

**РАЗРАБОТКА И ИССЛЕДОВАНИЕ АЛГОРИТМОВ ОБУЧЕНИЯ БОЛЬШИХ  
БАЙЕСОВСКИХ СЕТЕЙ**

**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 обучения больших Байесовских сетей с помощью локальных структур и эволюционных алгоритмов.

**Введение.** 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.

**Основная часть.** Разделение большого набора данных на локальные структуры даёт возможность значительно ускорить обучение Байесовской сети. А эволюционные алгоритмы позволяют соединить локальные структуры между собой в большую сеть. Одной из важнейших задач являлся поиск оптимального способа соединения локальных структур. Наиболее успешным оказалось соединение структур через “скрытые” переменные. “Скрытые” переменные являются результатом кластеризации методом  $k$ -средних внутри каждой локальной структуры. Кроме того, необходимо подобрать такой алгоритм разделения на локальные структуры, чтобы ни одна из локальных структур не включала в себя слишком большое количество переменных исходного набора данных. 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 показали его эффективность и превосходство над некоторыми аналогами не только по времени на структурное обучение, но по качеству выстроенной сети.

**Выводы.** В ходе исследования были получены границы применимости алгоритма. Алгоритм показывает наилучшие результаты на особенно больших наборах данных (от 100 узлов), однако на наборах данных среднего размера (50-100 узлов), алгоритмы на основе ограничения пространства поиска часто работают эффективнее и быстрее. Качество сети определялось с помощью специальной метрики – структурное расстояние Хэмминга. Алгоритм на основе локальных структур и эволюционных вычислений имеет большое количество перспектив для дальнейшего развития и улучшения. Эффективность такого подхода подтверждена экспериментально на разных наборах данных с различными характеристиками.

#### Список использованных источников

1. Suter P. et al. Bayesian structure learning and sampling of Bayesian networks with the R package BiDAG //arXiv preprint arXiv:2105.00488. – 2021.
2. Aragam B., Gu J., Zhou Q. Learning large-scale Bayesian networks with the sparsebn package //arXiv preprint arXiv:1703.04025. – 2017.
3. Xu R. et al. PEWOBS: An efficient Bayesian network learning approach based on permutation and extensible ordering-based search //Future Generation Computer Systems. – 2022. – Т. 128. – С. 505-520

Каминский Ю.К. (автор)

Деева И.Ю. (научный руководитель)