

УДК 519.226.3

**АЛГОРИТМ ОБУЧЕНИЯ БОЛЬШИХ БС НА ОСНОВЕ ОГРАНИЧЕНИЙ
ПРОСТРАНСТВА ПОИСКА С ПОМОЩЬЮ КОЭФФИЦИЕНТА БРАВЕ**

Каминский Ю.К. (Университет ИТМО)

Научный руководитель – Деева И.Ю.

(Университет ИТМО)

В данном исследовании проведён комплексный аналитический обзор существующих методов обучения больших байесовских сетей. Доклад направлен на исследование возможностей ограничения пространства поиска возможных решений с помощью коэффициента Браве и сравнение алгоритма на основе коэффициента Браве с существующими решениями в области больших байесовских сетей.

Существует два типа больших байесовских сетей: байесовские сети с большим количеством строк в наборе данных (миллионы строк) и большим количеством узлов (десятки, сотни узлов). Данное исследование сфокусировано на байесовских сетях с большим количеством узлов.

Байесовские сети это мощный и универсальный инструмент для моделирования многомерных распределений и генерации синтетических данных, однако структурное обучение больших байесовских сетей является сложной вычислительной задачей, а с ростом количества узлов байесовской сети, количество сетей-кандидатов и границы поиска растут сверхэкспоненциально. Таким образом, существует серьезная необходимость в разработке более быстрых алгоритмов структурного обучения байесовских сетей.

Коэффициент Браве даёт возможность оценить взаимную встречаемость двух переменных в группах, сформированных с помощью методов корреляции или взаимной информации. Следовательно, одной из важнейших задач являлась оценка наилучшего количества ближайших соседей каждой переменной набора данных для формирования таких групп. Кроме того, необходимо было оценить наиболее эффективное пороговое значение коэффициента Браве для включения ребра – кандидата в список возможных ребер, при котором качество структурного обучения будет удовлетворительным, а время на обучение останется приемлемым.

В данном исследовании структурное обучение байесовских сетей проведено более, чем на десяти различных наборах данных, предназначенных специально для тестирования алгоритмов структурного обучения. Эксперименты, проведенные в ходе работы над алгоритмом на основе коэффициента Браве показали его эффективность и превосходство над некоторыми аналогами не только по времени на структурное обучение, но по качеству выстроенной сети. Качество сети определялось с помощью специальной метрики – структурное расстояние Хэмминга.

Алгоритм на основе коэффициента Браве имеет большое количество перспектив для дальнейшего развития и улучшения. Эффективность такого подхода подтверждена экспериментально на разных наборах данных с различными характеристиками.

Каминский Ю.К. (автор)

Деева И.Ю. (научный руководитель)