

Исследование существующих визуальных моделей многомерных данных и их метафор визуализации

А.С. Курилов
kac.kurilov@yandex.ru

Брянский Государственный Технический Университет, Брянск, Россия

Анализ и интерпретация многомерных данных является очень важной проблемой во многих сферах жизни человека, таких как производство продукции, диагностика заболеваний, маркетинг, информационная безопасность и т. п. Однако иногда анализ многомерных данных с использованием традиционных подходов оказывается затруднительным. Так как человек лучше и быстрее воспринимает визуальную информацию, для анализа данных и выявления закономерностей в данных применяется их визуализация. Существует множество способов визуализировать многомерные данные. Однако, визуальные модели, построенные этими способами, показывают разную степень эффективности для разных задач. Поэтому необходимо выявить критерии, по которым можно оценить эффективность получаемой визуальной модели. В данной статье описывается процесс получения критериев сравнения метафор визуализации путем исследования визуальных моделей многомерных данных и их метафор визуализации.

Ключевые слова: визуализация многомерных данных, анализ данных, многомерные данные, визуализация, метафора визуализации, система поддержки метафор визуализации.

Research of existing visual models of multidimensional data and their visualization metaphors

A.S. Kurilov
kac.kurilov@yandex.ru

Bryansk state technical university, Bryansk, Russia

Analysis and interpretation of multidimensional data is a very important problem in many spheres of human life, such as production of products, diagnosis of diseases, marketing, information security, etc. However, sometimes the analysis of multidimensional data using traditional approaches is difficult. Since a person perceives visual information better and faster, visualization is used to analyze the data and identify patterns in the data. There are many ways to visualize multidimensional data. However, visual models built in these ways show different degrees of effectiveness for different tasks. Therefore, it is necessary to identify criteria by which to evaluate the effectiveness of the resulting visual model. This article describes the process of obtaining comparison criteria for visualization metaphors by examining visual models of multidimensional data and their visualization metaphors.

Ключевые слова: multivariate data visualization, data analysis, multivariate data, visualization, visualization metaphor, visualization metaphor support system.

1. Введение

В настоящее время человек вынужден изучать и анализировать огромное количество информации. Информация может быть представлена в различных формах. Однако, лучше всего люди воспринимают информацию в виде визуальных объектов. Существует множество способов построить визуальную модель многомерных данных [1]. Для оценки эффективности этих методов необходимо рассмотреть некоторые из них и выделить их достоинства и недостатки.

Цель данной работы — изучить визуальные модели многомерных данных и соответствующие им метафоры визуализации, выделить их достоинства и недостатки, выделить на их основе критерии оценки методов визуализации.

2. Исследование визуальных моделей многомерных данных и соответствующих им метафор визуализации

Для исследования, выполняемого в рамках данной работы, были выбраны следующие визуальные модели: свечной график, линейный график, спиральный график, столбиковая диаграмма, пузырьковая диаграмма, гистограмма.

2.1. Свечной график

Свечной график - вид интервального графика и технический индикатор, применяемый главным образом для отображения изменений биржевых котировок акций, цен на сырьё и т. д. Он основан на представлении диапазонов значений показателей в виде прямоугольников вдоль временной оси и цветовом представлении диапазонов в зависимости от их характеристик (Рис. 1).



Рис. 1. Свечной график

Данный тип графика используется как инструмент биржевой торговли для визуализации и анализа дина-

мики цен на ценные бумаги, деривативы, валюту, акции, облигации, биржевые товары и т.п. за определенный период времени [2, 3].

В роли качественных показателей используются значения времени различных масштабов (годы, месяцы, дни и т.п.).

В роли количественных показателей используются цены в различных валютах.

Основной прямоугольник графического элемента называется телом. Оно отображает диапазон между ценой открытия и ценой закрытия в определенный период времени. Линии, исходящие из верхней и нижней части тела, называются нижней и верхней тенью. Каждая тень представляет собой максимальную или минимальную цену торгов за указанный период времени. Если рынок бычий (цена закрытия выше цены открытия), то тело, как правило, окрашено в белый или зеленый цвет. А если рынок медвежий (цена закрытия ниже цены открытия), то тело обычно имеет черный или красный цвет [4, 6].

Свечной график обладает следующими преимуществами:

- он прост для чтения, так как не перегружен графическими и текстовыми элементами;
- информацию, отображаемую данным графиком, легко воспринимать, так как текстовые и графические элементы хорошо дополняют друг друга.

Свечной график обладает следующими недостатками:

- он обладает недостаточной информативностью, так как свечные графики не отображают события, происходящий в промежутке между ценой открытия и ценой закрытия – только отношения между этими двумя ценами. Поэтому по такому графику нельзя определить волатильность торгов за указанный период времени.

2.2. Линейный график

Линейные графики используются для отображения количественных показателей за непрерывный интервал или определенный период времени. Он основан на представлении значений количественных показателей в виде точек, соединенных отрезками, вдоль прямой качественных показателей (Рис. 2).

Чтобы нарисовать линейный график, необходимо сначала отметить точки данных на декартовой системе координат, а затем провести между этими точками линию. Как правило, на оси Y отмечаются количественные значения, а на оси X либо качественные значения, либо шкала последовательностей [1, 3].

Данный график используется для демонстрации трендов и соотношения показателей (при использовании нескольких линий). Линейные графики очень полезны для получения «общей картины» за определенный промежуток времени и наблюдения за развитием в этот период времени.

Количественные и качественные показатели зависят от визуализируемых данных [4, 6].

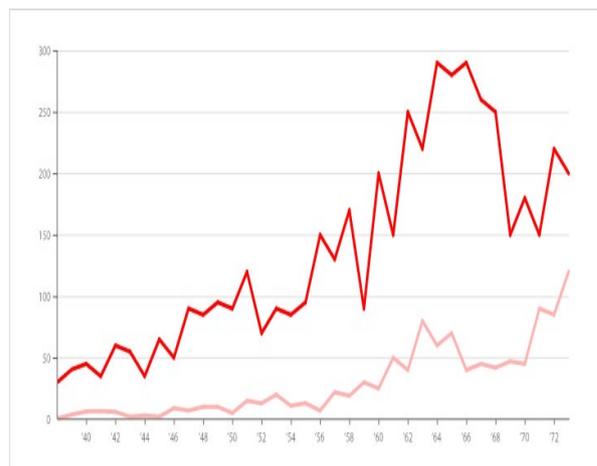


Рис. 2. Линейный график

Линейный график обладает следующими преимуществами:

- он прост для чтения, так как не перегружен графическими и текстовыми элементами;
- информацию, отображаемую данным графиком, легко воспринимать, так как текстовые и графические элементы хорошо дополняют друг друга;
- визуализированные данные легко сравнивать между собой, так как четко видна разница в количественном показателе для различных качественных;
- может быть использован для визуализации данных из различных областей деятельности, так как не ограничен определенными количественными и качественными показателями.

Линейный график обладает следующими недостатками:

- он способен визуализировать данные, показывающие «общую картину» определенных событий.

2.3. Спиральный график

Также известен как спираль динамических рядов. Он основан на представлении значений количественных показателей в виде столбцов, длина которых зависит от соответствующего значения количественного показателя, вдоль временной оси (Рис. 3).

На данном инструменте визуализации данные с временным критерием наносятся вдоль архимедовой спирали. График начинается из центра спирали и развивается вовне. Спиральные графики допускают разнообразие элементов и могут использовать полосы, линии и точки, которые размещаются по спиральной траектории [2,4].

В роли качественного показателя используется временные единицы.

Количественный показатель зависит от визуализируемых данных.



Рис. 3. Спиральный график

Спиральные графики идеально подходят для демонстрации больших наборов данных, чтобы, как правило, выявить общую тенденцию на протяжении длительного временного периода [4, 6].

Спиральный график имеет следующие преимущества:

- данный тип графика прост для чтения, так как не перегружен графическими и текстовыми элементами;
- информацию, отображаемую данным графиком, легко воспринимать, так как текстовые и графические элементы хорошо дополняют друг друга;
- он хорошо подходит для отображения циклических паттернов, в силу своей спиральной формы.

Спиральный график имеет следующие недостатки:

- визуализируемые данные сложно сравнивать из-за формы графика. Хотя для разделения временных периодов и удобства их сравнения можно использовать цвет. Например, анализируя данные за год, можно на графике каждому месяцу года присвоить определенный цвет.

2.4. Столбиковая диаграмма

Также известна как линейчатая или полосчатая диаграмма. Она основана на представлении значений количественных показателей в виде столбцов, длина которых зависит от соответствующего значения количественного показателя, исходящих из точек вертикальной или горизонтальной оси, соответствующих некоторым качественным показателям (Рис. 4).

Классическая столбиковая диаграмма оперирует горизонтальными или вертикальными столбцами для демонстрации дискретных, числовых сравнений между разными категориями. На одной оси диаграммы представлены конкретные сравниваемые категории, а на другой – шкала дискретных значений [1,2].

Столбчатые диаграммы обеспечивают визуальное представление категорических данных. Категорические данные — это данные, сгруппированные в дискретные группы, как например, месяцы года, возрастные группы, размеры обуви, виды животных и т. п.

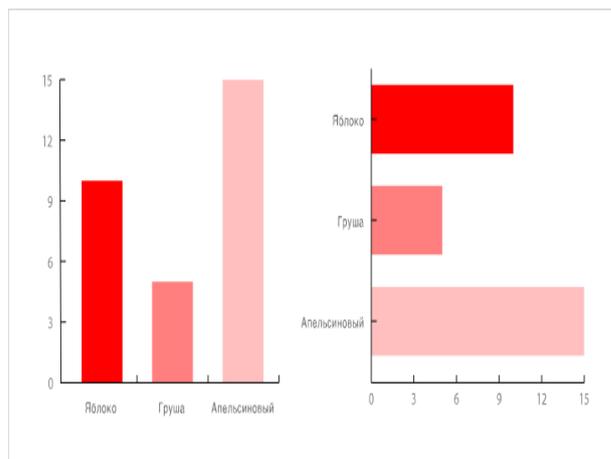


Рис. 4. Столбиковая диаграмма

Достоинства данной диаграммы:

- она проста для чтения, так как не перегружен графическими и текстовыми элементами;
- информацию, отображаемую данной диаграммой, легко воспринимать, так как текстовые и графические элементы хорошо дополняют друг друга;
- визуализированные данные легко сравнивать между собой, так как четко видна разница в количественном показателе для различных качественных.

Недостатки данной диаграммы:

- сложность оформления условных обозначений при большом количестве полос.

2.5. Пузырьковая диаграмма

Пузырьковая диаграмма – это многомерный график, который находится на пересечении диаграммы рассеяния и Диаграмма с пропорциональными областями. Она основана на представлении количественных показателей в виде окружностей разных диаметров и дополнительным цветовым или текстовым обозначением (Рис. 5).

Пузырьковая диаграмма использует декартову систему координат для нанесения на график точек, где оси X и Y – это отдельные переменные. Однако каждая точка имеет собственное название или является собственной категорией (подписанной рядом или в сопроводительном тексте). Соответственно, каждая точка представляет область своей окружности третью переменную. На графике можно использовать цвета для разграничения категорий или для введения дополнительной переменной [1,4]. Время отображается либо в качестве одной из переменных на осях, либо через динамическое изображение изменяющихся во времени данных.



Рис. 5. Пузырьковая диаграмма

Пузырьковые диаграммы, как правило, используются для сравнения и отображения взаимосвязей

между отмеченными/классифицированными окружностями с помощью определения их местоположения и пропорций. Общий вид пузырьковых диаграмм можно использовать для анализа паттернов/корреляций [1].

Достоинства данной диаграммы:

- она проста для чтения, так как не перегружена графическими и текстовыми элементами;
- информацию, отображаемую данной диаграммой, легко воспринимать, так как текстовые и графические элементы хорошо дополняют друг друга.

Недостатки данной диаграммы:

- визуализируемые данные сложно сравнивать между собой;
- слишком большое количество пузырьков может сделать график плохо читаемым, поэтому пузырьковые диаграммы имеют ограничения по объему отображаемых данных. Возможно скорректировать за счет использования интерактивных средств, как-то: нажатие или наведение мыши на пузырьки для отображения скрытой информации.

2.6. Гистограмма

Гистограмма визуализирует распределение данных в рамках непрерывного интервала или ограниченного периода времени. Каждая полоса на гистограмме представляет в табличной форме частотность за определенный интервал/бин. Общая площадь гистограммы равна количеству данных (Рис. 6) [2].

Она основана на представлении значений количественных показателей в виде столбцов, длина которых зависит от соответствующего значения количественного показателя, исходящих из точек вертикальной или горизонтальной оси, соответствующих некоторым количественным показателям.

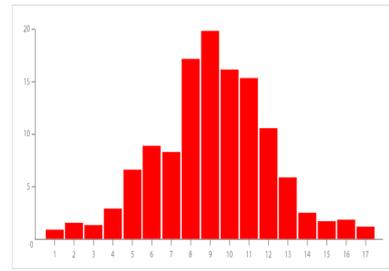


Рис. 6. Гистограмма

Гистограммы помогают определить концентрацию значений, предельные значения и наличие пробелов или отклонений. Кроме того, они удобны для составления приблизительного обзора распределения вероятностей [3].

Достоинства данной диаграммы:

- данный тип диаграммы прост для чтения, так как не перегружен графическими и текстовыми элементами;
- информацию, отображаемую данной диаграммой, легко воспринимать, так как текстовые и графические элементы хорошо дополняют друг друга;
- визуализированные данные легко сравнивать между собой, так как четко видна разница в количественном показателе для различных качественных.

2.7. Результаты исследования визуальных моделей и их метафор визуализации

Изучив достоинства и недостатки визуальных моделей и их метафор визуализации, представленных выше, были выделены следующие критерии, по которым можно сравнивать метафоры визуализации (Таблица 1) [1,4].

Также были составлены следующие рекомендации по использованию визуальных моделей и их метафор визуализации (Таблица 2).

Таблица 1. Критерии сравнения метафор

Критерий	Наихудшее значение	Наилучшее значение
Информативность	недостаточная насыщенность информацией/ перенасыщенность информацией	оптимальность баланса между объемом отображаемой информации и простотой ее восприятия
Насыщенность визуальной модели графическими элементами	недостаточная насыщенность графическими элементами, перенасыщенность графическими элементами	оптимальность баланса между количеством графических элементов, используемых для визуализации информации и простотой ее восприятия
Насыщенность визуальной модели текстовыми элементами	недостаточная насыщенность текстовыми элементами/ перенасыщенность текстовыми элементами.	оптимальность баланса между количеством текстовых элементов, используемых для визуализации информации и простотой ее восприятия
Согласованность графических элементов и текста	полная несогласованность графических и текстовых элементов	полная согласованность графических элементов
Корректность метафоры визуализации для определенной цели	полная некорректность метафоры для определенной цели	корректность метафоры для определенной цели
Простота сравнения данных	визуализированные данные не сравнимы	визуализируемые данные легко сравнить

Таблица 2. Рекомендации по использованию метафор

Метафора	Пути использования
свечной график	визуализация данных, отражающих изменения биржевых котировок акций, цен на сырьё и т. д.
линейный график	визуализация данных, отражающих «общую картину» за определенный промежуток времени и наблюдения за развитием в этот период времени.
спиральный график	визуализация больших наборов данных, чтобы, как правило, выявить общую тенденцию на протяжении длительного временного периода.

Метафора	Пути использования
столбиковая диаграмма	визуализация категорических данных.
пузырьковая диаграмма	визуализация и сравнение количественных данных, принадлежащих различным категориям.
гистограмма	визуализация и сравнение количественных данных, принадлежащих различным категориям

3. Заключение

Таким образом, в ходе работы были рассмотрены следующие визуальные модели и соответствующие им метафоры визуализации: свечной график, линейный график, спиральный график, столбиковая диаграмма, пузырьковая диаграмма, гистограмма.

В ходе исследования были выделены критерии, по которым можно сравнивать метафоры визуализации, а также были составлены следующие рекомендации по использованию метафор.

Дальнейшим развитием данной работы является выделение разработку методики сравнения метафор визуализации.

Библиографический список

- [1] Бондарев А.Е., Галактионов В.А., Шапиро Л.А. Анализ и визуализация многомерных данных в задачах вычислительной газовой динамики // Программирование. - 2015. - №5. - С. 5-12.
- [2] Мамаев, А. Н. Визуализация данных в презентациях, отчетах и исследованиях / А.Н. Мамаев, Д.А. Кудлай. - Л.: Практическая медицина, 2011. - 386 с.
- [3] Шкляр А.В., Захарова А.А. Структурный подход к визуализации данных // Материалы XX Юбилейной Международной конференции по вычислительной механике и современным прикладным системам. - М.: Издательство МАИ-Принт, 2017. - С. 609-611.
- [4] Bondarev A.E., Galaktionov V. A. Multidimensional data analysis and visualization for time-dependent CFD problems // Program. Comput. Softw, Vol. 41, № 5, 2015, pp. 247–252.
- [5] Бондарев А.Е., Бондаренко А.В., Галактионов В.А. Процедуры визуального анализа многомерных данных // Научная визуализация. - 2018. - № 4. - С. 120 - 133.
- [6] Захарова А.А., Шкляр А.В., Ризен Ю.С. Измеряемые характеристики задач визуализации // Научная визуализация. - 2016. - №1. - С. 95-107.
- [7] Манаков Д., Авербух В. Верификация визуализации // Научная визуализация. - 2016. - № 1. - С. 58-94.