}$, соответствующих данному узлу;
 - б. выбрать подмножество из обучающей выборки $\tilde{D} \subseteq D$ такое, что метки классов для всех объектов из \tilde{D} принадлежат множеству $\{C_{i_1}, \dots, C_{i_m}\}$;
 - с. получить список потомков узла $t: \{t_1, \dots, t_K\}$ и соответствующих им наборов категорий $\{CG_1, \dots, CG_K\}$;
 - д. обучить классификатор на наборе данных \tilde{D} , для которого вектора 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).

5. Вычислительный эксперимент

В рамках работы была проведена первоначальная апробация предложенных алгоритмов. Алгоритм построения случайного леса с модифицированной матрицей штрафа (RF_COST_MATRIX) и иерархический алгоритм классификации изображений с использованием случайного леса (RF_HIERARCHICAL) сравнивались с алгоритмом, объединяющим в себе построение словаря кодовых слов с использованием алгоритма KMeans и линейного SVM в качестве классификатора (BOW+SVM). Сравнение проводилось на базе данных ILSVRC 2010 [10], являющейся подмножеством базы данных ImageNet, содержащим 1000 категорий. Данные делились в соотношении 50:50 на обучающую и тестовую выборки. Каждое изображение масштабировалось таким образом, чтобы максимальная размерность не превосходила 300 пикселей, затем извлекались dense SIFT признаки (перекрывающиеся патчи размером 20x20 пикселей с шагом в 10 пикселей). Из извлеченных таким образом признаков выбиралось случайным образом 10^7 , на которых выполнялось построение словаря (1000 слов, соответствующих центроидам кластеров, найденным алгоритмом KMeans). При обучении машины опорных векторов использовался метод «один-против-всех», таким образом, общее число обученных классификаторов равнялось 1000 (число категорий). Обучение алгоритма случайного леса выполнялось с параметрами, выбираемыми по умолчанию. Результаты приведены в таблице 1.

Таблица 1. Ошибка классификации на тестовой выборке на базе данных ILSVRC2010

Алгоритм	Значение ошибки на тестовой выборке
RF_COST_MATRIX	0.53
RF_HIERARCHICAL	0.44
BOW+SVM	0.42

Как видно из табл. 1, иерархический алгоритм классификации на основе случайного леса дает примерно одинаковые результаты по сравнению с алгоритмом, объединяющим в себе построение словаря кодовых слов с использованием алгоритма KMeans и линейного SVM в качестве классификатора. Среди возможных подходов для повышения точности предсказания с использованием иерархического алгоритма следует выделить отдельное обучение словаря в каждом узле дерева, которое описывает иерархию объектов.

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.

Литература

1. Breiman L. Random Forests // Machine Learning. 2001. V. 45, №. 1, P. 5–32.
2. Breiman L., Friedman J.H., Olshen R.A., Stone C.J. Classification and Regression Trees. Wadsworth & Brooks, 1984.
3. CaltechDatasets – [<http://www.vision.caltech.edu/html-files/archive.html>].
4. Deng J., Berg A., Fei-Fei L. Hierarchical Semantic Indexing for Large Scale Image Retrieval // IEEE Computer Vision and Pattern Recognition (CVPR) – 2011.
5. Deng J., Berg A., Li K., Fei-Fei L. What does classifying more than 10,000 image categories tell us? // Proceedings of the 12th European Conference of Computer Vision (ECCV) – 2010.

6. Deng J., Dong W., Socher R., Li L., Li K., Fei-Fei L. ImageNet: A large-scale hierarchical image database // In CVPR09. – 2009.
7. Fellbaum C. WordNet: An Electronic Lexical Database. – MIT Press, 1998.
8. Griffin G., Perona P. Learning and using taxonomies for fast visual categorization // In CVPR, 2008.
9. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. – Springer, 2008.
10. Large Scale Visual Recognition Challenge 2010 (ILSVRC2010) [<http://www.image-net.org/challenges/LSVRC/2010/index>].