
Аналитический обзор методов визуализации данных

Берченко Дианна Андреевна

Магистрант,
Московский технологический университет (МИРЭА),
Россия, г. Москва
E-mail: dianna-dda@yandex.ru

Научный руководитель: **Круг Петр Германович**

д. т. н., профессор.
Кафедра «Автоматизированные системы управления»
Московский технологический университет (МИРЭА),
Россия, г. Москва

Ключевые слова: АНАЛИЗ, МЕТОДЫ ВИЗУАЛИЗАЦИИ, ЭКСПЕРТНЫЕ СИСТЕМЫ, ОБРАБОТКА ДАННЫХ.

Метод визуализации можно отнести к методам, рассчитанным на экспертные системы. Потому что, именно это метод позволяет сочетать находчивость человеческого ума, гибкость восприятия человеком «среды» и невероятные вычислительные ресурсы современных технологических решений.

Метод визуализации рассматривается как системное, основанное на правилах, динамическое и/или статическое графическое представление информации, способствующее «рождению» идей, помогающее разобраться в сложных понятиях, нацеленное на обобщение, анализ теории и опыта

То есть, основная идея методов визуализации заключается в предоставлении человеку-эксперту большой объем данных в форме, в которой будет удобно воспринимать информацию и проводить анализ.

Пользователю наиболее удобно работать с данными напрямую, рассматривать их с разных сторон и под различными углами зрения. Благодаря этому пользователь получает дополнительную информацию, которая ему помогает более четко сформулировать цели и задачи исследования, прийти к их решению. Для этих задач наиболее удобным является представление в виде визуальных образов. Полезность визуального анализа наиболее велика, если цель самого исследования не определена до конца и недостаточно информации о самих данных. Таким образом, можно сказать, что визуальный анализ данных является процессом генерации гипотез. Сгенерированные таким образом гипотезы можно проверить или автоматическими средствами используя Data Mining, или вновь прибегнуть к визуальному анализу для уточнения.

Такой подход имеет два основных преимущества:

- легко работать с неоднородными и зашумленными данными, не многие автоматические средства обработки данных могут это делать;
- представление данных с помощью визуальных образов интуитивно понятен и не требует сложных математических или статистических алгоритмов.

Визуальный анализ данных можно разделить на три этапа:

1. Беглый анализ — позволяет идентифицировать интересные шаблоны и сфокусироваться на одном или нескольких из них;
2. Увеличение и фильтрация — идентифицированные на предыдущем этапе шаблоны

отфильтровываются в большом масштабе;

3. Детализация по необходимости — если пользователю нужно получить дополнительную информацию, он может визуализировать более детальные данные.

На самом деле исследователь сам решает на каком этапе он получает достаточное количество знаний и на каком уровне детализации он может остановиться.

Выделяют следующие виды данных, с которыми могут работать средства визуализации:

- Одномерные данные;
- Двумерные данные;
- Многомерные данные(в состав многомерных данных можно включить и двумерные);
- Тексты и гипертексты;
- Иерархические и связанные структуры;
- Алгоритмы и программы.

Для визуализации перечисленных типов данных используются различные визуальные образы и методы их создания.

Методы визуализации можно разделить на следующие типы:

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

Графики, гистограммы, диаграммы, и т.п. — самые простые методы визуализации. Основной недостаток этого метода — невозможность легко воспринимаемой визуализации сложных данных и данных в большом количестве.

Методы геометрического преобразования направлены на трансформацию многомерных наборов данных с целью отображения их геометрических пространствах в декартовом и в недекартовом.

Другим классом визуализации данных являются методы отображения иконок. Их основной идеей является отображение значений элементов многомерных данных в свойства образов, которые могут представлять собой: человеческие лица, стрелки, звезды и т.д. Визуализация генерируется отображением атрибутов элементов данных в свойства образов. Такие образы можно группировать для целостного анализа данных. Результирующая визуализация представляет собой шаблон текстур, которые имеют различия, соответствующие характеристикам данных.

Основной идеей методов, ориентированных на пиксели, является отображение каждого измерения значения в цветной пиксель и из группировки по принадлежности к измерению. Так как один пиксель используется для отображения одного значения, то, следовательно, данный метод позволяет визуализировать большое количество данных (свыше одного миллиона значений).

Методы иерархических образов предназначены для представления данных, имеющих иерархическую структуру. В случае многомерных данных должны быть правильно выбраны измерения, которые используются для построения иерархии.

Эти методы можно сравнить по основным параметрам (Таблица 1): типу данных, к которым тот или иной метод можно применить и возможность применения этого метода для обработки больших данных.

Метод визуализации	Обрабатываемые данные					Обработка больших данных
	Одномерные	Многомерные	Тексты /гипертексты	Иерархические и связанные структуры	Алгоритмы и программы	
Стандартные 2D/3D-образы	+	–	–	–	–	–
Геометрические преобразования	+	+	–	–	+	+
Отображение иконок	–	+	+	–	–	+
Методы, ориентированные на пиксели	+	+	+	+	+	+
Иерархические образы	–	+	–	+	–	–

Таблица 1 — Сравнение методов визуализации

Нельзя однозначно выделить один наилучший и универсальный метод визуализации данных, так как каждый метод предназначен для своих целей, и только исследователь может самостоятельно выбрать метод, подходящий для решения его задач.

Ни один метод визуализации, конечно, не совершенен, как и не совершенен сам по себе ни один другой метод анализа данных. Но у исследователя всегда есть возможность совместить различные способы обработки и анализа данных, чтобы получить максимальные знания — экспертизу. Так же важно помнить, что новая информация — дает толчок к развитию технологий, а новые технологии — помогают улучшить качество получаемых из информации знаний.

Список используемых источников

1. Feldman R., Sanger J. The Text Mining Handbook, Cambridge University Press, 2007.
2. Причины и источники сетевых аномалий // Молодой ученый. Оладько В. С., Микова С. Ю., Нестеренко М. А., Садовник Е. А.— 2015. — № 22. — С. 158-161.