

Я.К. САЛАЛ, С.М. АБДУЛЛАЕВ

Южно-Уральский государственный университет, Челябинск, Россия
Yasskhudheirsalal@gmail.com, abdullaevsm@susu.ru

АНСАМБЛЕВЫЙ ПОДХОД К ПРОГНОЗУ УСПЕХОВ ОБУЧЕНИЯ (НА ПРИМЕРЕ ШКОЛ ИРАКА)

Общепринятый коллегиальный прогноз индивидуальной и коллективной успеваемости учащихся приводит к решению задач ансамблевой классификации и квантификации с обучением и тестированием ансамблей на несбалансированных выборках. Теоретически и экспериментально показано, что задача бинарной классификации и квантификации решается гетрогенным мета-ансамблем с тренировкой различных классификаторов процедурами бустинга, бэггинга и стэкинга на искусственно сбалансированных тренировочных выборках. Обосновано, что используя вероятностную форму ансамблевого прогноза можно осуществить прогнозу распределения классов успеваемости.

The computer models of collegiate solutions associated with forecast of individual and collective student performance are ensemble classification and quantification learning. The predictive sample in the educational process is always not balanced - it leads to a shift in estimates. The basic idea of this work is that the binary forecasting on unbalanced datasets is carried out by a meta-classifier - an ensemble of different classification algorithm (J48, NB, MLP, NN and SVM) which trained by stacking, bagging and boosting procedures on SMOTE, RUS, ROS balanced samples. The experiments conducted on moderately unbalanced dataset of 350 tuples of student data with using simple voting meta-ensemble models show that the best overall accuracy and F1 measure of minority class were achieved by the meta-ensembles containing 5-15 members classifiers trained on samples prepared by SMOTE. It is also proposed that these of meta-ensembles can be used to resolve quantification learning problem. As in our recent works the deterministic outputs of meta-ensemble members can be transformed to probabilistic outputs and then to prediction of class frequency. After analyzing the results of quantification, it becomes obvious that the most optimal design of the meta-ensembles is to collect pre-boosted classifiers trained on the SMOTE sample.

Ключевые слова: Ансамбли классификаторов и квантификаторов, несбалансированные выборки, мета-ансамбль, вероятностный прогноз

Во многих сферах человеческой деятельности (образование, медицина, законотворчество и др.), сложные решения, касающиеся как судьбы отдельного индивидуума, так и больших групп людей, принимаются коллегиальными экспертами. Очевидно, что коллегиальные решения не всегда являются верными, но общество пока не изобрело лучших вариантов. В нашей работе, рассматриваются два типа прогнозов необходимых для управления образовательным процессом - это прогноз индивидуальной и прогноз коллективной успеваемости учащихся. Алгоритмическими аналогами коллегиальных решений являются задача ансамблевой классификации с обучением в первом типе прогнозов и задача ансамблевой квантификации (прогноз распределения классов успеваемости), во втором случае. Обе задачи осложняются дисбалансом между классами успеваемости.

Нами показано, что в случае бинарной классификации и квантификации обе задачи могут быть решены мета-ансамблем с тренировкой различных классификаторов (J48, MLP, NB, k-NN и SVM) процедурами бустинга, бэггинга и стэкинга на искусственно сбалансированных тренировочных выборках RUS, ROS, SMOTE. Новизна подхода в том, что гетерогенный мета-ансамбль дает возможность различных сочетаний его сильных и слабых компонент, сохраняя одновременно стабильность и вариативность решений. При этом в ансамбле с большим числом компонент всегда от детерминированного прогноза можно перейти к адекватному вероятностному прогнозу и затем к квантификации. Результаты иллюстрируются на реальном примере бинарной классификации успехов 350 учеников средних школ г. Насирия, Ирак.

Данная работа является развитием исследований [1-3].

Список литературы

1. Abdullaev S.M, Salal Y.K. An economic deterministic ensemble classifiers with probabilistic output using for robust quantification: study of unbalanced educational datasets//1st Inter. Scientific and Practical Conf. on Digital Economy (ISCDE 2019), Advances in Economics, Business and Management Res., 2019, v. 105, P. 658-665.
2. Salal Y.K., Abdullaev S.M. Optimization of Classifiers Ensemble Construction: Case Study of Educational Data Mining //Bulletin of the South Ural State University. Ser. Computer Tech., Automatic Control, Radio Electronics. 2019. v. 19, No. 4. P. 139–143.
3. Абдуллаев, С.М. Салал Я.К. Ансамблевая классификация и квантификация: к вопросу о прогнозе индивидуальных и коллективных показателей качества обучения // Непрерывное образование. Материалы пятой международной научно-практической конференции «Университет XXI века в системе непрерывного образования», 22 ноября 2019 г., 2019. том 1. С. 3–13.