

## ИСПОЛЬЗОВАНИЕ ГЕНЕРАТОРА ИНФОРМАЦИОННЫХ СИСТЕМ В ЗАДАЧЕ СОКРАЩЕНИЯ АТТРИБУТОВ

*Глухих Олег Юрьевич*

*Магистрант*

*Институт ФО УрФУ имени первого Президента России Б. Н. Ельцина,  
Екатеринбург, Россия*

*E-mail: oleg.glukhikh@urfu.me*

*Научный руководитель — Васькин Павел Ильич*

В современном мире распространены информационные системы, выполняющие прикладные задачи: системы бухгалтерского учёта, маркетинговые, обработки данных и другие. С ростом количества данных в мире возрастает нагрузка на информационные системы, их эффективная работа начинает требовать больше денег и ресурсов. Решением данной проблемы является сокращение количества параметров, принимаемых в обработку, выявление значимых атрибутов.

Задача нахождения всех сокращений (редуктов), а также нахождения минимального сокращения для заданной таблицы данных, оказалась NP-трудной [1]. Исходя из этого, в последние годы разрабатываются новые эвристические методы сокращения. Основываясь на подходе, предложенном Ю. И. Журавлевым [2] для нахождения существенных переменных булевых функций, предлагается использовать таблицы решений для обозначения информационных систем через  $S = (U, C, D)$ , где  $U$  представляет собой конечное непустое множество объектов, также называемое универсумом,  $C$  является атрибутами условия, а  $D$  — атрибутами решения. При данном подходе задача сокращения атрибутов информационной системы сводится к решению задачи покрытия.

Для сравнения работы различных эвристик часто осуществляют проверку основанных на них алгоритмов на одинаковом множестве задач. Использование генератора задач в этом деле может упростить процесс подбора подходящих множеств и предоставить возможность статистически протестировать алгоритмы. В качестве псевдослучайного генератора информационных систем предлагается эталонная таблица решений — информационная система специального типа  $S_p = (U_p, C, D)$ .

**Определение 1.** *Эталонная информационная система (ЭИС) — это информационная система, при обращении к которой можно получить множество таблиц решения, то есть при подаче на*

*вход ЭИС набора значений атрибутов условия она выводит значения атрибутов решения. Таким образом, ЭИС заменяет экспертов и экспериментаторов, сопоставляющих значения атрибутов условия значениям атрибутов решения.*

Следует обозначить состав метаинформации, подаваемой на вход генератору. Она позволяет генератору эталонных информационных систем формировать таблицы решений с указанными свойствами. Это  $m$  — количество существенных входных переменных;  $l$  — количество зависящих от них атрибутов решения (функций); вероятности того, что функции зависят от такого-то числа переменных. Важной характеристикой информационной системы также является показатель степени пересечения множеств существенных переменных функций. В качестве такого показателя будем использовать  $\lambda$ . Показатель  $\lambda$  принимает значения в диапазоне  $[0,1]$ . Если  $\lambda$  равно 0, то функции не имеют пересечений по множествам существенных переменных. Если же  $\lambda$  равно 1, то все функции зависят от всего множества переменных.

Полученные генератором ИС могут использоваться как задачи для тестирования различных алгоритмов сокращения, а их генерируемый характер позволяет производить тестирование на огромном количестве объектов, не ограниченном типовыми примерами. Как следствие, это позволяет проверить различные алгоритмы в равных, статистически значимых условиях.

В заключение автор выражает признательность доценту кафедры интеллектуальных информационных технологий института фундаментального образования УрФУ к.т.н. П. И. Васькину за помощь в проведении исследований и подбор литературы.

### **Литература**

1. Thuy N. Wongthanavasus S. A new approach for reduction of attributes based on stripped quotient sets. Pattern Recognition. 2019.
2. Журавлев Ю. И. Дискретный анализ. Формальные системы и алгоритмы // Ю. И. Журавлев, Ю. А. Флеров, М. Н. Вялый. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2019.