

**О подходах к синтезу случайных и решающих лесов***Дюличева Ю. Ю.*

dyulichева\_yu@mail.ru

Симферополь, Таврический национальный университет им. В. И. Вернадского

Увеличение сложности структуры решающего дерева (РД) и уменьшение его обобщающей способности наблюдаются при все более точной, безошибочной «настройке» РД на исходную обучающую информацию. Такое поведение, характерное для большинства алгоритмов обучения распознаванию, называется переобучением (overfitting) или переобучением. Разработка критериев редукции (pruning) [5] и наращивания (grafting) [9] ветвей решающих деревьев, а также алгоритмов синтеза совокупностей решающих деревьев относительно «простой» структуры — лесов — направлена на поиск компромисса между излишним усложнением структуры РД и получением как можно более высокой оценки качества решающих правил, синтезируемых по обучающей выборке.

Перечислим вкратце основные подходы к синтезу случайных и решающих лесов. Существенным требованием при построении лесов является поиск и синтез различных решающих деревьев, входящих в состав леса. Корректные на обучающей выборке деревья случайного решающего леса (random decision forest), предложенного в работе [6], строятся на подмножествах признаков, случайно выбранных из исходного множества признаков. В работе [4] предложена модель случайных лесов (random forests). При построении каждого дерева случайного леса осуществляется выбор наиболее информативного признака на основе случайным образом сгенерированного для каждой вершины РД подмножества признаков. В работе [8] предложен алгоритм «дровосека» (lumberjack algorithm) для синтеза решающего леса с переходами по ссылкам (linked decision forest), основанный на анализе структуры РД с установкой ссылки перехода на другое дерево леса при появлении в синтезируемом дереве одинаковых поддеревьев. В работе [7] предложен эволюционный подход к построению решающего леса. В работах [1, 2, 3] предложена индуктивная модель эмпирического решающего леса. При построении каждого дерева решающего леса формируется область отказа. Для построения следующего дерева леса используются признаки, не участвовавшие в обучении предыдущих деревьев, и объекты, попавшие в пересечение областей отказа всех деревьев, входящих в состав леса, после чего производится дообучение дерева на оставшихся объектах обучающей выборки.

В докладе приводится сравнительный анализ описанных выше подходов к синтезу случайных и решающих лесов, а также обсуждается ряд

улучшений алгоритма синтеза эмпирического решающего леса, предложенного в работах [1, 2, 3].

### Литература

- [1] *Донской В. И., Дюличева Ю. Ю.* Индуктивная модель  $r$ -корректного эмпирического леса // Труды международной конференции по индуктивному моделированию, Львов. — 2002. — № 2. — С. 54–58.
- [2] *Дюличева Ю. Ю.* Оценка VCD  $r$ -редуцированного эмпирического леса // Таврический вестник информатики и математики. — 2003. — № 2. — С. 35–43.
- [3] *Дюличева Ю. Ю.* Применение эмпирического решающего леса для фильтрации обучающих данных // Таврический вестник информатики и математики. — 2006. — № 1. — С. 55–61.
- [4] *Breiman L.* Random Forests // Machine Learning. — 2001. — No 45. — Pp. 5–32.
- [5] *Esposito F., Malerba D., Semeraro G. A.* Comparative Analysis of Methods for Pruning Decision Trees // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 1997. — Vol. 19, No 5. — Pp. 476–491.
- [6] *Ho T. K.* The Random Subspace Method for Constructing Decision Forests // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 1998. — Vol. 20, No 8. — Pp. 832–844.
- [7] *Rouwhorst S. E., Engelbrecht A. P.* Searching the Forest: Using Decision Trees as Building Blocks for Evolutionary Search in Classification Databases // Congress on Evolutionary Computation CECOO. — 2000. — Vol. 1. — Pp. 633–638.
- [8] *Uther William T. B., Veloso Manuela M.* The Lumberjack Algorithm for Learning Linked Decision Forests // Symp. on Abstraction, Reformulation and Approximation, LNAI. — Springer Verlag, 2000. — Vol. 1864. — Pp. 219–230.
- [9] *Webb G. I.* Decision Tree Grafting // 15th Int. Joint Conf. on Artificial Intelligence, Nagoya, Japan. — Morgan Kaufmann, 1997. — Pp. 846–851.