

СИНЕРГИЯ АЛГОРИТМОВ КЛАССИФИКАЦИИ (SVM MULTIMODELLING)

Сергей Иванович, Александр Адуенко

sergeyivanychev@gmail.com, aduenko1@gmail.com

В данной статье рассматривается проблема агрегирования небольшого количества сильных классификаторов с целью улучшения решений задач классификации и регрессии. В качестве примера подобной системы рассматривается система SVM алгоритмов использующая kernel-trick с различными ядрами. Для комбинации решений и улучшения качества прогнозирования в задачах классификации и регрессии (SVR) авторы предлагают способ формирования новых признаков на основе сгенерированных отступов (*margins*) каждым классификатором, приводят алгоритм обучения на полученных объектах и анализируют отличия множеств опорных объектов для различных ядер. В качестве практической проверки были проведены эксперименты на различных реальных данных из репозитория UCI.

Ключевые слова: SVM combination, kernel-trick, агрегирование алгоритмов, синергия

Введение

Работа посвящена комбинированию небольшого количества сильных SVM, использующих kernel-trick с различными ядрами и получению агрегированного классификатора для улучшения решений задач классификации и регрессии.

SVM(Support Vector Machine) или *метод опорных векторов* [18][6] [3] — это один из наиболее распространенных и эффективных методов в машинном обучении. Основными достоинствами метода являются:

- Эффективность численных методов решения
- Обозначение опорных объектов из обучающих выборок
- Обобщение на нелинейные классификаторы (kernel-trick)

SVM используются для задач классификации и регрессии (SVR). Задача математического программирования сводится к двойственной задаче, функционалы в которой не зависят от векторов признаков как таковых, а лишь от их попарных скалярных произведений [20]. Использование особых функций, *ядер*, то есть скалярных произведений в сопряженном пространстве, позволяет получить разделяющие поверхности между классами более сложной формы [17]. Наша цель — скомбинировать SVM с различными примененными ядрами для улучшения решения, а также анализ множеств опорных объектов в случае разных использованных ядер.

Наиболее классическими методами агрегирования алго-

ритмов являются бэггинг (*bagging*) [4] и бустинг (*boosting*) [10], и их вариации, однако они работают только с большим количеством слабых классификаторов, что делает невозможным использование его использование для указанного множества базовых алгоритмов.

Среди способов агрегации для небольшого количества классификаторов можно выделить, например, выбор большинства классификаторов [9], комбинирование ранжирований (rankings) по классам, сделанных различными классификаторами [12]. В дальнейшем было показано, что все подобные методы есть особые случаи составного классификатора из [13], появляющиеся при особых условиях или способах аппроксимации.

Различные способы агрегации SVM используются во многих задачах анализа данных. Martin-merino и Mart использовали совокупность SVM для уменьшения ошибочно негативных классификаций (FP) в задаче фильтрации спама среди электронных писем. Для этого на электронных письмах были введены различные метрики, для каждой из них был приспособлен SVM, а затем результат получался голосованием [13]. Gorgevik и Sakmakov, решавшие задачу распознавания написанных рукой символов, делили множество признаков на четыре непересекающихся подмножества, и на каждом из них обучали SVM, увеличив этим самым коэффициент распознавания по сравнению с одним SVM.

В последнее время стал набирать популярность *метод многоядерного обучения* (MKL, multiple kernel learning) [7] [5][1], который основывается на том, что линейная комбинация ядер также является ядром. Данный метод хорош при объединении данных из нескольких источников и полной автоматизации, так как суперпозиция функций может быть оптимизирована любым методом валидации

(например кросс-валидацией).

Мы также предлагаем использовать накопившийся банк ядер, однако не на этапе обучения SVM, а на этапе агрегирования обученных алгоритмов. Известно, что алгоритм b_i для объекта x_j обучающей выборки генерирует *отступ* (margin). По отступу в общем случае можно определить не только предсказанный класс, но и насколько «уверен» в своем решении алгоритм. В случае банка с n ядрами и обучающей выборки с m сэмплами мы получим матрицу отступов $M \in \mathbb{R}^{m \times n}$. Отнормировав ее, мы получим новую матрицу «объект-признак», где вектором признаков каждого объекта будет вектор отнормированных отступов.

В этой работе предложен алгоритм обучения на матрице отступов, проведен анализ опорных объектов, генерируемые различными ядрами, а также проведено тестирование полученного алгоритма на реальных данных репозитория UCI.

Список литературы

- [1] Salah Althloothi и др. «Human activity recognition using multi-features and multiple kernel learning». В: *Pattern Recognition* 47.5 (2014), с. 1800—1812. ISSN: 00313203. DOI: [10.1016/j.patcog.2013.11.032](https://doi.org/10.1016/j.patcog.2013.11.032). URL: <http://dx.doi.org/10.1016/j.patcog.2013.11.032>.
- [2] S. Amari и S. Wu. «Improving support vector machine classifiers by modifying kernel functions». В: *Neural Networks* 12.6 (1999), с. 783—789. ISSN: 08936080. DOI: [10.1016/S0893-6080\(99\)00032-5](https://doi.org/10.1016/S0893-6080(99)00032-5).
- [3] Bernhard E. Boser, Isabelle M. Guyon и Vladimir N. Vapnik. «A Training Algorithm for Optimal Margin Classifiers». В: *Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory* (1992), с. 144—152. ISSN: 0-89791-497-X. DOI: [10.1.1.21.3818](https://doi.org/10.1.1.21.3818). arXiv: [arXiv:1011.1669v3](https://arxiv.org/abs/1011.1669v3). URL: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.21.3818>.
- [4] Leo Breiman. «Bagging Predictors». В: *Machine Learning* 24.421 (1996), с. 123—140. ISSN: 0885-6125. DOI: [10.1007/BF00058655](https://doi.org/10.1007/BF00058655).
- [5] S.S. Bucak, R. Jin и Ak. Jain. «Multiple Kernel Learning for Visual Object Recognition: A Review». В: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36.7 (2014), с. 1354—1369. ISSN: 0162-8828. DOI: [10.1109/TPAMI.2013.212](https://doi.org/10.1109/TPAMI.2013.212). URL: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6654166>.
- [6] Corinna Cortes и Vladimir Vapnik. «Support-Vector Networks». В: *Machine Learning* 20.3 (1995), с. 273—297. ISSN: 15730565. DOI: [10.1023/A:1022627411411](https://doi.org/10.1023/A:1022627411411). arXiv: [arXiv:1011.1669v3](https://arxiv.org/abs/1011.1669v3).
- [7] Martin Dyrba и др. «Multimodal analysis of functional and structural disconnection in Alzheimer's disease using multiple kernel SVM». В: *Human Brain Mapping* 36.6 (2015), с. 2118—2131. ISSN: 10970193. DOI: [10.1002/hbm.22759](https://doi.org/10.1002/hbm.22759).
- [8] Ben Fei и Jinbai Liu. «Binary tree of SVM: A new fast multiclass training and classification algorithm». В: *IEEE Transactions on Neural Networks* 17.3 (2006), с. 696—704. ISSN: 10459227. DOI: [10.1109/TNN.2006.872343](https://doi.org/10.1109/TNN.2006.872343).
- [9] J. Franke и E. Mandler. «A comparison of two approaches for combining the votes of cooperating classifiers». В: *Proceedings, 11th IAPR International Conference on Pattern Recognition. Vol.II. Conference B: Pattern Recognition Methodology and Systems* (1992), с. 1—4. DOI: [10.1109/ICPR.1992.201786](https://doi.org/10.1109/ICPR.1992.201786).
- [10] Y. Freund. *Boosting a Weak Learning Algorithm by Majority*. 1995. DOI: [10.1006/inco.1995.1136](https://doi.org/10.1006/inco.1995.1136). URL: <http://linkinghub.elsevier.com/retrieve/pii/S0890540185711364>.
- [11] D. Gorgevik и D. Cakmakov. «Handwritten Digit Recognition by Combining SVM Classifiers». В: *EUROCON 2005 - The International Conference on Computer as a Tool 2*. February (2005), с. 1393—1396. DOI: [10.1109/EURCON.2005.1630221](https://doi.org/10.1109/EURCON.2005.1630221).
- [12] Tin Kam T.K. Ho и др. «Decision combination in multiple classifier systems». В: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16.1 (1994), с. 66—75. URL: <http://scholar.google.com/scholar?hl=en%7B%5C%7DbtnG=Search%7B%5C%7Dq=intitle:Decision+Combination+in+Multiple+Classifier+Systems%7B%5C%7D0>.
- [13] J. Kittler, M. Hater и R. P W Duin. «Combining classifiers». В: *Proceedings - International Conference on Pattern Recognition* 2.3 (1996), с. 897—901. ISSN: 10514651. DOI: [10.1109/ICPR.1996.547205](https://doi.org/10.1109/ICPR.1996.547205).
- [14] Manuel Martin-merino и Manuel Mart. «Combining SVM Classifiers for Email Anti-spam Filtering». В: 4507. February (2007). DOI: [10.1007/978-3-540-73007-1](https://doi.org/10.1007/978-3-540-73007-1). URL: <http://link.springer.com/10.1007/978-3-540-73007-1>.

- [15] Catherine A. Shipp и Ludmila I. Kuncheva. «Relationships between combination methods and measures of diversity in combining classifiers». В: *Information Fusion* 3.2 (2002), с. 135—148. ISSN: 15662535. DOI: [10.1016/S1566-2535\(02\)00051-9](https://doi.org/10.1016/S1566-2535(02)00051-9).
- [16] European Signal и Processing Conference. «BOOTSTRAP-BASED SVM AGGREGATION FOR CLASS IMBALANCE PROBLEMS S . Sukhanov , A . Merentitis , C . Debes AGT International J . Hahn , A . M . Zoubir Signal Processing Group Technische Universit " at Darmstadt , Germany». В: (2015), с. 165—169.
- [17] Alex J Smola, Bernhard Sch и B Schölkopf. «A Tutorial on Support Vector Regression». В: *Statistics and Computing* 14.3 (2004), с. 199—222. ISSN: 09603174. DOI: [10.1023 / B : STCO . 0000035301.49549.88](https://doi.org/10.1023/B:STCO.0000035301.49549.88).
- [18] Vladimir N Vapnik. «Statistical Learning Theory». В: *Adaptive and learning Systems for Signal Processing, Communications and Control* 2 (1998), с. 1—740. ISSN: 10762787. DOI: [10.2307/1271368](https://doi.org/10.2307/1271368). URL: [http : // eu . wiley . com / WileyCDA / WileyTitle / productCd - 0471030031 . html](http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0471030031.html).
- [19] Vladimir Vapnik и Akshay Vashist. «A new learning paradigm: learning using privileged information.» В: *Neural networks : the official journal of the International Neural Network Society* 22.5-6 (январь. 2009), с. 544—57. ISSN: 1879-2782. DOI: [10 . 1016 / j . neunet . 2009 . 06 . 042](https://doi.org/10.1016/j.neunet.2009.06.042). URL: [http : // www . sciencedirect . com / science / article/pii/S0893608009001130](http://www.sciencedirect.com/science/article/pii/S0893608009001130).
- [20] Константин Вячеславович Воронцов. «Математические методы обучения по прецедентам (теория обучения машин)». В: (). URL: [www . machinelearning . ru / wiki / images / 6/6d/Voron-ML-1.pdf](http://www.machinelearning.ru/wiki/images/6/6d/Voron-ML-1.pdf).