

Об одном значимом результате в распознавании образов рукописных цифр

Кузьмицкий Н.Н.

Кафедра "ЭВМ и системы"

Брестский государственный технический университет

Брест, Республика Беларусь

e-mail: knnbrest@yandex.ru

Аннотация— Исследованы сверточные нейронные сети (СНС) архитектуры LeNet-5 в распознавании образов рукописных цифр. Показана перспективность комитета СНС как средства интеграции их знаний и повышения точности. Подтверждено, что разнообразие и селекция членов – важные факторы эффективности комитета. Создан классификатор, имеющий уникальную точность распознавания тестового MNIST – 99.64%.

Ключевые слова: сверточная нейронная сеть, MNIST; комитет, схема голосования, разнообразие, селекция.

I. ВВЕДЕНИЕ

Многие исследователи в области машинного обучения сходятся во мнении, что в большинстве случаев принятие решений на основе объединения (комитета, ансамбля, смеси и др.) классификаторов является более эффективным, чем с помощью лучшего из них, но единственного [1]. Однако достижение потенциальных преимуществ объединения неизбежно связано с решением двух непростых задач: выбором схемы голосования и множества членов.

Ввиду существования ограниченного числа схем учета голосов и наличия рекомендаций к их выбору [2] имеются хорошие предпосылки для эффективного решения первой задачи. Главным же требованием при выборе членов является соблюдение баланса между их точностью и разнообразием. И если точность можно оценить на валидационной выборке, то измерение разнообразия является не столь очевидным, ввиду большого числа вариантов формализации данного понятия [3]. Обычно под ним подразумевается степень коррелированности ошибок членов объединения, существенное значение которой подтверждается результатами ряда экспериментов. Так, Чжан показал преимущество объединения трех классификаторов, имеющих низкий уровень коррелированности ошибок и точности в районе 67%, перед аналогичным объединением с точностью членов около 95% [4].

Целью описываемого исследования было изучение возможности повышения эффективности одиночных нейросетевых классификаторов путем их объединения в комитеты, для решения задачи распознавания изображений цифр на примере рукописных образов базы MNIST. При этом в работе рассматриваются вопросы создания сетей, выбора схем голосования в комитете, разнообразия его членов и их селекции.

II. СВЕРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ

В качестве базового классификатора исследования была выбрана *сверточная нейронная сеть (СНС, CNN)* архитектуры LeNet-5, изображенная на рис. 1.

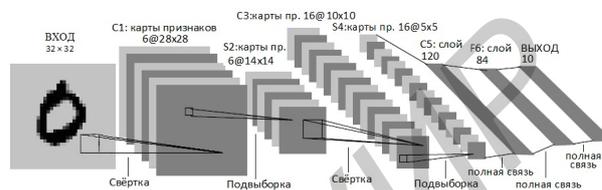


Рис.1 – Архитектура сверточной нейронной сети LeNet-5 [5]

Разработанный Лекуном, данный тип нейронных сетей, по мнению ряда авторов, наилучшим образом подходит для задач визуального анализа документов [6]. Особенности СНС (локальные рецептивные поля, разделяемые веса и пространственные подвыборки), позволяют строить высокоуровневые признаки в автоматическом режиме. Отметим, что применяемая в исследовании сверточная архитектура не содержала слой F6, используемый при большем числе классов.

Обучение СНС осуществлялось модификацией алгоритма обратного распространения ошибки методом Левенберга-Марквардта, в online-режиме. Параметр обучения равномерно уменьшался от 0.001 до 0.000001 в течение 68 эпох. Перед каждой из них тренировочные образы подвергались эластичным (с параметрами: $\sigma = 8$, $\alpha = 36$, [6]) и аффинным искажениям (поворот на $\pm 15^\circ$ и масштабирование в пределах $\pm 15\%$ для каждой размерности отдельно).

В качестве основной базы данных исследования был выбран MNIST [7], являющейся важной точкой отсчета в OCR задачах. Он состоит из тренировочной и тестовой части, с 60000 и 10000 образами, которые использовались для обучения нейронных сетей и контроля их обобщаемости соответственно.

III. КОМИТЕТЫ НЕЙРОННЫХ СЕТЕЙ

Различные модели классификаторов обладают параметрами, вариация которых способна обеспечить хороший уровень баланса между точностью и разнообразием членов объединения. По отношению к нейронным сетям ими являются: 1) инициализация весов; 2) число нейронов; 3) архитектура слоев; 4) тренировочные данные (перечислены в порядке увеличения значимости [8]). Т.к. в качестве базовой модели была выбрана СНС архитектуры LeNet-5, допустимой являлась вариация лишь четвертого, но при этом самого существенного параметра.

С этой целью была применена процедура "регулярного масштабирования": выберем диапазон изменения высоты (h) и ширины (w) символа (учитывая, что стандартный размер образов MNIST 20x20 пикселей, в качестве такового был взят [16, 24] для h , [10, 18] для w с шагом в 2 пикселя) и выполним масштабирование к каждому допустимому размеру

(всего возможны 25 комбинаций с фиксацией h и w , 10 – одного из них и сохранением отношения w/h). В результате можно построить 35 подмножеств базы, с помощью которых обучить 35 СНС и объединить их в комитет. Процесс его применения изображен на рис. 2.

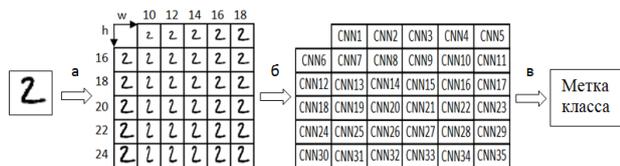


Рис.2 – Схема применения комитета СНС:

- а) регулярное масштабирование; б) распознавание образов; в) объединение решений

Ввиду длительности создания одной СНС (от 2 до 5 часов на стандартном ЭВМ) из требуемых 35 к моменту написания данной статьи были обучены 10 сетей, при следующих размерах образов: ($h = 16, 18, 20, 22, 24, w/h$), ($h = 20, w = 10, 12, 14, 16, 18$). Их точность на тестовой части MNIST составила: 99.14%, 99.39%, 99.39%, 99.35%, 99.36%, 99.28%, 99.33%, 99.41%, 99.34%, 99.44% соответственно.

При создании комитетов СНС использовались наиболее универсальные схемы голосования [2], для которых результатом распознавания является класс: 1) *максимальное* – с максимальным откликом членов комитета; 2) *усредняющее* – с наибольшим средним откликом членов комитета; 3) *большинством* – с наибольшим числом голосов членов в свою пользу. Точность сформированных на их основе комитетов составила: 99.51%, 99.58%, 99.57%, что на 0.07%, 0.14% и 0.13% выше, чем у лучшей одиночной СНС.

IV. СЕЛЕКЦИЯ ЧЛЕНОВ КОМИТЕТА

Построенное с помощью процедуры регулярного масштабирования множество СНС являлось не полным, однако даже в своей сокращенной форме оно позволило создать высокоточные комитеты. Данный факт наряду с вычислительной сложностью обучения СНС, схожестью характеристик сетей множества привели к выводу: формирование комитета должно сопровождаться селекцией членов, целью которой является выбор подкомитета с более высокой или, по крайней мере, равной точностью.

Для исследования данного вопроса был выбран алгоритм EPIC, обладающий рядом преимуществ по сравнению с другими [9]. Его выходом является перечень членов в порядке убывания их вкладов в общую эффективность, интерпретируемый двумя способами: 1) путем пошагового добавления членов строилось десять подкомитетов и выбирался самый точный; 2) подкомитет инициализировался первым членом, добавление остальных велось пока мог быть найден член, повышающий точность. Алгоритмом были сформированы два подкомитета с усредняющей схемой учета голосов, СНС в которых обучались на образах следующих размеров: первым способом – ($h = 16, 18, 24, w/h$), ($h = 20, w = 10, 14$), вторым – те же + ($h = 20, w/h$). Точность распознавания ими тестового MNIST (99.59% и 99.6%), не сильно превышала

уровень всего комитета (99.58%), однако количество членов, по сравнению с ним было уменьшено в 2 раза.

В EPIC корректность распознавания членами тестового образа измеряется бинарными величинами. Возникло предположение о целесообразности их замены значениями выходного вектора СНС, нормированными к $[0, 1]$. Так, модифицированная версия EPIC позволила сформировать комитет с 99.64% точного распознавания тестового MNIST! Отметим, что выбор СНС, обученных на образах размеров: ($h = 16, 24, w/h$), ($h = 20, w = 10, 18$), выглядел не совсем логичным с позиции точности (10-ая, 1-ая, 9-ая и 6-ая), однако с точки зрения разнообразия он не являлся случайным, т.к. данные сети были созданы при максимальных вариациях размеров символов в исследовании.

Подчеркнем также, что полученный результат достаточно уникален, т.к. является лучшим для классификаторов архитектуры LeNet-5 и других не нейросетевых, включая SVM, K-NN и др., при этом международный конкурс по распознаванию образов базы MNIST, проводимый с 1998 г. по настоящее время, является одним из наиболее престижных в области машинного обучения. Более высокими являются лишь результаты швейцарской команды исследователей (0.35%, 0.27%, и 0.23%, 2010 - 2012 гг [10]), полученные благодаря применению громоздких нейросетевых архитектур и мощного оборудования, в частности, нескольких расчетных GPU.

V. ЗАКЛЮЧЕНИЕ

Основными результатами исследования являются: 1) показана перспективность использования СНС как базовой модели для создания систем распознавания образов цифр; 2) предложена эффективная технология построения комитетов СНС на основе множеств образов с различным варьированием их ширины и высоты; 3) показана существенная роль селекции в создании емких и эффективных комитетов СНС; 4) создан классификатор, имеющий один из лучших в мире результатов в распознавании тестового MNIST – 0.36% ошибок на 10000 примерах.

- [1] T. G. Dietterich, "Ensemble methods in machine learning", in: Proceedings of the 1st International Workshop on Multiple Classifier Systems, 2000, pp. 1 – 15.
- [2] L. I. Kuncheva, "Combining Pattern Classifiers. Methods and Algorithms", Wiley, 2004.
- [3] L. I. Kuncheva and C. J. Whitaker, "Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy", Machine Learning, 51(2), 2003, pp. 181 – 207.
- [4] Y. Zhang, S. Burer, and W. N. Street, "Ensemble pruning via semi-definite programming", The Journal of Machine Learning Research, 7, 2006, pp. 1315 – 1338.
- [5] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-Based Learning Applied to Document Recognition", Proceedings of the IEEE, 86(11), 1998, pp. 2278-2324.
- [6] P. Simard, D. Steinkraus, and J. Platt, "Best practices for convolutional neural networks applied to visual document analysis", ICDAR, 2003, pp. 958–963.
- [7] MNIST database, <http://yann.lecun.com/exdb/mnist/index.html>.
- [8] M. S. Haghghi, A. Vahedian, "Making Diversity Enhancement Based on Multiple", Neural Process Lett, 35, 2012, pp. 61-80.
- [9] Z. Lu, X. Wu, X. Zhu, J. Bongard, "Ensemble pruning via individual contribution ordering", KDD, 2010, pp. 871-880.
- [10] D. Ciresan, U. Meier, J. Schmidhuber, "Multi-column deep neural networks for image classification", CVPR, 2012, pp. 3642-3649.