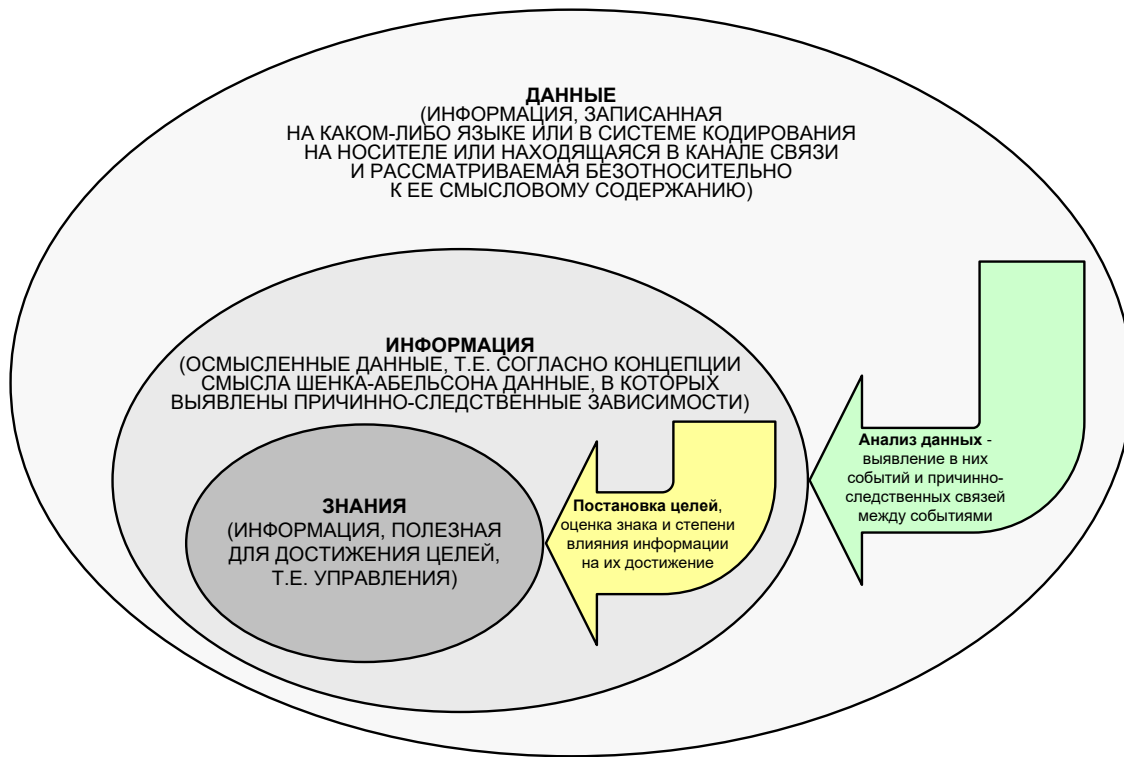


О соотношении смыслового содержания понятий: «Данные», «Информация» и «Знания»

О соотношении содержания понятий: «Данные», «Информация» и «Знания»



Знания – это информация, полезная для достижения целей, т.е. для решения задачи управления и других задач.

Значит для преобразования информации в знания необходимо:

1. Оценить полезность информации для достижения этой цели (знак и силу влияния).
2. Поставить цель (классифицировать будущие состояния моделируемого объекта как целевые и нежелательные).

Первый пункт, по сути, выполнен при преобразовании данных в информацию. Поэтому остается выполнить только второй пункт, т.к. классифицировать будущие состояния объекта управления как желательные (целевые) и нежелательные.

Знания могут быть представлены в различных формах, характеризующихся различной степенью формализации:

- вообще неформализованные знания, т.е. знания в своей собственной форме, ноу-хау (мышление без вербализации есть медитация);
- знания, формализованные в естественном вербальном языке;
- знания, формализованные в виде различных методик, схем, алгоритмов, планов, таблиц и отношений между ними (базы данных);
- знания в форме технологий, организационных, производственных, социально-экономических и политических структур;
- знания, формализованные в виде математических моделей и методов представления знаний в автоматизированных интеллектуальных системах (логическая, фреймовая, сетевая, продукционная, нейросетевая, нечеткая и другие).

Таким образом, для решения сформулированной проблемы необходимо осознанно и целенаправленно **последовательно повышать степень формализации** исходных **данных** до уровня, который позволяет ввести исходные данные в интеллектуальную систему, а затем:

- преобразовать исходные данные в информацию;
- преобразовать информацию в знания;
- использовать знания для решения задач управления, принятия решений и исследования предметной области.

Данные – это информация, записанная на каком-либо носителе или находящаяся в каналах связи и представленная на каком-то языке или в системе кодирования и рассматриваемая безотносительно к ее смысловому содержанию. Это любые изменения значений каких-либо свойств объектов и явлений.

Исходные данные об объекте управления обычно представлены в форме баз данных, чаще всего временных рядов, т.е. данных, привязанных ко времени. В соответствии с методологией и технологией автоматизированного системно-когнитивного анализа (АСК-анализ), развиваемой проф. Е.В.Луценко, для управления и принятия решений использовать непосредственно исходные данные не представляется возможным. Точнее сделать это можно, но результат управления при таком подходе оказывается мало чем отличающимся от случайного. Для реального же решения задачи управления необходимо предварительно преобразовать данные в информацию, а ее в знания о том, какие воздействия на объект управления к каким его изменениям обычно, как показывает опыт, приводят.

Информация есть осмысленные данные.

Смысл данных, в соответствии с концепцией смысла Шенка-Абельсона, состоит в том, что известны причинно-следственные зависимости между событиями, которые описываются этими данными. Таким образом, данные преобразуются в информацию в результате операции, которая называется «Анализ данных», которая состоит из двух этапов:

1. Выявление событий в данных (разработка классификационных и описательных шкал и градаций и преобразование с их использованием исходных данных в обучающую выборку, т.е. в базу событий – эвентологическую базу).
2. Выявление причинно-следственных зависимостей между событиями.

В случае систем управления событиями в данных являются совпадения определенных значений входных факторов и выходных параметров объекта управления, т.е. по сути, случаи перехода объекта управления в определенные будущие состояния под действием определенных сочетаний значений управляющих факторов. Качественные значения входных факторов и выходных параметров естественно формализовать в форме лингвистических переменных. Если же входные факторы и выходные параметры являются числовыми, то их значения измеряются с некоторой погрешностью и фактически представляют собой интервальные числовые значения, которые также могут быть представлены или формализованы в форме лингвистических переменных (типа: «малые», «средние», «большие» значения экономических показателей).

Какие же математические меры могут быть использованы для количественного измерения силы и направления причинно-следственных зависимостей?

Наиболее очевидным ответом на этот вопрос, который обычно первым всем приходит на ум, является: «Корреляция». Однако, в статистике хорошо известно, что это совершенно не так. Для преобразования исходных данных в информацию необходимо не только выявить события в этих данных, но и найти причинно-следственные связи между этими событиями. В АСК-анализе предлагается 7 количественных мер причинно-следственных связей, основной из которых является семантическая мера целесообразности информации по А.Харкевичу.

Когнитивно-целевая структуризация предметной области (единственный неавтоматизированный в системе «Эйдос-Х++» этап АСК-анализа)
На этом этапе разработчик интеллектуального Эйдос-приложения решает, что он рассматривает как объект моделирования и управления, что как факторы, действующие на этот объект, а что как будущие состояния, в которые объект управления может перейти под действием этих факторов

Формализация предметной области (реж.2.3.2.2)

Средства автоматизации кодирования исходных данных - программные интерфейсы (API)

Inp_data, Inp_data.xls
Исходные данные

Class_Sc, Gr_CISc
Классификационные шкалы и градации (реж.2.1)

Opis_Sc, Gr_OpSc
Описательные шкалы и градации (реж.2.2)

Обучающая выборка, эвентологическая база данных (реж.2.3.1)

Obi_zag
Заголовки объектов обучающей выборки

Obi_Kcl
Коды классов объектов обучающей выборки

Obi_Krg
Коды признаков объектов обучающей выборки

Синтез и верификация моделей (реж.3.5)

Способы нормализации моделей 2-го и 3-го уровней (2nd_level, 3rd_level):

1. Дисперсия, сравнение путем вычитания (Var_comp_sub).
2. Дисперсия, сравнение путем деления (Var_comp_div).
3. Среднее, сравнение путем вычит. (Avr_comp_sub)
4. Среднее, сравнение путем делен. (Avr_comp_div)
5. Средн.модуль откл.от средн., сравн.путем вычит. (MAD_comp_sub)
6. Средн.модуль откл.от средн., сравн.путем делен (MAD_comp_div)

Выбор уровня моделей и способа нормализации

Уровни моделей:

- | | |
|---------------------------|------------|
| 1-й уровень (basic_level) | 10 моделей |
| 2-й уровень (2nd_level) | 60 моделей |
| 3-й уровень (3rd_level) | 60 моделей |

Имена моделей формируются из имени: Abs, Prc1, Prc2, Inf1, Inf2, Inf3, Inf4, Inf5, Inf6, Inf7, уровня модели и, если модель не базовая, то способа нормализации

Статистические модели (реж.5.5)

Prc1
Матрица условных и безусловных процентных распределений, рассчитанная по числу признаков классов

Abs
Матрица абсолютных частот (матрица сопряженности, корреляционная матрица)

Prc2
Матрица условных и безусловных процентных распределений, рассчитанная по числу объектов классов

Системно-когнитивные модели (модели знаний) (многопараметрическая типизация) (реж.5.5)

INF1
Количество знаний по А.Харкевичу

INF4
ROI-return on investment

INF6
Разность условной и безусловной вероятностей

INF3
Разности между фактическими и теоретическими частотами

INF2
Количество знаний по А.Харкевичу

INF5
ROI-return on investment

INF7
Разность условной и безусловной вероятностей

Выбор наиболее достоверной модели (реж.3.4)

ВЫХОД,
т.к. достоверной модели нет

Решение задач (реж.4)

Решение задач распознавания системной идентификации и прогнозирования (реж.4.1.2)

Решение задач принятия решений (управления) (реж. 4.4.8, 6.3)

Решение задач исследования предметной области путем исследования ее модели

Если модель адекватна, **достоверна**, т.е. соответствует действительности, то и результаты решения задач в этой модели также соответствуют действительности. Это значит, что если достоверность модели низка или неизвестна, то применять ее для решения реальных задач нельзя. Если же это делается, то является авантюризмом и профанацией науки.

О соотношении задач.

- **распознавание, классификация, идентификация и диагностика** (это одно и то же, т.е. синонимы). При решении этих задач определяется степень сходства/различия образа конкретного объекта с обобщенными образами классов.
- **идентификация и прогнозирование** (при идентификации значения свойств и принадлежность объекта к классу относятся к одному моменту времени, а при прогнозировании значения факторов относятся к прошлому, а переход объекта под действием этих факторов в состояние, соответствующее классу относится к будущему, по сути, **прогнозирование** - это **идентификация будущих состояний**, т.е. это тоже **идентификация, но не в пространстве (настоящем), а в пространстве-времени**;
- **прогнозирование и принятие решений** (при прогнозировании по значениям факторов, действующих на объект моделирования, определяется в какое будущее состояние он перейдет под их действием. При принятии решений, наоборот, по будущему целевому состоянию объекта моделирования определяются значения факторов, которые обуславливают его переход в это будущее целевое состояние. Таким образом **задача принятия решений является обратной по отношению к задаче прогнозирования**);
- **принятие решений путем многократного многовариантного прогнозирования** при различных сочетаниях значений факторов невозможно из-за комбинаторного взрыва. Прогнозирование может быть **элементом** принятия решения, т.е. применено для оценки **адекватности** рассматриваемого уже ранее сформированного другим методом варианта решения, но оно в реальных случаях, т.е. когда много факторов, не может быть применено для выработки самого варианта решения;
- **принятие решений и исследование моделируемой предметной области** (задача принятия решений является обратной по отношению к задаче прогнозирования только в простейшем случае: в случае использования SWOT-анализа. Однако SWOT-анализ имеет свои ограничения: может быть задано только **одно** будущее целевое состояние, некоторые рекомендуемые факторы может **не быть** технологической и финансовой **возможности использовать**. Поэтому в АСК-анализе и системе «Эйдос» реализован развитый алгоритм принятия решений п.6.3 в котором кроме SWOT-анализа используются также результаты решения задачи прогнозирования и результаты кластерно-конструктивного анализа классов и значений факторов, т.е. некоторые результаты решения задачи исследования предметной области.)

Основные публикации автора по вопросам выявления, представления и использования знаний:

Луценко Е.В. Методологические аспекты выявления, представления и использования знаний в АСК-анализе и интеллектуальной системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №06(070). С. 233 – 280. – Шифр Информрегистра: 0421100012/0197, IDA [article ID]: 0701106018. – Режим доступа: <http://ej.kubagro.ru/2011/06/pdf/18.pdf>, 3 у.п.л.