

1. Как мог бы выглядеть выход нейронной сети, если бы она должна была выдавать неориентированный граф (количество вершин ограничено)?

С помощью слоя Conv2D с сигмоидной функцией активации на выходе неориентированный граф может быть представлен как матрица смежности.

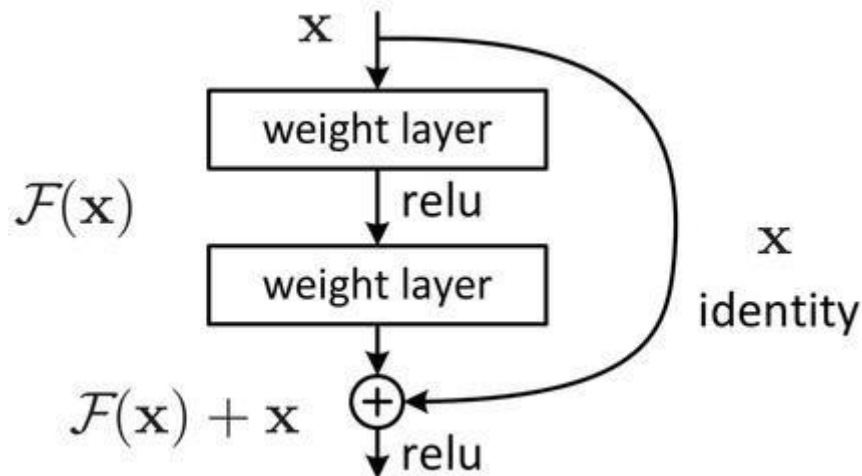
2. Что такое проблема долговременной зависимости в рекуррентных сетях?

Идея RNN сетей состоит в умении связывать предыдущую информацию с текущей задачей. Например, если стоит задача предсказать последнее слово в предложении “облака плывут по небу”, не нужен более широкий контекст; в этом случае довольно очевидно, что последним словом будет “небу”. В этом случае, когда дистанция между актуальной информацией и местом, где она понадобилась, невелика, RNN могут обучиться использованию информации из прошлого. Однако, сеть теряет способность связывать информацию при наличии долговременных зависимостей. Например, если стоит задача предсказать последнее слово в тексте “Я вырос во Франции... Я бегло говорю по-французски”. Ближайший контекст предполагает, что последним словом будет название языка, но, чтобы установить, какого именно языка, нужен контекст Франции из более отдаленного прошлого. Таким образом, разрыв между актуальной информацией и точкой ее применения может стать очень большим. К сожалению, по мере роста этого расстояния, RNN теряют способность связывать информацию. В теории проблемы с обработкой долговременных зависимостей у RNN быть не должно. Человек может аккуратно подбирать параметры сети для решения искусственных задач такого типа. К сожалению, на практике обучить RNN этим параметрам кажется невозможным.

3. Что такое shortcut соединение и для чего оно нужно?

Проблема деградации, которая возникает, когда более глубокая сеть начинает сворачиваться и точность сначала увеличивается, а затем быстро ухудшается, подразумевает, что сложная нелинейная функция, полученная добавлением нескольких слоев, должна выучить тождественное

преобразование, в случае если на предыдущих слоях был достигнут предел точности. Чтобы преодолеть проблему деградации, Microsoft ввела глубокую «остаточную» структуру обучения, при которой нейронная сеть обучается предсказывать функцию  $F(x) - x$ , вместо функции  $F(x)$ . Для компенсации этой разницы и вводится это замыкающее соединение (shortcut), которое добавляет недостающий  $x$  к функции.



4. Какая максимальная точность была получена? Как ее можно повысить?

Наибольшая точность составила примерно 75%. Для повышения точности можно обучать сеть на большем количестве эпох, повысить вероятность исключения нейронов на слоях Dropout, найти другую комбинацию слоев и их параметров.

5. Что будет происходить, если значения Dropout будет равно 0?

Слой Dropout получает вероятность строго между 0 и 1. Поведение Dropout при значении 0 не определено.

6. Можно ли помещать слой MaxPooling друг за другом?

Не учитывая целесообразность такого действия, поместить слои MaxPooling друг за другом можно.