Обучение признаковых представлений для вершин в ориентированных графах

Иванов Олег Юрьевич[1]
Бартунов Сергей Олегович[2]

1: *Студент, факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия*
2: *Аспирант, Вычислительный Центр РАН им. А. А. Дородницына*
E-mail: tigvarts@gmail.com, sbos@sbos.in

В задачах машинного обучения объекты могут являться вершинами некоторого графа. Примером таких задач может быть анализ социальных сетей, веб–графов, графов цитирования и так далее. Стандартные широко используемые методы машинного обучения, такие как метод опорных векторов, нейронные сети, случайные леса, предназначены для работы с выборкой независимых вещественных векторов. Для того, чтобы применить эти методы для анализа образовывающих граф объектов, предлагается метод, обучающий для каждой веришны ориентированного графа представление — вещественный вектор заданной размерности. Эти представления содержат информацию о локальной и глобальной структуре графа, поэтому можно работать с объектами графа как с выборкой независимых вещественных векторов, используя эти представления как признаки объектов.

Наиболее популярный метод отображения вершин графа в вещественные вектора называется спектральной проекцией графа [1]. Однако несмотря на интересные теоретические свойства, спектральные проекции имеют ряд недостатков в качестве векторов признаков объектов, что показано в данной работе.

Пусть V — множество всех вершин графа, $E = \{(u_i, v_i)\}_{i=1}^{|E|}$ — множество всех его ребер. Каждой вершине $u \in V$ ставится в соответствие два её представления: $\mathrm{In}_u \in \mathbb{R}^D$ (входное) и $\mathrm{Out}_u \in \mathbb{R}^D$ (выходное). Обозначим совокупность всех таких представлений для всех вершин графа как $\theta = \{\mathrm{In}_u, \mathrm{Out}_u\}_{u \in V}$.

В данной работе рассматривается вероятностная билинейная модель связности (БМС) $p(v|u, \theta)$, которая задает вероятность ориентированного ребра между двумя вершинами u и v при фиксированной первой вершине u через скрытые представления вершин в графе:

$$p(v|u, \theta) = \frac{\exp(\mathrm{In}_u^{\mathrm{T}} \mathrm{Out}_v)}{\sum_{w \in V} \exp(\mathrm{In}_u^{\mathrm{T}} \mathrm{Out}_w)} \tag{1}$$

Заметим, что описанный в работе метод позволяет обучать представления для широкого класса вероятностных моделей $p(v|u, \theta)$, а не только модели (1).

На основе вероятности ссылки с фиксированной первой вершиной $p(v|u, \theta)$ можно найти вероятность ссылки без фиксированной первой вершины $p(u, v|\theta)$ и использовать принцип максимального правдоподобия:

$$J(\theta) = \sum_{i=1}^{|E|} \log p(u_i, v_i|\theta)$$

Оптимизируя $J(\theta)$ с использованием L_2 регуляризатора на θ, можно получить обученные представления. Для быстрой минимизации используется метод Noise Contrastive Estimation [2].

БМС позволяет автоматически обучать представления для вершин ориентированного графа больших размерностей (миллионы вершин и ребер). Также он легко адаптируется для параллельного обучения, обучения на лету (то есть при появлении новых ребер в процессе обучения). Из-за итерационной природы метода, представления приемлемого качества можно получить намного раньше, чем $J(\theta)$ окончательно сойдется.

Полученные представления позволяют не только с некоторой точностью восстановить исходный граф, но и пригодны для предсказания новых ребер. На данный момент наилучшие результаты в задаче предсказания ребер показывают случайные блуждания (СБ) [3]. AUC для предсказания новых ребер с использованием представлений БМС сравним с результатами наилучших известных на данный момент методов предсказания ребер (таблица 1), однако позволяет вычислять вероятность ссылок после обученния представлений намного быстрее (таблица 2).

Можно заметить, что в БМС входное и выходное представление вектора неравноправны. Если из вершины u не выходит ребер, то $J(\theta)$ не зависит от входного представления In_u, и входное представление In_u для этой вершины не обучается. Однако $J(\theta)$ зависит от выходного представления Out_u для любой вершины u. Поэтому в качестве обученных признаков для вершин лучше использовать выходные представления.

Полученные выходные представления могут быть использованы для визуализации графа. Предварительные эксперименты показывают, что классификаторы, обученные на этих представлениях, позволяют эффективно выделять, например, сообщества в социальных

сетях или определенную категорию документов по их ссылкам друг на друга.

Данные [4]	БМС, $D = 30$	Джаккард	ЛСБ(3)	ССБ(3)
soc-LiveJournal	0.975	0.938	0.986	0.985
soc-Pocek	0.978	0.850	0.966	0.967
web-Google	0.961	0.945	0.977	0.978
web-BerkStan	0.979	0.960	0.996	0.996
cit-HepPh	0.983	0.962	0.988	0.989

Таблица 1: Предсказание ребер, AUC
Джаккард — метрика Джаккарда (англ. Jaccard metric [3])
ЛСБ(3) — локальные СБ (англ. Local Random Walk [3]) с 3 шагами
ССБ(3) — совмещенные СБ (англ. Superposed Random Walk [3]) с 3 шагами

	БМС	Джаккард	ЛСБ (T шагов)	ССБ (T шагов)
Асимптотика	$O(D)$	$O\left(\frac{\|E\|}{\|V\|}\right)$	$O(\|E\|T)$	$O(\|E\|T)$
Реальное время, сек.	$D = 30$: 10^{-6}	$4.65 \cdot 10^{-5}$	$T = 3$: 3.13	$T = 3$: 3.13

Таблица 2: Время вычисления вероятности одной ссылки для разных методов. Время было измерерно для социальной сети LiveJournal [4].
Расшифровку см. в предыдущей таблице.

Литература

1. Chung F. R. K. Spectral Graph Theory (CBMS Regional Conference Series in Mathematics, No. 92), USA, American Mathematical Society, 1996.
2. Gutmann M., Hyvärinen A. Noise-Contrastive Estimation of Unnormalized Statistical Models, with Applications to Natural Image Statistics // In Journal of Machine Learning Research, 2012, T. 13, № 1, P. 307–361.
3. Lü L., Zhou T. Link prediction in complex networks: A survey // In Physica A Statistical Mechanics and its Applications, 2011, T. 390, P. 1150–1170.
4. Leskovec J., Krevl A. Stanford Large Network Dataset Collection: `http://snap.stanford.edu/data`