

УДК 004.942

УЛУЧШЕНИЕ ЭКСТРАПОЛИРУЮЩИХ СПОСОБНОСТЕЙ ГРАФОВЫХ НЕЙРОННЫХ СЕТЕЙ С ИСПОЛЬЗОВАНИЕМ ПРИЧИННО-СЛЕДСТВЕННЫХ МОДЕЛЕЙ

Андреева П. О. (Университет ИТМО)

Научный руководитель – Шиков Е.Н. (Университет ИТМО)

В данной работе рассмотрено текущее положение исследований экстраполирующих способностей графовых нейронных сетей, определены основные проблемы существующих методов. Также предложены основные характеристические свойства графов, которые потенциально могут влиять на метки, проведены эксперименты по восстановлению причинно-следственной модели и проверке улучшения результатов классификации графов при использовании только прямых причин.

Введение. Несмотря на то, что в настоящее время алгоритмы глубокого обучения показывают высокие результаты во многих задачах, результаты значительно ухудшаются если тренировочные и тестовые данные были взяты из различных распределений. Такая проблема называется экстраполированием нейронных сетей. В настоящее время существует множество методов для улучшения экстраполирующих способностей нейронных сетей для задач на табличных данных и намного меньше – на графах. Кроме того, поскольку сложно определить, что значит распределение данных для графов, то существующие методы для улучшения обобщения графовых нейронных сетей предлагают различные постановки и позволяют решать задачи только для случаев, когда известна некоторая априорная информация про графы и их метки. В данной работе предложено использовать причинно-следственные модели для определения того, какие именно части информации в графе влияют на метку в задачах классификации графов, а также проведены эксперименты по подтверждению работоспособности данного подхода.

Основная часть. Причинно-следственные модели предлагают определить различные контексты как вмешательство в модель структурных уравнений, являющихся ключевым понятием в причинно-следственной теории. Если нейронная сеть обучается на данных из нескольких контекстов, а не на всех, то она может учитывать возникающие поверхностные корреляции в данных. Таким образом, в данных из другого контекста, где нет тех же поверхностных корреляций результат значительно ухудшается. Для того, чтоб нейронные сети могли имитировать когнитивные человеческие способности, необходимо, чтоб алгоритмы имели возможность отделять истинные корреляции между входными данными и результатом от поверхностных. Стабильное обучение предлагает декоррелировать всевозможные переменные за счет взвешивания данных. На графовых данных предложена работа по декоррелированию каждой размерности векторного представления графов, однако такой подход все еще не дает понимания того, что именно влияет на результат. Так как существуют подходы по восстановлению структуры направленных ациклических графов, отражающих модель структурных уравнений, то предлагается выделить основные элементы данных, которые могут влиять на метку и по ним восстановить причинно-следственную модель. Тогда при обучении графовых нейронных сетей необходимо использовать только прямые причины, что приведет к более стабильным результатам на тестовых данных. Такими элементами данных были определены: каждая размерность вектора атрибутов и каждый «цвет» в тесте Вейсфейлера-Лемана, отражающего структуру графа.

В качестве экспериментов, подтверждающих работоспособность данного подхода была проведена классификация графов с помощью всех данных и с помощью только прямых причин; зафиксированы улучшения точности классификации.

Выводы. В работе были рассмотрены существующие подходы к улучшению экстраполирующих способностей графовых нейронных сетей, выделены основные проблемы существующих методов, а также предложен новый подход, использующий причинно-следственную теорию. Кроме того, предложены элементы данных, которые могут влиять на результат и имеют доступную для человеческого восприятия интерпретацию. Также были проведены эксперименты, показывающие улучшение результатов при использовании только истинных причинно-следственных связей. Дальнейшая работа будет заключаться в построении новой графовой нейронной сети, учитывающей истинные причинно-следственные связи без предварительной обработки данных.

Андреева П.О. (автор)

Шиков Е.Н. (научный руководитель)