

Применение активного обучения к графовым моделям на примере оценки рисков распространения эпидемии

Антон Юрьевич Бишук

Московский физико-технический институт

Отчет по НИР/Группа 774, осень 2020

Научный руководитель: Зухба Анастасия Викторовна

Цель работы

Построить модель, способную определять множество вершин графа, получение информация о которых больше всего уменьшит энтропию всей системы.

Задачи

- Выбор метрики качества;
- Выбор подхода к решению задачи.

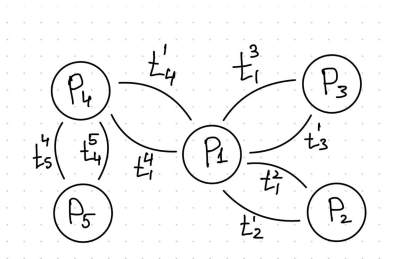
- ❶ i' – статус i -го человека (0 – здоров, 1 – болен);
- ❷ p_i – вероятность того, что человек i – болен;
- ❸ k_i – число контактов i -го человека;
- ❹ t_i^j – время контакта i -го человека с j -ым.

Тогда во всей системе $n + 2 \sum_{i=1}^n k_i + 2 \sum_{i=1}^n [t_i]$

Модель: инициализация

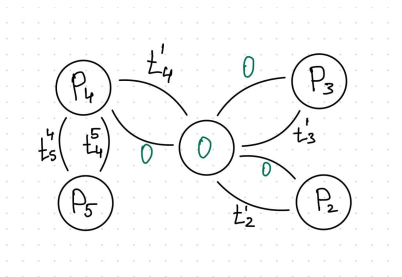
Инициализируем каждого пользователя вероятностью того, что он был на текущий момент зараженным.

Такой подход позволяет указывать заведомо больных и здоровых людей просто ставя вероятность болезни 1 и 0 соответственно.



Модель: обновление параметров (1)

$$f(t_1^2) = f(t_1^3) = f(t_1^4) = 0$$

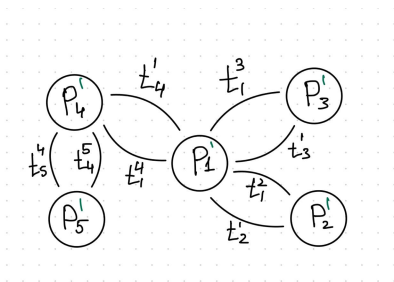


проверили человека и он оказался здоров, тогда граф на следующей итерации будет выглядеть следующим образом

Модель: обновление параметров (2)

$$p'_1 = 1 \text{ далее всегда}$$

На этого человека теперь так же не влияют другие.



проверили человека и он оказался больным, тогда граф на следующей итерации будет выглядеть следующим образом

- ① Если мы проверили i -го человека и он оказался здоров, то мы определили $k_i + [t_i]$ параметров системы;
- ② Если мы проверили i -го человека и он оказался болен, то мы определили $1 + 2k_i + [t_i]$ параметров системы;

Тогда мы получаем, что:

$$\begin{cases} \text{с вероятностью } p_i & \text{определим } 1+2k_i + [t_i] \text{ параметров} \\ \text{с вероятностью } 1-p_i & \text{определим } k_i + [t_i] \text{ параметров} \end{cases}$$

Задачу можно переформулировать на языке теории информации. Вся система имеет свою энтропию. При проверке пользователя мы уменьшаем её. При этом чем больше параметров мы своим выбором определяем, тем сильнее уменьшится энтропия всей системы.

- ❶ Поиск нужных людей можно переформулировать в задачу поиска вершин с наибольшей центральностью. Поэтому необходимо определить какие центральности можно для этого использовать.
- ❷ Поскольку мы имеем дело с не замкнутой системой, то наш алгоритм не должен быть детерминированным. Необходимо иногда делать случайные блуждания в графе.