

УДК 519.226.3

**СРАВНИТЕЛЬНЫЙ АНАЛИЗ ПОДХОДОВ К ПАРАМЕТРИЧЕСКОМУ ОБУЧЕНИЮ
БАЙЕСОВСКИХ СЕТЕЙ НА ДАННЫХ СМЕШАННОГО ТИПА**

Деева И.Ю. (Университет ИТМО)

Научный руководитель – к.т.н., Калюжная А.В.

(Университет ИТМО)

В данном исследовании проведён комплексный аналитический обзор существующих методов обучения байесовской сети, а также предложен гибкий метод параметрического обучения, подходящий для представления данных смешанного типа. Сравнительный анализ представлен на задаче моделирования социальных данных с последующим семплированием из байесовской сети. Проведённый анализ показывает преимущества гибкого метода обучения на смешанных данных с точки зрения точности моделирования многомерных распределений.

Байесовские сети являются мощным и универсальным инструментом для моделирования многомерных распределений. Однако структурное и параметрическое обучение байесовских сетей на данных является вычислительно сложной задачей. Так как реальные данные зачастую обладают смешанным типом (дискретным и непрерывным), то обучение байесовских сетей усложняется. Большинство существующих решений сводят задачу к работе с данными одного типа, это достигается либо дискретизацией непрерывных переменных, либо работой только с непрерывными данными без учёта дискретных. Оба подхода приводят к значительным потерям информации и как следствие, неточности моделирования. Именно поэтому важна разработка эффективных методов обучения, способных работать со смешанными данными.

Задача байесовского вывода позволяет решать множество прикладных проблем, связанных с моделированием многомерного распределения. В данном исследовании существующие подходы к обучению байесовских сетей рассмотрены на задаче моделирования социальных данных с последующим семплированием из байесовской сети. Качество семплирования – это показатель качества обучения байесовской сети. Предложенный метод обучения параметров байесовской сети на смешанных данных основан на использовании трёх типов распределений:

- 1) Таблицы условных вероятностей (СРТ), если значения в узле дискретны, а его родительские элементы имеют дискретное распределение;
- 2) Гауссово распределение, если значения в узле непрерывны, а его родители имеют непрерывное распределение;
- 3) Условное гауссовское распределение, если значения узла непрерывны, а его родители имеют дискретное и непрерывное распределение.

Проведённые эксперименты показывают эффективность предложенного подхода к обучению параметров байесовской сети с точки зрения точности моделирования многомерного распределения.

Предложенный подход имеет множество перспектив развития прежде всего в сторону усложнения моделирования непрерывных распределений. Полученная модель может быть использована как для моделирования многомерных социальных данных, так и для последующей работы с моделью (генерация синтетических данных, моделирование данных социальных групп).