

ОЦЕНКА ЭФФЕКТИВНОСТИ АЛГОРИТМОВ ДЛЯ АНАЛИЗА УСПЕВАЕМОСТИ В ВУЗЕ

Сторожев Д.А.

*Белорусский государственный университет информатики и радиоэлектроники, г. Минск, Беларусь,
d.storozhev@gmail.com*

Abstract. The article provides material in which, with the help of the Weka program, the effectiveness of selected algorithms for analyzing performance in a university is evaluated.

Среди задач управления современным вузом обеспечение высокого качества процесса обучения является одной из ключевых.

Качественное обучение поможет сократить время, затрачиваемое студентом на изучение определенного материала, позволит студентам получить больше полезной информации и определить приоритеты в освоении учебных дисциплин.

Одним из показателей качества образования в вузе является успеваемость студентов.

Однако, как показывает практика, контроль успеваемости в вузе усложняется необходимостью анализа образовательных данных, накопленных за длительный период времени.

В настоящее время инструментом, наиболее полезным для поддержки принятия управленческих решений, направленных на повышение качества обучения в вузе, является технология анализа образовательных данных – Educational Data Mining (EDM).

Вместе с тем, чтобы обеспечить высокое качество анализа массивов образовательных данных необходимо использовать соответствующие аналитические инструменты, в том числе эффективные алгоритмы, способные выявить основные закономерности в созданных массивах.

По мнению специалистов, среди алгоритмов классификации наиболее эффективными для решения задач АОД являются алгоритмы на основе деревьев решения, в частности алгоритм J48. Наиболее эффективным алгоритмом кластеризации для решения задач анализа данных является алгоритм неиерархической кластеризации k-means [1].

Для анализа и оценки эффективности алгоритмов J48 и k-means использована программа интеллектуального анализа данных Weka. Программа Weka (Waikato Environment for Knowledge Analysis) – свободное программное обеспечение для анализа данных и машинного обучения, написанное на Java в Университете Уайкато (Новая Зеландия), распространяющееся по лицензии GNU GPL (см. рисунок 1).

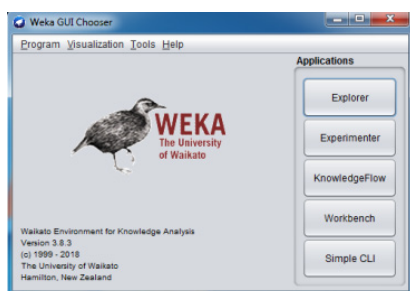


Рисунок 1 –Главное окно программы Weka

Функционально Weka – это набор алгоритмов машинного обучения для задач интеллектуального анализа данных. Она содержит инструменты для подготовки, классификации, регрессии, кластеризации данных, анализа правил сопоставления и визуализации.

Преимущества Weka:

- свободное распространение;
- переносимость, поскольку она полностью реализована на языке программирования Java и, следовательно, работает практически на любой современной вычислительной платформе;
- обширная коллекция методов предварительной обработки данных и моделирования;
- простота использования благодаря графическому интерфейсу пользователя.

Weka поддерживает несколько стандартных задач интеллектуального анализа данных, в частности, предварительную обработку данных, кластеризацию, классификацию, регрессию, визуализацию и выбор функций.

Все методы программы основаны на предположении, что данные доступны в виде одного плоского файла или отношения, где каждая точка данных описывается фиксированным количеством атрибутов (обычно это числовые или номинальные атрибуты, но также поддерживаются некоторые другие типы атрибутов).

Фрагмент структурной схемы платформы Weka представлен на рисунке 2.



Рисунок 2 – Фрагмент структурной схемы платформы Weka

Программа обеспечивает доступ к базам данных SQL с помощью Java Database Connectivity и может обрабатывать результат, возвращаемый запросом к базе данных. Weka предоставляет доступ к глубокому обучению на основе библиотеки DeepLearning4j.

Для оценки эффективности алгоритма классификации J48 предварительно необходимо создать тестовый файл успеваемости студентов. Создаем таблицу успеваемости в книге Excel (см. рисунок 3).

	A	B	C
1	Math	Physic	IT
2	3	4	4
3	3	4	4
4	5	4	4
5	3	4	4
6	5	4	4
7	3	3	3
8	3	4	4
9	3	4	4
10	5	4	4
11	5	4	4
12	3	3	3
13	4	4	4
14	5	5	5
15	4	4	4
16	3	4	4
17	4	3	4
18	4	3	4
19	4	4	4
20	3	3	3
21	4	4	4

Рисунок 3 – Таблица успеваемости студентов по 3-м дисциплинам (по 5-й шкале)

Выполняем экспорт файла в книги Excel в формат CSV для загрузки в программу Weka (см. рисунок 4).

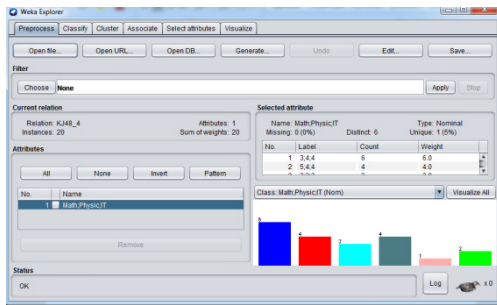


Рисунок 4 – Экран формы подготовки данных для анализа

Выбираем алгоритм J48 и запускаем процедуру классификации (рисунок 5).

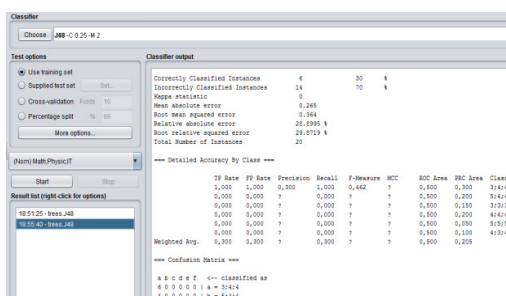


Рисунок 5 – Экран результатов классификации

На рисунке 6 изображена кривая прироста классификации.

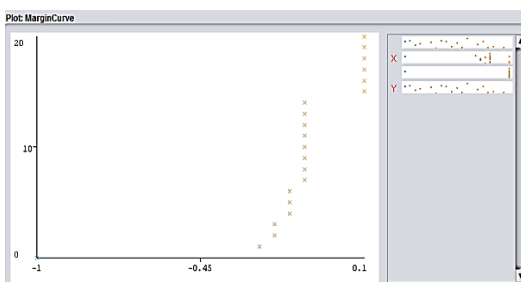


Рисунок 6 – Кривая прироста классификации

Как следует из отчета анализа средняя ошибка классификации не превышает 0.265 и среднеквадратичная

ошибка равна 0.364, что позволяет сделать вывод о высокой точности алгоритма J48.

Для оценки эффективности алгоритма кластеризации k-means выбираем алгоритм SimpleKMeans и запускаем процедуру кластеризации (см. рисунок 7).

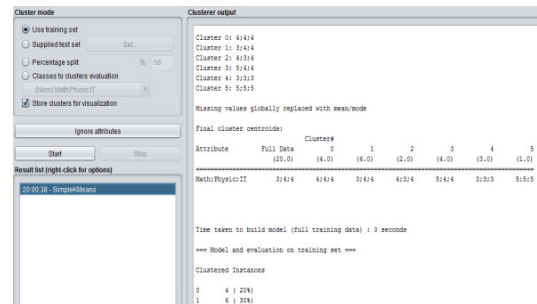


Рисунок 7 – Экран результатов кластеризации

На рисунке 8 представлен график визуализации кластеров.

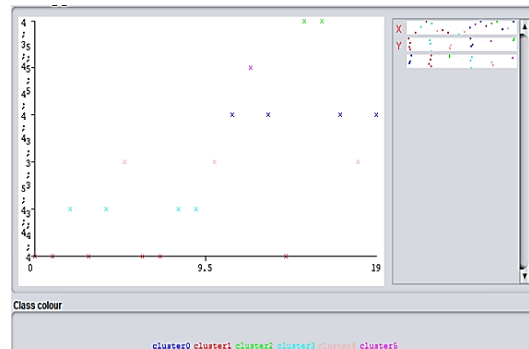


Рисунок 8 – График визуализации кластеров

Как следует из отчета анализа внутри кластера сумма квадратов ошибок равна 0, что позволяет сделать вывод о высокой точности алгоритма k-means.

Полученные результаты позволяют сделать вывод, что выбранные алгоритмы J48 и k-means могут использоваться для анализа успеваемости студентов в вузе.

Таким образом, программа Weka поддерживает такие стандартные задачи интеллектуального анализа данных как кластеризация и классификация, что позволяет ее использовать для оценки эффективности алгоритмов J48 и k-means [2].

Как показал анализ тестовых данных, алгоритмы J48 и k-means обеспечивают высокую точность классификации и кластеризации, что позволяет сделать вывод о высокой эффективности указанных алгоритмов и возможности их использования для анализа успеваемости студентов в вузе.

Литература

1. Прошкина, Е.Н. Анализ и прогнозирование успеваемости студентов на основе радиальной базисной нейронной сети / Е.Н. Прошкина, И.Ю. Балашова. // Материалы III Междунар. науч. конф. (г. Самара, март 2018 г.). – Казань: Молодой ученый, 2018. – С.24-27.

2. Data Mining and Big Data Analytics Technical Committee [Электронный ресурс]. – Режим доступа: <https://cis.ieee.org/technical-committees/data-mining-and-big-data-analytics-technical-committee>.