

Статистический анализ данных ВИЧ-инфицированных пациентов с применением Байесовских сетей

Научный руководитель – Яровая Елена Борисовна

Орловский А.А.¹, Балабин Н.М.²

1 - Московский государственный университет имени М.В.Ломоносова, Москва, Россия; 2 - Московский государственный университет имени М.В.Ломоносова, Механико-математический факультет, Кафедра теории вероятностей, Москва, Россия

В последнее десятилетие в прикладных исследованиях распространено применение различных графических моделей [1][2][3]. Одной из самых популярных являются Байесовские сети (БС). БС находит условные и безусловные зависимости, что помогает отобрать наиболее информативные факторы из включенных в модель. Применение БС зависит от распределения этих факторов [4]. В работе с помощью БС проанализированы данные пациентов с ВИЧ-инфекцией. Особенностью является то, что в качестве изучаемых показателей берутся вирусы-возбудители вторичных заболеваний, которые измеряются либо количественно (число копий ДНК вируса на мл биоматериала), либо качественно (обнаружен ли вирус в биоматериале). Сложность работы с количественными показателями состоит в том, что в случае их обнаружения количество копий ДНК может варьироваться от 10^2 копий до 10^7 копий в мл биоматериала и не поддается параметризации с помощью непрерывных распределений.

Цель работы: формализовать данные для применения БС и с их помощью выявить и изучить зависимости между исследуемыми показателями у ВИЧ-инфицированных пациентов. Затем путем фиксации клинически обоснованных связей скорректировать полученные сети для отображения наиболее правдоподобного, с медицинской точки зрения, взаимодействия вирусов-возбудителей и болезней с сохранением качества модели.

В данном исследовании алгоритм построения Байесовской сети будет основан на совместности исследуемых показателей с вероятностным распределением получающейся модели. Существует несколько оценок того, насколько хорошо сеть согласована с данными [5]. Продемонстрировано преимущество применения Байесовского информационного критерия для анализа полученных данных. **Основным результатом** можно считать способ формализации исходных данных для применения БС и построение моделей взаимосвязей вторичных заболеваний и вирусов-возбудителей в обследованных биоматериалах.

Данные, использованные в работе, получены на базе ИКБ №2 г. Москвы. Работа проведена совместно со старшим научным сотрудником федерального научно-методического центра по профилактике и борьбе со СПИДом, кандидатом медицинских наук Василием Иосифовичем Шахгельдяном.

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

- 3 Korsunsky, Илья; Ramazzotti, Daniele; Caravagna, Giulio; Mishra, Bud. Inference of Cancer Progression Models with Biological Noise — eprint arXiv:1408.6032. 08/2014
- 4 Barber, D. Bayesian Reasoning and Machine Learning. — Cambridge University Press. 2012.
- 5 Nagarajan R, Scutari M, Lebre S (2013). Bayesian Networks in R with Applications in Systems Biology. — Springer. 2013