

УДК 004.4'414

ВЫБОР МЕТОДОВ ДЛЯ АНАЛИЗА ДАННЫХ В ИНФОРМАЦИОННЫХ СИСТЕМАХ

К. Д. Елизенцев

Сибирский государственный университет науки и технологий имени академика М. Ф. Решетнева
Российская Федерация, 660037, г. Красноярск, просп. им. газ. «Красноярский рабочий», 31
E-mail: bringer34@gmail.com

Рассмотрены основные проблемы выбора методов анализа данных в информационных системах в современных условиях.

Ключевые слова: информационная система, информация, анализ данных.

CHOOSING METHODS FOR ANALYZING DATA IN INFORMATION SYSTEMS

K. D. Elizentsev

Reshetnev Siberian State University of Science and Technology
31, Krasnoyarsky Rabochy Av., Krasnoyarsk, 660037, Russian Federation
E-mail: bringer34@gmail.com

The main problems of choosing methods of data analysis in information systems in modern conditions are considered.

Keywords: information system, information, data analysis.

С развитием сферы информационных технологий, информационные системы стали неотъемлемой частью технической базы большинства организаций. Применение современных информационных систем в организации позволяет оптимизировать процессы ее деятельности. «Машинный труд» доказал свою эффективность на практике благодаря его возможности быстро и точно обрабатывать данные.

На данный момент рынок предлагает множество товаров и услуг. Для того чтобы организация успешно вписывалась в рамки конкуренции, необходимо адаптировать все ее внутренние процессы. Поскольку информационные системы являются частью организации и частью этих процессов, то они также могут подвергаться изменениям.

Основная составляющая информационной системы – это информация. По этой причине существует необходимость в понимании того, как работать с ней, для получения на выходе необходимого результата. Поскольку информационная система – это программный продукт, то и механизм восприятия информации будет иметь специфический вид [2]. Необходимо указать «машине», как и какую информацию обрабатывать. Для решения этой задачи выделяют ряд критериев к предоставляемой информации:

Однородность. Информация должна быть представлена в однородном виде, то есть содержать общую структуру. В первую очередь стоит отметить, что информация может быть представлена в разных типах: например, текстовая, цифровая или смешанная. Обуславливаясь конкретной задачей, необходимо чтобы система получала тот или иной конкретный тип данных.

Ограничение объема данных. Необходимо понять с каким объемом данных приходится работать, так как если информации много, то и количество информационного шума может быть достаточно большим. Большие показатели информационного шума сводят на нет целесообразность работы с данными.

Ценность данных. Необходимо определить ценность данных, то есть понять насколько данные взаимосвязаны между собой. В случае, когда информация не содержит взаимосвязей,

снижается эффективность работы с такой информацией и такая деятельность может не привести к требуемым результатам.

Полезность. Информация должна обладать полезностью. Это значит, что, проанализировав информацию, выводы приведут к конкретному значению и будут применяться на практике.

Существуют следующие современные и актуальные методы анализа данных в информационных системах:

- искусственные нейронные сети (ИНС);
- факторный анализ;
- регрессионный анализ;
- анализ панельных данных;
- корреляционный анализ;
- семантический анализ;
- Data Mining [4].

Прежде чем говорить о выборе метода данных, следует выделить фундаментальную основу любого программного продукта – основная цель или задача, которую будет решать этот продукт. От направления зависит, какую информацию необходимо подавать на вход информационной системы, а от того какой она будет на входе уже зависит сам выбор метода. Таким образом, если вектор имеет, например, экономическую направленность, то на вход будет подаваться экономическая информация (чаще выраженная в числах), а в качестве метода будет использоваться метод, работающий с такой информацией.

Наиболее часто в организациях используются экономические информационные системы, то есть системы, работающие с экономической информацией и на выходе дающие соответствующего рода результаты. Под экономической информацией понимаются сведения о процессах производства, материальных ресурсов, трудовых ресурсов, финансовых процессах и о состоянии объектов управления на определенное время. Она возникает в процессе производственной хозяйственной деятельности хозяйствующих субъектов, а ее назначение состоит в применении для управленческих нужд. Особенности экономической информации определяют необходимость и экономическую целесообразность применения специальных технических средств, в том числе средств вычислительной техники, при её сборе, накоплении, передаче и обработке. Поэтому экономическая информация является как предметом, так и продуктом автоматизированной обработки. Данный вид информации чаще всего выражается в числовом формате, поэтому методам для его обработки необходимо находить причинно-следственные связи в числовых структурах для этих целей могут использоваться методы факторного анализа, регрессионного анализа и анализа панельных данных [3].

Информация, подаваемая на вход информационной системы может быть представлена в смешанном виде или иметь отфильтрованный вид, то есть однородную числовую или текстовую структуру. В случае, когда информация выражена в смешанном виде, в зависимости от целей ее приводят к однородному виду с помощью метода Data mining или вручную, или используют методы способные без участия человека обрабатывать такую информацию – искусственные нейронные сети или искусственный интеллект [1].

Задачи по обработки смешанной информации могут сводиться к «смешиванию» или комбинированию методов, в зависимости от поставленных задач. Таким образом, отфильтровав информацию с помощью метода Data Mining и получив однородную числовую структуру, можно использовать метод факторного анализа.

Возвращаясь к типизации информации стоит отметить такой вид как текстовая информация. Обработывая такого рода информацию, используются методы семантического анализа. Данные методы является сложной математической задачей для «машины» и как правило применяются в процессе создания искусственного интеллекта, однако являются эффективным помощником для информационных систем сфер торговли, маркетинга, журналистики и т. д.

Таким образом, перед тем как выбирать метод для анализа данных в информационных системах, необходимо определиться с целью или задачей, которая будет выполняться с помощью программного продукта. После того как выявлена цель, необходимо обозначить ряд критериев к входной информации, а затем привести ее к необходимому виду. И только после того как

информация имеет необходимый вид, программист переходит к этапу применения метода анализа данных.

Библиографические ссылки

1. Ефимов С. Н., Тынченко В. С. Модели и алгоритмы формирования GRID-систем для структурно-параметрического синтеза нейросетевых моделей // Вестник СибГАУ. 2008. №. 4.
2. Антамошкин О. А., Кукарцев В. В. Модели и методы формирования надежных структур информационных систем обработки информации // Информационные технологии и математическое моделирование в экономике, технике, экологии, образовании, педагогике и торговле. 2014. № 7. С. 51–94.
3. Экономическая информация [Электронный ресурс]. URL: <https://studfiles.net/preview/5679014> (дата обращения: 19.03.2018).
4. Современные методы анализа данных [Электронный ресурс]. URL: <http://riep.ru/upload/iblock/031/031173bb40e099800b248497db44cb88.pdf> (дата обращения: 19.03.2018).

© Елизенцев К. Д., 2018