

СОВРЕМЕННЫЕ ИНТЕЛЛЕКТУАЛЬНЫЕ ТЕХНОЛОГИИ АНАЛИЗА ДАННЫХ В РАЗВИТИИ ДИАГНОСТИЧЕСКИХ БАЗ ЗНАНИЙ

Карп В.П., Саруханов Б.А.

Московский государственный институт радиотехники, электроники и автоматики,
(МИРЭА), кафедра Информационных систем.

Россия, 117454, Москва, пр. Вернадского, 78, E-mail: karpvica@mail.ru

В процессе решения задачи классификации сложноорганизованных объектов (практически любой предметной области) возникает проблема учета особенностей их описания. Это связано с тем, что сложные объекты, как правило, характеризуются большим количеством качественных (неизмеряемых) показателей (признаков), или, например, наличием пропущенных данных, порожденных невозможностью непосредственного измерения интересующего показателя. Последнее вынуждает исследователей дополнять пространство основных признаков косвенными, в расчете на то, что при отсутствии части основных сведений можно будет воспользоваться имеющимися косвенными. Однако диагностические правила, построенные на таком смешанном пространстве, могут оказаться статистически и/или содержательно коррелированными между собой, что может исказить и саму базу знаний и, как следствие, исказить результаты диагностики.

Суть данной проблемы состоит в том, что такие косвенные признаки и их симптомы, включены под разными номерами с основными, но по сути несут в себе одну и ту же или очень близкую семантическую информацию, которая практически не учитывается при формальном (компьютерном) построении диагностической базы знаний (ДБЗ).

Целью исследования явилось, с одной стороны, обнаружение неявно зависимых симптомов, которые образуют близкие, практически тождественные по смыслу диагностические правила, а с другой – выявить скрытые множественные связи, обобщенно характеризующие проблемную область исследования.

В данной работе рассматривается алгоритм, обеспечивающий содержательный (семантический) анализ ДБЗ, построенной методом перебора конъюнкций и представленной в виде продукционных правил. Каждое диагностическое правило ДБЗ (за каждый различаемый класс объектов) состоит из двух взаимосвязанных элементов: номеров симптомов, образующих конъюнкцию, и смыслового значения комбинации этих симптомов.

Идея алгоритма состоит в том, чтобы провести поэтапный анализ количества обнаруженных связей симптомов, входящих в ДБЗ, установить ранжирование этих связей и построить графическое отображение обобщенной структуры этих связей.

Практическое применение алгоритма для конкретной медицинской задачи показало его эффективность в решении задачи оптимизации ДБЗ, что позволяет рассматривать разработанный алгоритм как один из способов повышения качества ДБЗ по конкретной проблеме.