

Анализ свойств ансамбля локально аппроксимирующих моделей

Р. И. Исламов¹, А. В. Грабовой¹, В. В. Стрижов¹

islamov.ri@phystech.edu; grabovoy.av@phystech.edu; strijov@ccas.ru

¹Московский физико-технический институт

Данная работа посвящена анализу свойств ансамбля локальных моделей. Для задачи регрессии предлагается использовать многоуровневый подход, согласно которому множество объектов разбивается на несколько подмножеств и каждому подмножеству соответствует одна локальная модель. Рассматривается задача построения универсального аппроксиматора — мультимодели, которая представлена в виде совокупности локальных моделей. В качестве решающей функции используется выпуклая комбинация локальных моделей. Коэффициенты выпуклой комбинации — шлюзовая функция — функция, значение которой зависит от объекта, для которого производится предсказание. Такой подход позволяет описывать те выборки, которые затруднительно описывать одной моделью. Для анализа свойств проводится вычислительный эксперимент. В качестве данных используются синтетические и реальные выборки. В данной работе реальные данные представлены выборками из boston house prices dataset, servo dataset.

Ключевые слова: локальная модель; линейные модели; ансамбль моделей.

1 Введение

В данной работе исследуется проблема построения мультимодели — ансамбля локальных моделей. *Локальная модель* — модель, которая обрабатывает объекты, находящиеся в определенной связной области в пространстве объектов. В качестве агрегирующей функции используется выпуклая комбинация локальных моделей, при этом веса локальных моделей не постоянны, а зависят от положения объекта в пространстве объектов.

Подход к мультимоделированию предполагает, что вклад каждой локальной модели в ответ зависит от рассматриваемого объекта. Мультимодель использует шлюзовую функцию, которая определяет значимость предсказания каждой локальной модели, входящей в ансамбль.

В данной работе каждая локальная модель является линейной. В качестве функционала качества рассматривается логарифм правдоподобия модели. Предлагается алгоритм нахождения оптимальных параметров ансамбля и локальных моделей.

Преимуществом данного подхода является его способность описывать те выборки, которые затруднительно описывать одной моделью, и разбивать выборку в соответствии с выбранными моделями.

Алгоритмы тестировались на синтетических и реальных данных. Реальные данные представляли собой boston house prices и servo datasets. Эксперименты показали преимущество использования многоуровневой модели и смеси моделей по сравнению с использованием одной модели.

В прикладных задачах данные порождены в результате использования нескольких источников, либо гипотеза порождения и вовсе не известна. В таких случаях качество предсказания можно повышать увеличивая количество моделей. Если моделей на самом деле меньше, чем предполагается, то веса лишних моделей будут малы и их вклад будет несущественен. Этим объясняется актуальность использования мультимоделирования.

2 Работы по теме

С момента своего появления мультимодельный подход стал предметом многих исследований. Были предложены различные типы архитектур локальных моделей, такие как SVM [1], Гауссовский процесс [2] и нейронные сети [3]. Другие работы были сосредоточены на различных конфигурациях, таких как иерархическая структура [4], бесконечное число экспертов [5] и последовательное добавление экспертов [6]. [7] предлагает модель ансамбля локальных моделей для машинного перевода. Стробирующая сеть обучается на предварительно обученной модели NMT ансамбля.

Ансамбль локальных моделей имеет множество приложений в прикладных задачах. Работы [8], [9] [10] посвящены применению смеси экспертов в задачах прогнозирования временных рядов. В работе [11] предложен метод распознавания рукописных цифр. Метод распознавания текстов при помощи ансамбля локальных моделей исследуется в работах [12], а для распознавания речи — в [13], [14]. В работе [15] исследуется смесь экспертов для задачи распознавания трехмерных движений человека.

Литература

- [1] Ronan Collobert, Samy Bengio, and Yoshua Bengio. A parallel mixture of SVMs for very large scale problems. *Neural Computation*, 14(5):1105–1114, may 2002.
- [2] Volker Tresp. Mixtures of gaussian processes. In *Advances in Neural Information Processing Systems 13*, pages 654–660. MIT Press, 2001.
- [3] Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarczyk, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer.
- [4] Michael I. Jordan and Robert A. Jacobs. Hierarchies of adaptive experts. In J. E. Moody, S. J. Hanson, and R. P. Lippmann, editors, *Advances in Neural Information Processing Systems 4*, pages 985–992. Morgan-Kaufmann, 1992.
- [5] Carl Edward Rasmussen and Zoubin Ghahramani. Infinite mixtures of gaussian process experts. In *Advances in Neural Information Processing Systems 14*, pages 881–888. MIT Press, 2001.
- [6] Rahaf Aljundi, Punarjay Chakravarty, and Tinne Tuytelaars. Lifelong learning with a network of experts.
- [7] Ekaterina Garmash and Christof Monz. Ensemble learning for multi-source neural machine translation. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1409–1418, Osaka, Japan, December 2016. The COLING 2016 Organizing Committee.
- [8] M. Serdar Yumlu, Fikret S. Gurgen, and Nesrin Okay. Financial time series prediction using mixture of experts. In *Computer and Information Sciences - ISCIS 2003*, pages 553–560. Springer Berlin Heidelberg, 2003.
- [9] Yiu-ming Cheung, Wai Leung, and Lei Xu. Application of mixture of experts model to financial time series forecasting. October 1995.
- [10] Andreas S. Weigend and Shanming Shi. Predicting daily probability distributions of s&p500 returns. *Journal of Forecasting*, 19(4):375–392, 2000.
- [11] Reza Ebrahimpour, Mohammad Moradian, Alireza Esmkhani, and Farzad Jafarlou. Recognition of persian handwritten digits using characterization loci and mixture of experts. *JDCTA*, 3:42–46, January 2009.
- [12] Andrew Estabrooks and Nathalie Japkowicz. A mixture-of-experts framework for text classification. In *Proceedings of the 2001 workshop on Computational Natural Language Learning*. Association for Computational Linguistics, 2001.

- 71 [13] S. Iman Mossavat, Oliver Amft, Bert de Vries, Petko N. Petkov, and W. Bastiaan Kleijn. A
72 bayesian hierarchical mixture of experts approach to estimate speech quality. In *2010 Second*
73 *International Workshop on Quality of Multimedia Experience (QoMEX)*. IEEE, jun 2010.
- 74 [14] Fengchun Peng, Robert A. Jacobs, and Martin A. Tanner. Bayesian inference in mixtures-of-
75 experts and hierarchical mixtures-of-experts models with an application to speech recognition.
76 *Journal of the American Statistical Association*, 91(435):953–960, sep 1996.
- 77 [15] Cristian Sminchisescu, Atul Kanaujia, and Dimitris N. Metaxas. BM^3e : Discriminative density
78 propagation for visual tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*,
79 29(11):2030–2044, nov 2007.

Поступила в редакцию