

АЛГОРИТМ ОБНАРУЖЕНИЯ И УЧЁТА СКРЫТЫХ СВЯЗЕЙ В ЗАКОНОМЕРНОСТЯХ БАЗЫ ЗНАНИЙ

Карп В.П., Саруханов Б.А., Сафарова Э.А.¹

Московский Государственный Институт Радиотехники Электроники и Автоматики (технический университет), ф-т Кибернетики, каф. Информационных систем, Россия, 119454, г. Москва, пр-т Вернадского д.78, Тел (903)2166547, e-mail: karpvica@mail.ru

¹ Медицинский центр ООО «Меда-21», г. Москва

В процессе решения задач классификации сложноорганизованных объектов возникает необходимость дополнительного анализа построенных диагностических правил, т.е. базы знаний (БЗ). Сложные объекты характеризуются наличием большого количества качественных признаков и наличием пропущенных данных, порожденных невозможностью непосредственного измерения интересующего показателя. Последнее вынуждает исследователей включать косвенные признаки, возможно, связанные между собой, что может значительно исказить результат компьютерной диагностики. Эти особенности требуется учитывать при формировании БЗ, вне зависимости от метода построения правил классификации. Разработанный алгоритм ориентирован на анализ диагностической БЗ, построенной методом перебора конъюнкций (МПК) и представленной в виде набора (упорядоченных по убыванию) диагностически значимых конъюнкций за каждый класс [1]. Каждая конъюнкция состоит из двух взаимосвязанных фрагментов: «номеров симптомов», образующих конъюнкцию; и ее «смыслового значения». Идея алгоритма заключается в дополнительном анализе построенной БЗ в двух обстоятельствах. Первое касается способности МПК допускать в БЗ несколько (не более трех) конъюнкций с одним и тем же «ведущим» симптомом. В этом случае для анализа используется фрагмент БЗ, состоящий из номеров симптомов, и применяется так называемое условие К-ограничения, по которому из набора конъюнкций с одинаковым ведущим симптомом для каждого конкретного объекта учитывается одна наиболее диагностически значимая. Вторая ситуация порождается наличием скрытых (косвенных/неочевидных) содержательных зависимостей отобранных в БЗ [2]. Для их обнаружения алгоритм подсчитывает коэффициенты парных корреляций между всеми конъюнкциями. Конъюнкции, образовавшие пары с «высокими» коэффициентами корреляций, подвергаются дополнительному анализу их смыслового содержания и семантическому преобразованию БЗ. Такой подход позволяет рассматривать данный алгоритм одним из способов повышения качества базы знаний.

Литература.

1. Карп. В. П. Интеллектуальный анализ данных в проблеме построения решающих правил классификации (на примере задач медицинской диагностики) // Новости искусственного интеллекта. – 2006. – №2. – с.57-75.
2. Карп В.П., Саруханов Б.А. Формирование проблемно-ориентированной базы данных в условиях пассивного эксперимента // Сборник трудов научно-технической конференции. Ч.1. «Информационные технологии и системы. Вычислительная техника». МИРЭА. – М., – 2007. – с. 53-57.