

Студ. П.Р. Гуринович
Науч. рук. доц. Н.В. Пацей
(кафедра программной инженерии, БГТУ)

СЕРВИС АНАЛИТИЧЕСКОЙ ОБРАБОТКИ И ВИЗУАЛИЗАЦИИ ДАННЫХ

В настоящее время мы владеем большим количеством информации. Чтобы улучшить процесс анализа и восприятия данных существуют алгоритмы преобразования и визуализации данных. Визуальная информация лучше воспринимается и позволяет быстро и эффективно донести собственные мысли и идеи. Физиологически, восприятие визуальной информации является основной для человека. Есть многочисленные исследования, подтверждающие, что:

1. 90% информации человек воспринимает через зрение
2. 70% сенсорных рецепторов находятся в глазах
3. на 19% меньше при работе с визуальными данными используется когнитивная функция мозга, отвечающая за обработку и анализ информации
4. на 17% выше производительность человека, работающего с визуальной информацией

Визуализация данных — это наглядное представление массивов различной информации. Существует несколько типов визуализации:

1. Графики — показывают зависимость данных друг от друга. Строятся по осям X и Y, хотя могут быть и трехмерными. Например, линейные графики, графики рассеивания.

2. Диаграммы сравнения — показывают соотношения набора данных. Во многих случаях строятся вокруг осей, хотя и необязательно. Например, столбцовая диаграмма (bar chart), гистограмма (histogram), круговая диаграмма (pie chart), кольцевая диаграмма (ring chart) и т. д.

3. Деревья и структурные диаграммы — показывают структуру набора данных и взаимосвязи между его элементами. Например, граф, дерево, плоское дерево (tree map), формализованные структурные диаграммы (uml-диаграммы)

4. Диаграммы времени — показывают распределение данных в зависимости от времени. Например, временная шкала, диаграмма Ганта.

5. Карты — показывают данные, зависящие от географии или архитектуры некоего объекта. Например, географическая карта, фотографическая карта, дорожная карта и т. д.

6. Диаграммы связей — показывают связи внутри набора данных, как правило, достаточно большого. Например, круговая диаграмма связей (network diagram, arc diagram), дендрограмма и т. д.

Для того чтобы визуализация данных было эффективно, необходимо применять алгоритмы преобразования (предобработки) данных. Существуют несколько основных разновидностей алгоритмов:

1. Классификации — цель данного алгоритма определить, к какому классу относятся те или иные данные; при этом множество классов, к одному из которых впоследствии можно отнести исследуемый объект, заранее известно. Каждый класс обладает определенными свойствами, которые характеризуют его объекты. Наглядный пример — задача выяснения кредитоспособности клиента. Банковский служащий при этом оперирует двумя известными ему классами — «кредитоспособный» и «некредитоспособный».

2. Кластеризации — это задача разбиения множества объектов на группы, называемые кластерами. Внутри каждой группы должны оказаться «похожие» объекты, а объекты разных группы должны быть как можно более отличны. Главное отличие кластеризации от классификации состоит в том, что перечень групп четко не задан и определяется в процессе работы алгоритма.

3. Регрессии — построение параметрической функции, описывающей изменение указанной числовой величины в указанный промежуток времени. Эта функция строится на основе известных данных, а затем используется для предсказания дальнейших значений этой же величины. На вход метода поступает последовательность пар вида «время — значение», описывающая поведение этой величины при заданных условиях.

4. Поиск закономерностей — суть метода заключается в нахождении правил, описывающих взаимозависимости между внутренними элементами данных. Классическим примером является анализ покупок в супермаркете и выявление правил вида «если человек покупает пельмени, то обычно он покупает еще и сметану».

5. Многомерный анализ — суть метода заключается в построении многомерного куба и получении его различных срезов. Результатом анализа, как правило, является таблица, в ячейках которой содержатся агрегированные показатели (количество, среднее, минимальное или максимальное значение и так далее).

Таким образом, цель выполненной научной работы заключается в улучшение процесса анализа пользовательских данных на основе оперативной предобработки и визуализации данных. Разработанный сервис должен применять схемы преобразования данных в форму,

усиливающую восприятия, а также реализовать алгоритмы визуализации. На рисунке 1 представлена структурная схема сервиса:

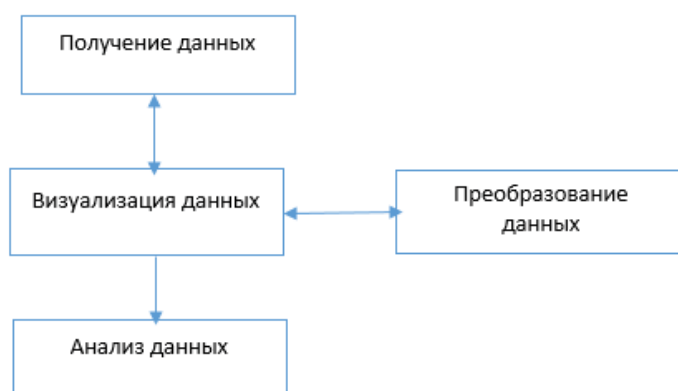


Рисунок 5 – Структурная схема сервиса

Модуль получения данных представляет собой загрузку в сервис пользовательских данных. Модуль преобразования данных – обработка пользовательских данных. Модуль визуализации данных – визуальное представление данных.

В заключение можно отметить, что разработанный сервис имеет преимущество перед его аналогами за счет того, что он предоставляет пользователю визуализировать как исходные данные, так и преобразованные в соответствующим алгоритмом.

ЛИТЕРАТУРА

1. Зильберман М. А. «Инфографика»- 2013г.
2. Методы интеллектуального анализа данных. [Электронный ресурс] - Режим доступа: <http://datareview.info/article/dannyie-kak-poleznyie-iskopaemyie-osnovnyie-metodyi-data-mining/> (дата обращения 29.03.2018).
3. Визуализация данных. Классификация [Электронный ресурс] - Режим доступа: <http://jvetrau.com/visualization-1/> (дата обращения 10.04.2018).