

МЕТОДЫ АНАЛИЗА И ВИЗУАЛИЗАЦИИ БОЛЬШИХ ДАННЫХ В ИНФОРМАЦИОННЫХ СИСТЕМАХ

В данной статье рассмотрены основные методы, на основе использования которых, может осуществляться обработка и визуализация данных любой сложности, что позволит создавать системы для автоматизирования процессов хранения, обработки и визуального представления больших данных, позволит ускорить процесс работы с этими данными, уменьшит количество данных с неполной информацией, минимизирует экономические затраты предприятия.

ВВЕДЕНИЕ

Вся история человечества – это огромный массив данных. Данные стали неотъемлемой частью истории, политики, науки, экономики и бизнеса, а сейчас и социальной жизни человека. В современных организациях различные датчики, видеокамеры, интеллектуальные счетчики и другие подключенные устройства генерируют огромные объемы данных, которые добавляются к уже хранящейся информации [1]. В связи с этим большую ценность приобретают системы, позволяющие автоматически осуществлять анализ этих данных, выявлять закономерности, которые не может обнаружить человеческий мозг и для более легкого восприятия результатов анализа представлять их данных в удобном виде. Верно подобранный метод визуализации данных может принести ряд следующих преимуществ: увеличится скорость принятия решений, улучшится понимание данных, уменьшится количество ошибок и данных с неполной информацией, минимизируются экономические издержки организации.

СОДЕРЖАНИЕ ДОКУМЕНТА

Массивы информации, которые невозможно обработать или проанализировать при помощи традиционных методов с использованием человеческого труда и настольных компьютеров носят название «Больших данных» [1]. Для грамотного выбора метода для анализа и визуализации больших объемов информации требуется определить тип данных.

Существуют следующие виды данных, с которыми могут работать системы анализа и визуализации данных [2]:

- Одномерные данные;
- Двумерные данные;
- Многомерные данные;
- Тексты и гипертексты;
- Иерархические и связанные структуры;
- Алгоритмы и программы.

Для визуализации перечисленных типов данных используются различные визуальные образы и методы их создания, которые можно разделить на несколько типов [2]:

- Стандартные 2D/3D-образы – гистограммы, линейные графики и т.п.;
- Иерархические образы – наложение измерений и древовидные карты;
- Геометрические преобразования – диаграмма разброса данных и т.п.;
- Отображение иконок – линейчатые фигуры и звезды;
- Методы, ориентированные на пиксели – рекурсивные шаблоны и т.п.

Для наиболее эффективных анализа и визуализации больших данных чаще всего используются методы геометрических преобразований, методы отображения иконок и методы, основанные на пиксели. Рассмотрим каждый из них подробнее.

Метод геометрических преобразований направлен на трансформацию многомерных наборов данных с целью отображения их в декартовом и в недекартовом геометрических пространствах. Рассмотрим частные виды применения данного метода.

График рассеяния или точечная диаграмма (см. рис. 1). Точечные диаграммы часто используются для визуализации взаимосвязей между данными в двух измерениях, помогают найти взаимосвязь между двумя показателями.

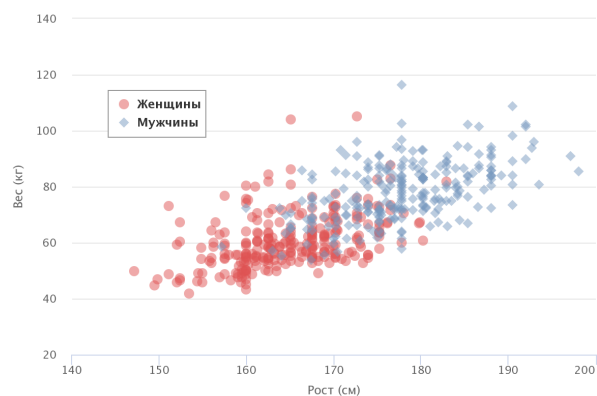


Рис. 1 – Точечная диаграмма зависимости роста и веса в зависимости от пола

Диаграмма на рисунке 1 визуализирует рост и вес в зависимости от пола, демонстрируя четкую тенденцию, согласно которой мужчины в среднем выше и тяжелее женщин.

Пузырьковая диаграмма (см. рис. 2). Данный график позволяет сравнить два параметра по третьему. Размеры пузырьков определяются значениями третьего ряда данных. Пузырьки разных размеров позволяют визуально выделить конкретные значения.

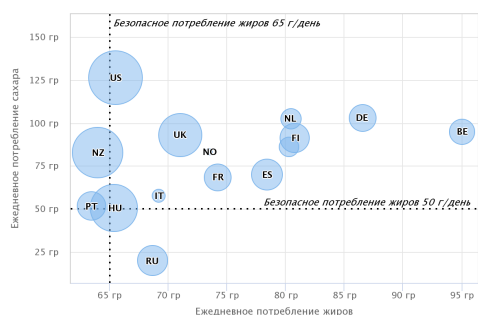


Рис. 2 – Пузырьковая диаграмма потребления сахара и жиров по странам

Географическая диаграмма используется для отображения распределения определенного показателя по регионам, странам, материкам (см. рис. 3).

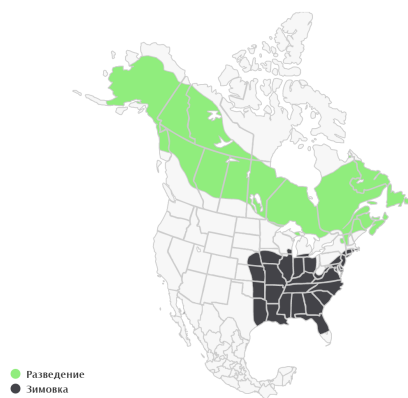


Рис. 3 – Географическая диаграмма, показывающая место разведения и зимовки ржавчатого малого трупяла

Следующие методы визуализации больших данных – это методы отображения иконок (см. рис. 4). Основная идея данного метода состоит в том, чтобы сопоставить значения многомерных элементов данных со свойствами образцов, которые могут представлять: человеческие лица, стрелки, звезды и т.д. Визуализация создается с помощью сопоставления атрибутов элементов данных со свойствами образцов. Такие образы можно сгруппировать для целостного анализа данных. Результирующая визуализация представляет собой образец текстур, которые име-

ют различия, соответствующие характеристикам данных [3].



Рис. 4 – Метод отображения иконок на примере социального взаимодействия в Австралии

Методы, ориентированные на пиксели (см. рис. 5). Основная идея заключается в том, чтобы отображать измерения значения в каждом в цветном пикселе и объединять их согласно принадлежности к измерению [3].

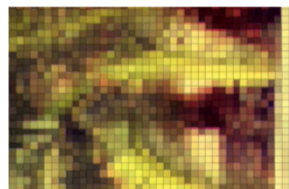


Рис. 5 – Пример использования метода, ориентированного на пиксели

Нельзя однозначно выделить один наилучший и универсальный метод визуализации данных, так как каждый метод предназначен для своих целей, и только разработчик систем анализа и визуализации данных может самостоятельно выбрать метод, подходящий для решения его задач [1].

СПИСОК ЛИТЕРАТУРЫ

1. Иванов, П. Д., Лопуховский, А. Г. Технологии Big Data и различные методы представления больших данных / Иванов П. Д., Лопуховский А. Г. // Инженерный журнал: наука и инновации – №9, 2014. – Режим доступа: <http://engjournal.ru/catalog/it/asu/1229.html>. – Дата доступа: 20.04.2021.
2. Берченко, Д. А. Аналитический обзор методов визуализации данных / Д. А. Берченко, П. Г. Круг // Евразийский Научный Журнал – №5, 2017. – Режим доступа: <http://journalpro.ru/pdf-article/?id=8506>. – Дата доступа: 20.04.2021.
3. Olshannikova, E., Ometov, A., Koucheryavy, Y., Olsson, T. Big Data In Data Technologies and Applications / Borko F., Flavio V. // Big Data Technologies and Applications – pp. 101-131.

Куликова Алина Валдимировна, студентка кафедры систем управления БГУИР, m.andreychuk@icloud.com.

Научный руководитель: Захарьев Вадим Анатольевич, преподаватель кафедры систем управления БГУИР, кандидат технических наук, доцент, zahariev@bsuir.by.