

КУЗНЕЦОВ С. О., ФИНН В. К.

ОБ ОДНОЙ МОДЕЛИ ОБУЧЕНИЯ И КЛАССИФИКАЦИИ, ОСНОВАННОЙ НА ОПЕРАЦИИ СХОДСТВА

§ 1. Введение

В этой статье мы дадим обзор некоторой системы поиска закономерностей, получившей в честь Джона Стюарта Милля название ДСМ-метода автоматического порождения гипотез (или ДСМ-АПГ) и впервые рассмотренной в работах [13], [14]. ДСМ-метод может быть изложен по-разному: как логическая теория, имитирующая правдоподобный вывод, или же как программная система, используемая для прогнозирования свойств объектов различной природы и анализа причин этих свойств. В данной статье мы представим ДСМ-метод как систему автоматического обучения и классификации, значительное внимание уделяя вычислительной сложности алгоритмов и задач, возникающих в модели. Читателя, интересующегося логической проблематикой ДСМ-метода, мы отсылаем, прежде всего, к статьям [16]–[18], в которых можно найти исчерпывающие ссылки. В обзорных работах [10], [18] читатель может найти ссылки на работы по приложению ДСМ-метода в задачах анализа биологической активности химических соединений, канцерогенеза, технической диагностики и др.

В отличие от предыдущих публикаций по ДСМ-методу, в данной работе терминология приближена к той, которая используется в теории формального анализа понятий (formal concept analysis) [32], [33], развиваемой в Германии и Франции. При этом, дав представление о модели для общего типа данных, мы остановимся на более подробном рассмотрении случая, когда данные представляются множествами некоторых «дескрипторов», каждый из которых соответствует определенному структурному элементу описываемых объектов.

Элементарная схема действий системы, основанной на ДСМ-методе, сходна со схемой вывода в типичных системах автоматического обучения (см., например, [24]). Гипотезы о закономерностях получают в результате таких обобщений положительных примеров, которые не являются обобщением отрицательных примеров. Специфическими для ДСМ-АПГ чертами являются способ выражения операции обоб-

щения, использование «усиленных» гипотез, соответствующих проверкам на гипотезах выполнения некоторых «эмпирических закономерностей» — условий здравого смысла, использование так называемого обобщенного метода, возможность получения сложных зависимостей на данных с помощью комбинации индуктивного и дедуктивного (логического) выводов. Можно выделить три основных уровня ДСМ-метода, которые можно назвать уровнями сходства, правил и рассуждения, соответственно.

На первом уровне, или *уровне сходства*, задается структура данных об объектах из предметной области, а также операция сходства, сопоставляющая двум объектам из предметной области третий объект, выражающий сходство первых двух.

На втором уровне, *уровне правил*, объекты из предметной области делятся на положительные примеры (объекты, вызывающие некоторый интересующий нас эффект W) и отрицательные примеры (объекты, не вызывающие эффекта W). Правила «находят» сходства положительных примеров и, быть может, проверяют некоторые другие условия, позволяющие называть найденные сходства гипотезами о структурных причинах эффекта W .

На третьем уровне, *уровне рассуждений*, составляются стратегии, т. е. последовательности применения правил правдоподобного вывода и проверок некоторых условий на множестве всех исходных данных и полученных гипотез, позволяющих делать вывод об обоснованности правдоподобного ДСМ-вывода. На уровне рассуждений получение некоторых гипотез возможно за счет дедуктивного вывода, использующего результаты правдоподобного вывода. Тем самым на этом уровне осуществляется синтез правдоподобного и чисто дедуктивного (достоверного) вывода, что сближает ДСМ-метод с системами обучения, основанными на объяснении [23], [28].

Поясним структуру статьи. В § 2 даются определения основных понятий ДСМ–АПГ: операции сходства, правил порождения гипотез и классификаций, обобщенного ДСМ-метода и рассуждений в ДСМ-методе, а также обсуждается концепт устойчивости гипотез. В § 3 обсуждаются вопросы сложности алгоритмов и задач ДСМ-метода.