

УДК 004.6

КОНСТРУКТОР ДЛЯ СОЗДАНИЯ ЦИФРОВЫХ ПОРТРЕТОВ И ПИКТОГРАФИЧЕСКОЙ ВИЗУАЛИЗАЦИИ МНОГОФАКТОРНЫХ ДАННЫХ

Турусина Е.А. (ИГУ)

Научный руководитель – кандидат физико-математических наук, доцент Балахчи А.Г.
(ИГУ)

Введение. Современные задачи анализа данных требуют инструментов, способных преобразовывать сложные многомерные наборы информации в интуитивно понятные визуальные представления. Традиционные методы визуализации, такие как гистограммы или диаграммы рассеяния, зачастую неэффективны для выявления скрытых взаимосвязей между десятками переменных. В этом контексте пиктографическая визуализация становится перспективным решением, позволяющим отображать каждое наблюдение в виде уникального символа, где параметры (форма, цвет, размер) кодируют значения переменных. За рубежом проекты, такие как «Capturing People Through Data» и дата-портреты, представленные на конференции TED, представляют собой узко специализированные пользовательские разработки, созданные в контексте конкретных мероприятий. Данные проекты не являются полноценными системами и не могут быть адаптированы для использования в различных отраслях. В России аналогичных инструментов практически нет, что обуславливает необходимость создания программного продукта, автоматизирующего процесс создания пиктографиков для визуализации многомерных данных.

Основная часть. Система создания цифровых портретов для визуализации данных представляет собой инструмент, предназначенный для преобразования числовых и текстовых данных в пиктографические визуализации. Пиктографики позволяют отображать данные в виде многомерных символов, где различные параметры, такие как форма, размер и цвет, соотносятся с значениями переменных. Этот подход способствует интуитивному восприятию сложных взаимосвязей между переменными и позволяет аналитикам более эффективно выявлять паттерны, требующие дальнейшего исследования.

Цель системы заключается в предоставлении пользователям средства для эффективного анализа данных посредством преобразования числовой и текстовой информации в пиктографические визуализации, что способствует улучшению понимания взаимосвязей между переменными и поддерживает процесс принятия обоснованных решений. Ключевые функции системы включают ускорение процесса выявления зависимостей, упрощение восприятия многомерных данных с использованием визуальных метафор, а также содействие переходу от интуитивного восприятия данных к формальному анализу. Визуализация данных позволяет пользователю выявить как простые, так и сложные взаимодействия между переменными на этапе предварительного анализа, что ускоряет процесс принятия решений.

Основные компоненты системы включают следующие модули:

- 1) Модуль импорта данных, который поддерживает загрузку данных через CSV-файлы и ручной ввод. Система автоматически выполняет валидацию структуры данных, включая проверку разделителей, кодировки и типов переменных, а также предоставляет предпросмотр загружаемых данных.
- 2) Модуль обработки данных, позволяющий выполнять фильтрацию данных по диапазонам или категориям, а также сортировку по возрастанию или убыванию. Пользователь может выбрать поля для анализа и настроить параметры агрегации данных, что позволяет гибко работать с входными данными.
- 3) Модуль визуализации генерирует пиктографические представления на основе шаблонов, где каждой переменной соответствует определённый визуальный атрибут: форма символа для категориальных данных (например, тип продукта), цвет и градиент для количественных или порядковых данных (например, уровень продаж), размер для отображения интенсивности показателя (например, частота событий).

4) Модуль экспорта, который позволяет сохранять результаты визуализации в форматах PNG или JPG, что обеспечивает возможность интеграции визуализаций в отчёты и презентации.

Выводы. Система может быть использована для анализа данных в разных областях, например, в маркетинге для исследования целевой аудитории, в образовании для визуализации характеристик студентов, в медицине для наглядного отображения медицинских показателей [1], в экологии — для мониторинга изменений климата или визуализации результатов ботанических исследований [2]. В каждой из этих областей разработка будет способствовать улучшению качества анализа данных, ускорению процесса принятия решений и повышению доступности информации для пользователей с разным уровнем подготовки.

Для внедрения системы необходимо провести этап тестирования с участием потенциальных пользователей из различных областей. Это позволит выявить ключевые потребности и оптимизировать функционал в зависимости от специфики задач. В процессе испытаний особое внимание следует уделить функционалу настройки визуализаций и удобству работы с большими объемами данных. Также рекомендовано интегрировать методы машинного обучения для улучшения точности предложений по визуализации и ускорения анализа данных.

Список использованных источников:

1. Осадчая Ирина Анатольевна, Берестнева Ольга Григорьевна, Немеров Евгений Владимирович Анализ многомерных медицинских данных с помощью пиктографиков «Лица Чернова» // Бюллетень сибирской медицины. 2014. №4. URL: <https://cyberleninka.ru/article/n/analiz-mnogomernyh-meditsinskih-dannyh-s-pomoschyu-piktografikov-litsa-chernova> (дата обращения: 15.02.2025).
2. Давиденко О.Н., Давиденко Т.Н., Невский С.А., Пискунов В.В. Способы визуализации данных в ботанических и экологических исследованиях // Учебно-методическое пособие для студентов биологического факультета. Саратов; 2013: 41 с. URL: http://elibrary.sgu.ru/uch_lit/855.pdf.