

ITMO

Структурное обучение КОМПОЗИТНЫХ ВЕРОЯТНОСТНЫХ МОДЕЛЕЙ

Шахкян Каринэ Артуровна,

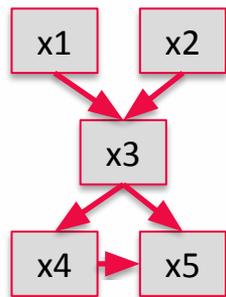
аспирант ИТМО, инженер ИЦ "Сильный ИИ в промышленности",

научный руководитель:

Калюжная Анна Владимировна, к.т.н., доцент, старший научный

сотрудник

Байесовская сеть – направленный ациклический граф, в узлах которого располагаются распределения признаков, а рёбра обозначают условную независимость между признаками.



$$x_1 \perp x_4 | x_3$$

$$p(X) = \prod_i p(x_i | x_1, \dots, x_{i-1})$$

$$p(x_1, \dots, x_5) = p(x_1)p(x_2|x_1) \times \dots \\ \dots \times p(x_3|x_1, x_2)p(x_4|x_3) p(x_5|x_3, x_4)$$

Обучение байесовских сетей

Обучение структуры –
нахождение
структуры БС из
данных

Обучение параметров –
нахождение параметров
распределений в узлах БС
после нахождения структуры

$$\underbrace{P(\mathcal{M} | \mathcal{D})}_{\text{learning}} = \underbrace{P(\mathcal{G} | \mathcal{D})}_{\text{structure learning}} \cdot \underbrace{P(\Theta | \mathcal{G}, \mathcal{D})}_{\text{parameter learning}}$$

- **Constraint-based algorithms:** они используют статистические тесты для изучения отношений условной независимости (называемых в данном случае «ограничениями») из данных и предполагают, что DAG - идеальная карта для определения правильной сетевой структуры.
- **Score-based algorithms:** каждому кандидату группы DAG присваивается оценка, отражающая степень соответствия, которая затем принимается в качестве целевой функции для максимизации.
- **Hybrid algorithms:** тесты условной независимости используются для изучения, по крайней мере, части отношений условной независимости из данных, тем самым ограничивая пространство поиска для последующего поиска на основе оценок. Последний определяет, какие ребра фактически присутствуют в графе и их направление.

Обучение классических БС

Задача структурного обучения байесовской сети формулируется как задача оптимизации:

$$E_{opt} = \arg \max_{G' \in G_{possible}} F(G'),$$

где E_{opt} — ребра найденной оптимальной структуры графа, $G_{possible}$ — множество возможных структур графа байесовской сети, F — функция оценки.

$$F(G) = \sum_{i=1}^N L(v_i | \theta_i),$$

где L — правдоподобие, N — количество узлов в графе, v_i — узел графа, θ_i — параметры распределения в узле v_i .

Композитная БС

В классической постановке модели для условных распределений задаются **сверху и одинаково** для всех узлов.

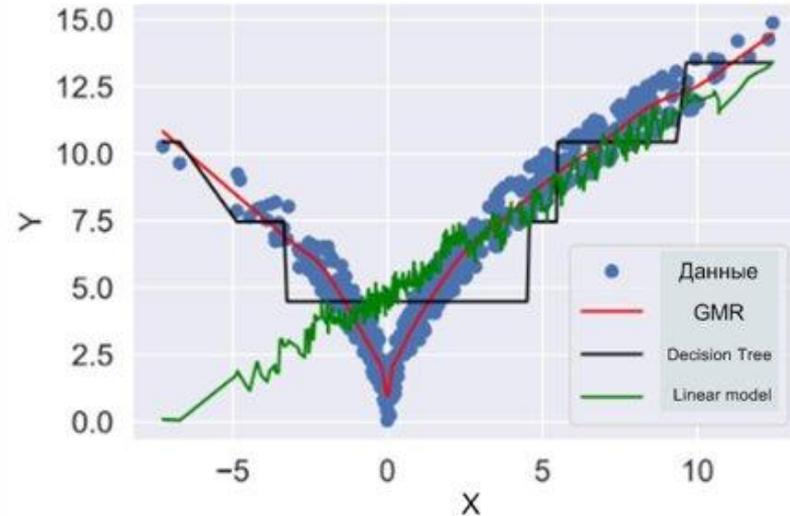
$$V_{opt}, E_{opt} = \operatorname{argmax}_{G' \in G_{possible}} F(G')$$

При **композитном подходе** модели МО подбираются **прямо во время обучения байесовских сетей** и являются частью модели байесовской сети.

В этом случаи узлы БС делятся на два типа:

- узлы-переменные
- узлы-модели

$$V_{opt}, E_{opt}, M_{opt} = \operatorname{argmax}_{G' \in G_{possible}} F(G')$$



BAMT, GOLEM — Scientific Open Source

BAMT — open-source фреймворк для вероятностного моделирования на основе байесовских сетей.

- Обучение структуры и параметров узлов.
- Поддержка разных типов данных.
- Генерация синтетических данных и поиск зависимостей.
- Модульность и интеграция с ML-моделями.
- Легковесный API и гибкая конфигурация.
- Документация [readthedocs](https://github.com/aimclub/BAMT).

GOLEM — open-source фреймворк для оптимизации графовых структур. В основе оптимизации лежит генетический алгоритм.

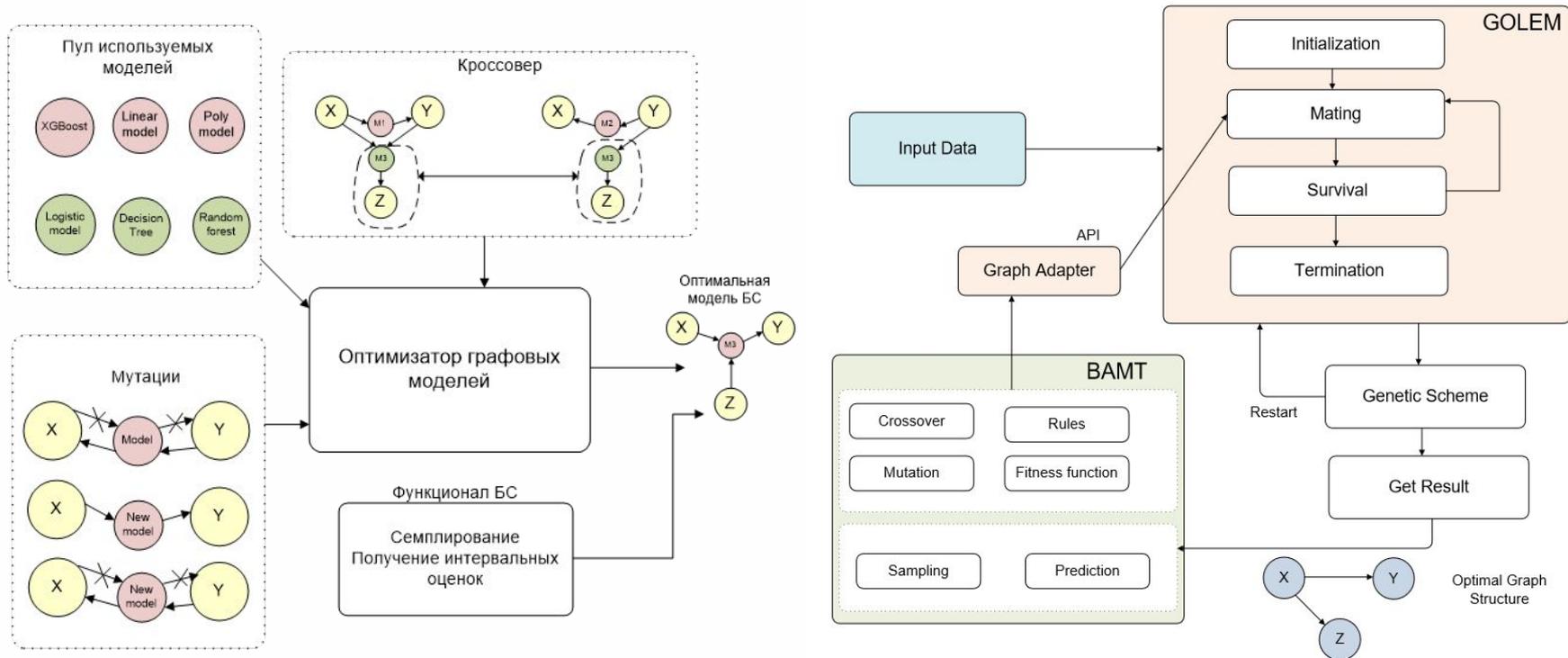


<https://github.com/aimclub/BAMT>



<https://github.com/aimclub/GOLEM>

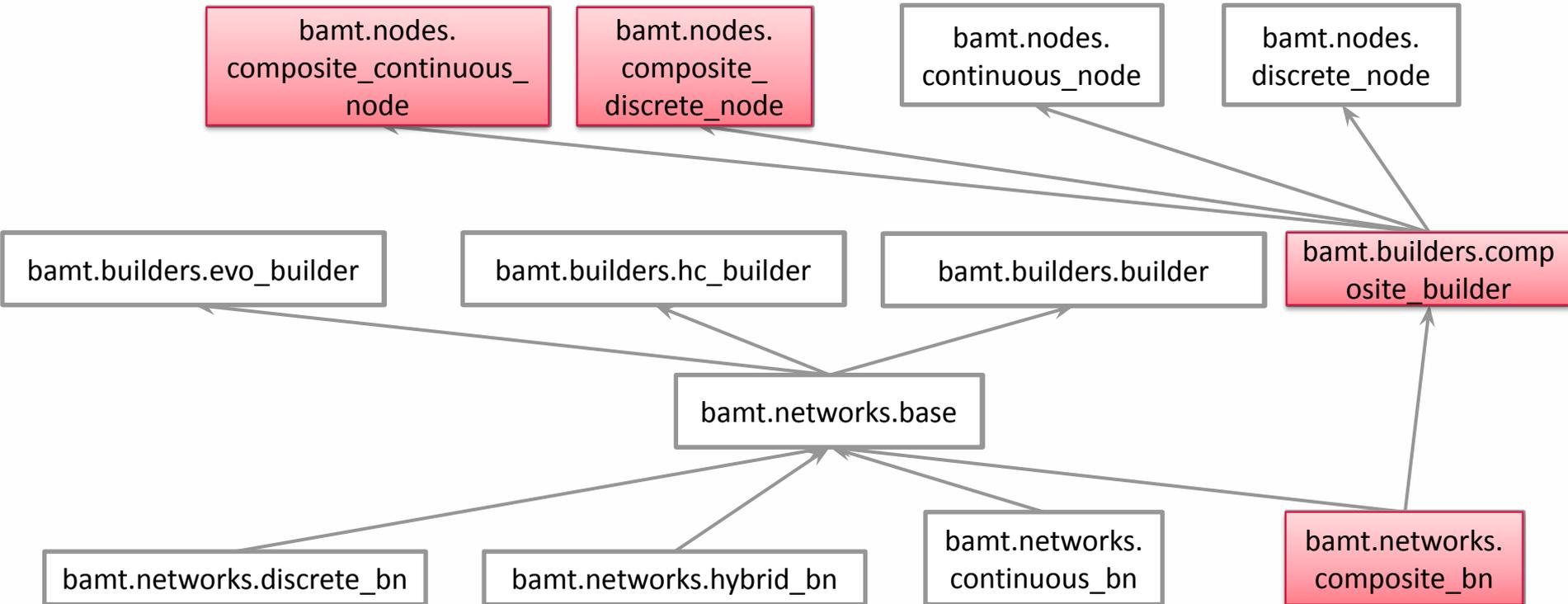
Реализация обучения композитных БС на основе GOLEM и BAMT



Другие библиотеки

BAMT	Python	✓	✓	✓	✓
bnlearn	R	✓	✓	✗	✓
BiDAG	R	✗	✓	✗	✓
pomegranate	Python	✗	✗	✗	✗
DEAL	R	✓	✓	✗	✗
pgmpy	Python	✓	✗	✗	✓

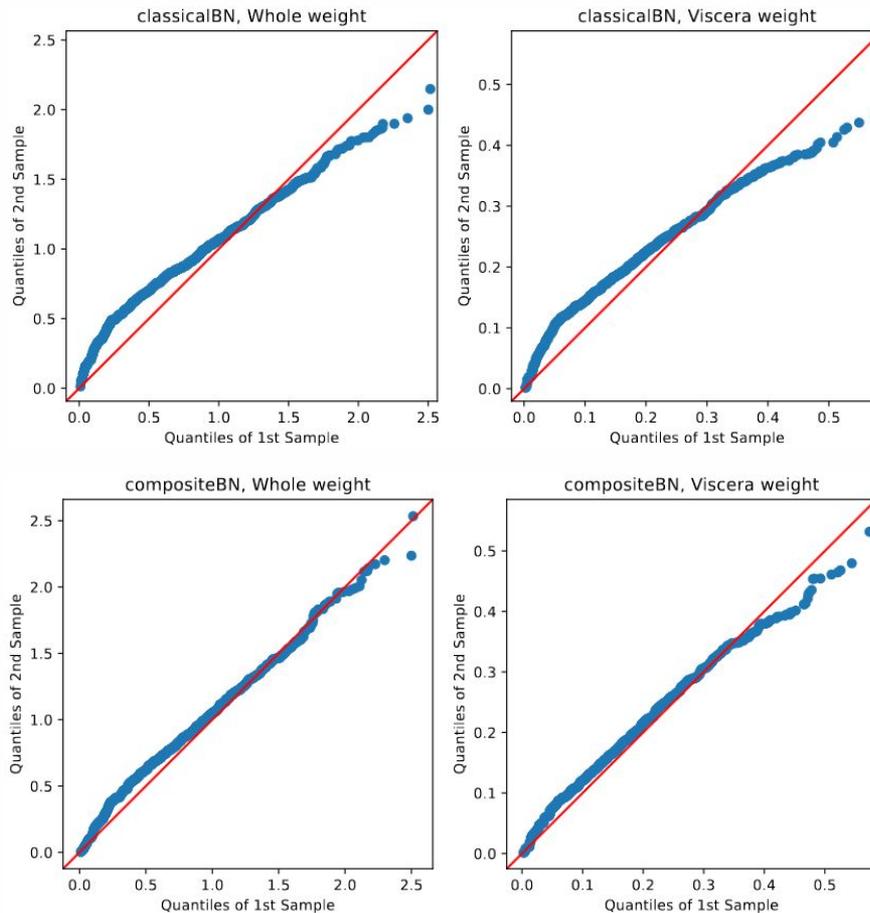
Диаграмма классов



Композитные БС. Likelihood

Datasets	PyBNesian	MoTBFs	CBN
Iris	-32,79	-54,94	-29,3
Yeast	1903,77	984,86	1388,3
QSAR fish toxicity	-399,72	-468,53	-574
Liver	-756,17	-764,35	-756,1
Glass	6,28	83,52	60,9
Ecoli	195,5	136,48	662,56
Balance	-386,98	-	-441,7
Breast_cancer	-922,14	-721,57	-1203,3
Parkinsons	906,74	-	1036,1
Vowel	-467,09	-1139,65	-697,1
QSAR Aquatic	-1072,87	-	-958,2
Wine	-357,55	-	-320,9
Block	-146578,02	-	-21746,9
Breast_tissue	-612,02	-	-393,8
CPU	-1035,04	-	-949,8
Sonar	1879,03	-	1563,2
Vehicle	-4261,32	-	-4219,4
Wdbc	1299,51	-	1093,9
Wdbc	162,61	-	77,4

Композитные БС. QQ-biplots



- Байесовские сети – удобный и многофункциональный инструмент построения многомерных распределений.
- Композитный взгляд на БС – новый шаг на пути развития теории БС.
- Композитными БС можно воспользоваться с помощью фреймворка ВАМТ.

Спасибо за внимание!

iTMO *re than a*
UNIVERSITY

Постановка экспериментов для композитных БС

PyBNesian:

структурное обучение происходит с помощью модифицированного Hill Climbing. Параметрическое обучение с помощью гауссиан и с помощью KDE оценки. Оптимизация структуры и параметров происходит одновременно.

MoTBFs:

Обучение байесовских сетей на основе смесей усеченных базисных функций (MoTBF), которые включают смеси полиномов (MOP) и смеси усеченных экспонент (MTE). Структура байесовской сети получена от PyBNesian.

CBN:

Оптимизация структуры и моделей машинного обучения композитной байесовской сети с помощью фреймворка BANT.