

УДК 004.942

ГЕНЕРАЦИЯ АТТРИБУТИРОВАННЫХ СЕТЕЙ С ПОМОЩЬЮ АВТОЭНКОДЕРОВ

Лысенко А. В. (Университет ИТМО)

Научный руководитель – к.ф.-м.н., Чунаев П. В.

(Университет ИТМО)

В текущей работе рассмотрена задача генерации атрибутированных синтетических графов, похожих на какой-то существующий граф или набор графов, где схожесть определяется через сходство топологических характеристик, а также через схожесть атрибутов. Разработана модель генерации синтетических атрибутированных сетей и проведены эксперименты, определяющие схожесть сгенерированных графов с исходным/исходными по топологическим характеристикам и по качеству генерации атрибутов вершин.

В настоящее время генерация синтетических графов применяется для ряда задач: синтез молекул, с целью оптимизации какого-либо свойства, генерация 3D моделей по точкам, генерация синтетических социальных сетей дружбы, цитирования и других взаимодействия для дальнейшего моделирования. До определенного времени в основном использовались классические методы, например, Эрдеша-Реньи или Барабаши-Альберта, но основная проблема таких моделей заключается в том, что получающиеся случайные графы моделируют только небольшое число характеристик, например распределение степеней вершин, теряя другие свойства. Чтобы улавливать более сложные графовые паттерны, начали активно применять нейронные сети, в том числе и графовые, которые позволили продвинуться в этой области.

В настоящее время существует несколько различных подходов к генерации синтетических графов с помощью нейронных сетей: методы, использующие автокодировщики; методы, использующие архитектуру GAN, авторегрессионные методы и обучение с подкреплением. Каждый из перечисленных имеет свои сильные стороны и свою область применения. Так, обучение с подкреплением используется при ограничительных правилах на генерируемые графы, например число связей для атома в молекуле. Методы, основанные на автокодировщиках используются в основном для небольших графов, где число вершин не превосходит 100, поскольку процесс генерации в таких моделях является затратным по памяти, в то время как авторегрессионные методы являются более масштабируемыми на текущий момент.

В данной работе используется подход, использующий автокодировщики, поскольку в отличие от авторегрессионных методов, получаемый граф не зависит от порядка нумерации вершин. В качестве энкодера используется графовая нейронная сеть, а скрытое пространство моделируется распределением, полученным на векторах вершин исходного графа в скрытом пространстве. Такой подход использован из-за того, что в случае графов, кодируются сами вершины, что не позволяет привести представления вершин к многомерному нормальному распределению, не теряя в качестве раскодированных графов.

В ряде экспериментов показано качество генерируемых графов, в которых в качестве исходных данных использовались различные атрибутированные графы, например сети цитирований или сети репостов в твиттере.

Лысенко А. В. (автор)

Чунаев П. В. (научный руководитель)