

4. Administrators - Administrators are responsible for allocating the resources for implementation in institutions.

EDM generally consists of four phases:

1. The first phase of the EDM process (not counting pre-processing) is discovering relationships in data. This involves searching through a repository of data from an educational environment with the goal of finding consistent relationships between variables. Several algorithms for identifying such relationships have been utilized, including classification, regression, clustering, factor analysis, social network analysis, association rule mining, and sequential pattern mining.

2. Discovered relationships must then be validated in order to avoid overfitting.

3. Validated relationships are applied to make predictions about future events in the learning environment.

4. Predictions are used to support decision-making processes and policy decisions.

During phases 3 and 4, data is often visualized or in some other way distilled for human judgment.

Literature

1. Educational data mining [Электронный ресурс]. – Режим доступа: https://en.wikipedia.org/wiki/Educational_data_mining

2. Educational Data Mining: введение [Электронный ресурс]. – Режим доступа: <https://habrahabr.ru/post/181053/>

International Educational Data Mining Society [Электронный ресурс]. – Режим доступа: <http://www.educationaldatamining.org/>

УДК 004.8.032.26; 517.9; 681.3

ОПРЕДЕЛЕНИЕ НАСТУПЛЕНИЯ СОБЫТИЯ С ПРИМЕНЕНИЕМ МОДИФИЦИРОВАННЫХ СТАТИСТИЧЕСКИХ КРИТЕРИЕВ

А.Л. Ляхов, Е.А. Бородина

*Полтавский национальный технический университет имени Юрия
Кондратюка, Украина*

Существует ряд задач в которых необходимо определить степень наступления некоторого события, а также понесённый ущерб от принятия решения, например при аварии на АЭС, определения наличия/отсутствия заболевания у пациента, состояния ценных бумаг на фондовой бирже и т.д.

Классически для определения наступления события проводят статистический анализ в процессе, которого следует сформулировать и проверить предположение. При этом проверка статистических гипотез осуществляется на основании выборочных данных, ограниченного ряда наблюдений, решения имеют вероятностный характер. Такое решение сопровождается вероятностью ошибочного заключения, тем самым вызывая ошибки первого /второго рода. Как известно увеличение вероятности одной из ошибок приводит к снижению другой, поэтому принято задавать вероятность ошибки первого рода, а именно уровень значимости. Необходимо всегда помнить, что процедура проверки статистических гипотез может гарантировать результаты с определенной вероятностью лишь на основе объективных данных. Для проверки статистических гипотез применяют статистические критерии, которые позволяют определить меру расхождения результатов выборочного наблюдения с высказанной гипотезой [1].

Известно, что правило принятия решения для статистических критериев одинаково и сводится к сравнению отношения правдоподобия с пороговыми значениями. Отличие заключается лишь в величине порогового значения отношения правдоподобия. Однако существующие статистические критерии направлены на определение вероятности наступления того или иного события. Данные критерии не позволяют определить и учесть полный материальный ущерб от принятия решения. Поэтому предложен подход к решению задачи модифицирования статистического критерия идеального наблюдателя и критерия Неймана-Пирсона в задачах распознавания образов с учетом особенностей предметной области объекта исследования. Данный подход позволяет осуществлять выбор порогов распределения классов в пространстве информативных признаков с минимизацией ожидаемых потерь в предметной области.

Поставленная задача формализуется выражением:

$$U_{og} = \varphi [(\alpha, \beta / X_{пор}) \cdot A_{jm}], \quad (1)$$

при условии, что $\alpha = P\{T \in \Omega_\alpha | H_0\}$ и $\beta = P\{T \notin \Omega_\alpha | H_1\}$; $X_{пор}$ – порог раздела признаков пространственных классов, $A_{jm} \in M_{nm}$, где M_{nm} – матрица потерь.

Порог принятия решения играет ключевую роль и в простом случае должен быть связан с точкой (или областью), которая делит величину ущерба пополам. Очевидно, что при этом будут отражены особенности

предметной области, которые учтены в элементах матрицы потерь. Выбор порога для разделения пространства признаков разных классов требует пояснений (рис.1).

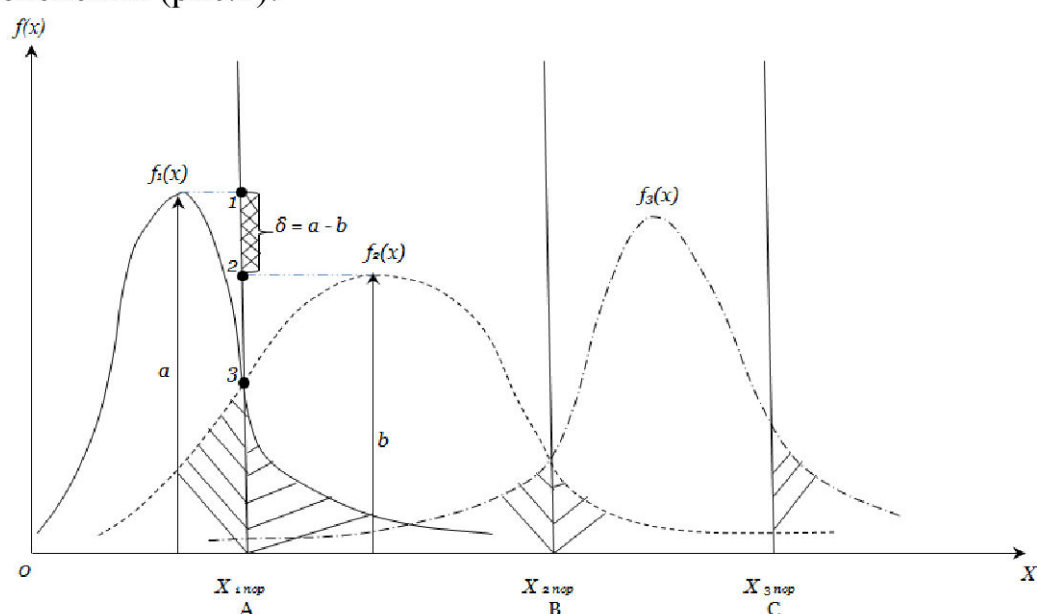


Рис.1. Демонстрация несовпадений вероятностей ошибок и физических потерь от них

Для определения порога разделения классов с учетом ошибок первого и второго рода разработан алгоритм, позволяющий оценивать вероятности ошибок первого и второго рода, рассчитывать весовые коэффициенты потерь. Весовые коэффициенты потерь реализованы с вероятностями соответствующих ошибок. Произведение максимально возможных потерь на вероятность появления этих потерь позволило количественно оценить ожидаемый материальный ущерб. Реализация этого произведения осуществлена на базе пакета технического анализа данных объединением вероятностных и линейных нейросетевых моделей. Следствием является модификация статистического критерия Неймана-Пирсона по тому же принципу [2].

Модификация статистических критериев, обеспечивающая минимизацию ожидаемых потерь при наступлении того или иного события, а также выбора порога разделения классов в пространстве информативных признаков, с учетом дифференциации потерь от ошибки первого и второго рода позволяет перейти к разработке ИТ, под которой понимается процесс обработки сгенерированной выборки, распределенной по классам для получения информации относительно распределения порогов между классами с учетом ущерба и без учета.

Результатом реализации ИТ является определение вероятности попадания каждого значения с пространства информативных признаков к определенному классу, нахождение порогового значения между всеми

классами без учета ущерба и с учетом, а также расчёт уровня ущерба от принятия решения. К основным этапам реализации разработанной информационной технологии относится:

1. Проведение анализа выборки, предназначенной для обучения нейронной сети, производится разбиение наблюдений на классы.
2. Обучение нейронной сети.
3. Определение плотности нормального распределения для каждого класса
4. Определение пороговых значений между классами представленной выборки.
5. Расчет выше указанных параметров с учетом ущерба.
6. Построение графика распределения значений для каждого класса с указанием порогов между классами.
7. Построение графика распределения значений для каждого класса с указанием порогов между классами с учетом ущерба.
8. Вычисление ошибок первого и второго рода с учетом установленных порогов.
9. Вычисление ошибок первого и второго рода с учетом установленных порогов, включающих значение ущерба.
10. Вычисление ущерба от принятия решения.

Модифицированные статистические критерии позволяют определять не вероятность наступления события, а понесенный ущерб от данного события. Что позволяет изменить установление пороговых значений, тем самым переводя решение задачи на новый уровень (рис. 2).

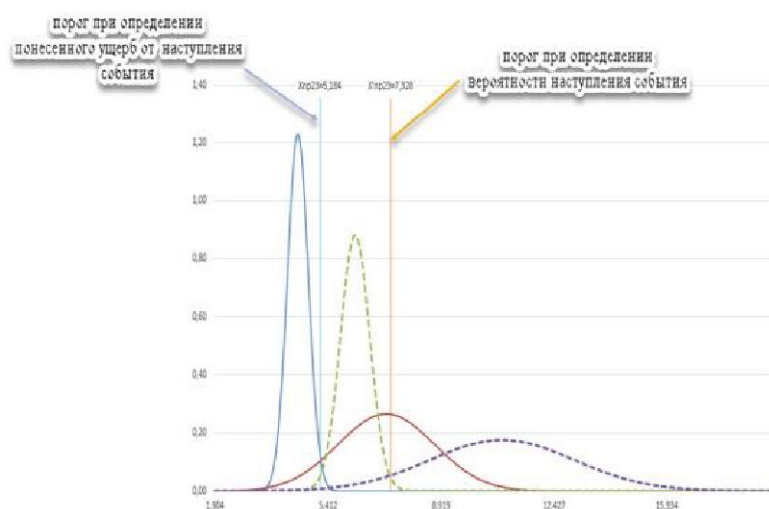


Рис. 2. Смысловое значение порогов с модифицированными статистическими критериями и без

Литература

1. Процесс и критерии проверки статистических гипотез [Электронный ресурс] – Режим доступа: <https://xreferat.com/113/1279-1-process-i-kriterii-proverki-statisticheskikh-gipotez.html>
2. Алёшин С.П., Ляхов А.Л., Бородин Е.А. Нейросетевая оценка предметных рисков принятия решений по величине прогноза предотвращенного ущерба. /Алёшин С.П., Ляхов А.Л., Бородин Е.А. // Научные ведомости БелГУ (Научный рецензируемый журнал). – 2014. – №21 (192) – Выпуск 32/1 – С. 142–147. – Белгород, (Россия).

УДК 004.94

ОБРОБКА МУЛЬТИМЕДІЙНОЇ ІНФОРМАЦІЇ З ВИКОРИСТАННЯМ ТЕХНОЛОГІЙ ОБРОБКИ НАДВЕЛИКИХ МАСИВІВ ІНФОРМАЦІЇ ТА ТЕХНОЛОГІЇ NVIDIA CUDA

В.М. Романчук

Національний технічний університет України “КПІ імені Ігоря Сікорського” кафедра автоматизованих систем обробки інформації та управління, України

Вступ. Проблема обробки великих даних з кожним днем стає все гострішою оскільки кількість інформації з кожним днем зростає все швидше. У даній роботі розглянуто використання Apache Spark та NVIDIA CUDA для обробки графічної та мультимедійної інформації та вирішення задачі класифікації..

Задачі дослідження.

– завдання визначити технологічний стек на основі системи обробки надвеликих масивів інформації.

– виконати експериментальне дослідження роботи комплексу задач на прикладі розпізнавання та класифікації облич на основі відеоряду чи окремих графічних файлів

Основна частина

Для порівняння вхідного зображення із збереженими шаблонами використовуємо відстань Хемінга. В якості порівнюваних векторів використовуємо вектори ключових рис облич визначені за методом Вахідом Каземі та Джозефін Салліван [1].

У якості програмної реалізації використовуємо бібліотеку OPENCV з підтримкою NVIDIA CUDA та Apache Spark. NVIDIA CUDA зарекомендувала себе як високопродуктивна технологія [1].