

Новый метод построения гибридных байесовских сетей на основе на взаимной информации и тестов на условную независимость

Балабин Никита Максимович

Студент (специалист)

Московский государственный университет имени М.В.Ломоносова,
Механико-математический факультет, Кафедра теории вероятностей, Москва, Россия
E-mail: nikita@mxl.ru

В последнее десятилетие в прикладных исследованиях распространено применение различных графических моделей для выявления зависимостей между переменными модели [1,6]. Одной из самых популярных являются байесовские сети (БС). БС выявляют условные и безусловные вероятностные зависимости между переменными, включенными в модель, что помогает отобрать наиболее информативные из них. Применение БС зависит от распределения этих переменных [2]. Для применения БС, например, в медицинских исследованиях, возникает необходимость в построении гибридных БС, то есть БС, построенных по смешанным данным, которые содержат как дискретные переменные, так и непрерывные переменные, распределенные по нормальному закону. На данный момент самым распространенным методом построения гибридной БС является Байесовский информационный критерий (BIC), введенный в [5]. В [3] был предложен новый способ построения БС, называющийся взаимным информационным тестом (MIT, mutual information test). Однако этот способ был рассмотрен только для дискретных переменных.

Цель работы: Обобщить взаимный информационный тест для моделей, как с непрерывными, так и дискретными переменными. Программно реализовать полученный алгоритм и провести ряд сравнений со стандартными подходами построения БС.

Основным результатом работы можно считать получение нового метода построения гибридной БС, полученный с помощью одновременного использования статистических тестов на условную независимость и метода максимизации функции правдоподобия, в то время как старые алгоритмы использовали только один из этих подходов [4]. Также в работе продемонстрирован сравнительный анализ с подходами из [4], подтверждающий качество данного алгоритма.

Источники и литература

- 1) Димитрова Л. К., Голубева О. А. Применение байесовской сети в дифференциальной диагностике артериальной гипертензии [Текст] // Технические науки в России и за рубежом: материалы III Междунар. науч. конф. (г. Москва, июль 2014 г.). — М.: БукиВеди, 2014. — С. 4-14.
- 2) Barber, D. Bayesian Reasoning and Machine Learning. — Cambridge University Press. 2012.
- 3) De Campos, L.M. (2006) A Scoring Function for Learning Bayesian Networks Based on Mutual Information and Conditional Independence Tests. Journal of Machine Learning Research, 7, 2149-2187.
- 4) Nagarajan R, Scutari M, Lebre S (2013). Bayesian Networks in R with Applications in Systems Biology. — Springer. 2013
- 5) Schwarz, Gideon E. (1978). "Estimating the dimension of a model". Annals of Statistics 6 (2): 461–464

- 6) Sucar, Luis Enrique; Gillies, Duncan F. Expressing Relational and Temporal Knowledge in Visual Probabilistic Networks — eprint arXiv:1303.5432. 03/2013