**LSTM**

Что такое нейронная сеть?

Нейронная сеть (также искусственная нейронная сеть, ИНС) — математическая модель, а также её программное или аппаратное воплощение, построенная по принципу организации и функционирования биологических нейронных сетей — сетей нервных клеток живого организма.

Deep learning algorithm

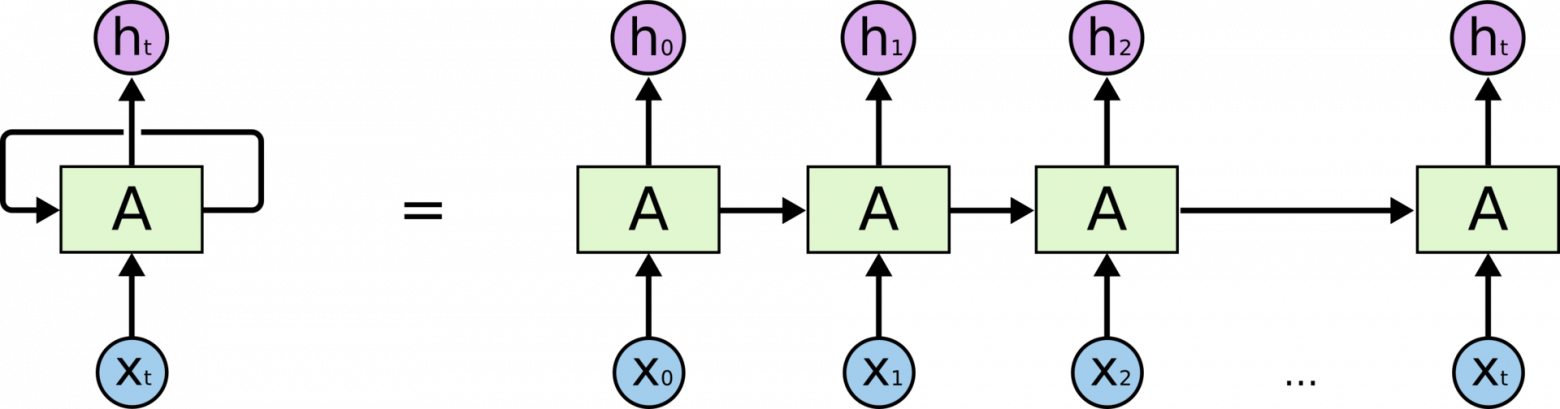
ML

Рекуррентный нейронные сети

Мы не думаем заново каждую секунду. Например, читая что-либо, мы понимаем каждое слово, основываясь на понимании предыдущего слова. То есть можно сказать, что наши мысли обладают постоянством.

Традиционные нейронные сети не обладают этим свойством. Решить эту проблему помогают рекуррентные нейронные сети. Это сети, содержащие обратные связи и позволяющие сохранять информацию.

На картинке можем увидеть, что фрагмент нейронной сети входное значение x\_t и возвращает значение h\_t.  Наличие обратной связи позволяет передавать информацию от одного шага сети к другому. Но все же они не сильно отличаются от обычных нейронных сетей. Рекуррентную сеть можно рассматривать, как несколько копий одной и той же сети, каждая из которых передает информацию последующей копии.



То, что RNN напоминают цепочку, говорит о том, что они тесно связаны с последовательностями и списками. RNN – самая естественная архитектура нейронных сетей для работы с данными такого типа.

Проблема рекуррентных нейронных сетей

Вся прелесть РНН состоит в том, что они умеют связывать предыдущую информацию с текущей задачей. Но всегда ли РНН предоставляют нам такую возможность? Это будет зависеть от некоторых обстоятельств.

Иногда для выполнения задачи нам необходима только недавняя информация. То есть в том случае, когда дистанция между актуальной информацией и местом, где она понадобилась, невелика, РНН будут очень даже полезны.

Но при увеличении этой условной дистанции (то есть для текущей задачи нам необходимо будет больше контекста) РНН уже теряют способность связывать информацию. Так мы медленно приблизились к LSTM – долгой краткосрочной памяти.

Сети LSTM

Долгая краткосрочная память – это разновидность архитектуры РНН, способная к обучению долговременным зависимостям.  Они прекрасно решают целый ряд разнообразных задач и в настоящее время широко используются.

LSTM разработаны специально, чтобы избежать проблемы долговременной зависимости. Запоминание информации на долгие периоды времени – это их обычное поведение.

Как мы уже выяснили, любая РНН имеет форму цепочки повторяющихся модулей. Но, скажем, в обычной РНН ячейка нейронной сети имеет всего один слой нейронов с функцией активации гиперболический тангенс.

*Лирическое отступление (немного о функциях активации)*

*Сначала ответим на вопрос, что делает искусственный нейрон? Если говорить просто, то он считает взвешенную сумму (сумма входных сигналов, умноженная на соответствующие веса, эта сумма характеризует поступивший на нейрон сигнал в целом), добавляет смещение (bias – это особый нейрон, особенность которого заключается в том, что его вход и выход всегда равняются 1 и он никогда не имеет входных синапсов) и решает, следует это значение исключать или использовать дальше.*

*Короче говоря, функция активации определяет выходное значение нейрона в зависимости от результата взвешенной суммы входов и порогового значения. Но суть в том, что получившееся значение может варьироваться от -бесконечности до +бесконечности. Поэтому нужно было придумать что-то, что помогло бы внешним связям определять активирован нейрон или его можно игнорировать.*

*В данном докладе будут рассматриваться две функции активации: сигмоида и гиперболический тангенс.*

Структура LSTM также напоминает цепочку, но модули выглядят иначе. Вместо одного слоя нейронной сети они содержат целых четыре. И эти слои по-особенному между собой взаимодействуют.

Изображение выглядит как текст, часы

Автоматически созданное описание

LSTM сеть развернута во времени, на схеме можно увидеть копию сети в предыдущий момент времени, копию сети в настоящий момент времени и копию сети в следующий момент времени.

На вход поступают элементы последовательности, и в каждый момент времени сеть LSTM дает какой-то результат. Причем сеть дает в свой следующий момент времени не одно значение, а два.

Архитектура ячейки LSTM: состоит из нескольких слоев - нейронов, которые соединяются между собой с помощью операции поэлементного умножения или сложения. Основой в ячейке является состояние, оно позволяет сохранить данные на длительный промежуток времени. Верхняя стрелка: сигнал, который поступает с предыдущего этапа сети и передается на следующий этап. Именно из-за этого ячейка может хранить данные очень долго.

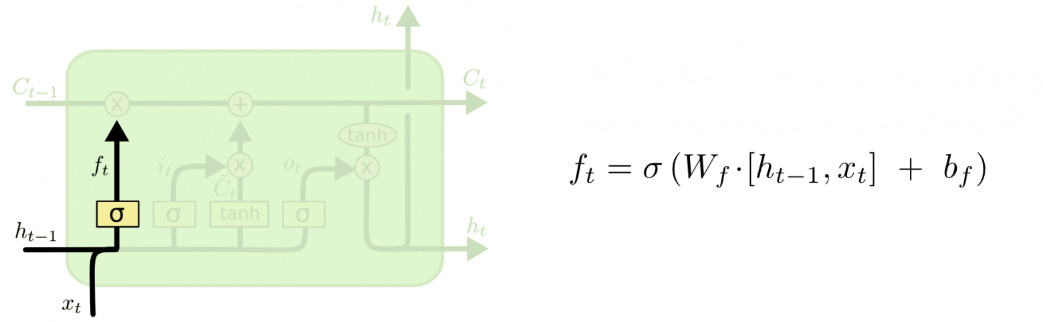
Изображение выглядит как текст, часы

Автоматически созданное описание

Для того, чтобы управлять состоянием ячейки, используются так называемые специальные вентили (gate). Это слой нейронов, который имеет на выходе операцию сигмоид, в котором значения меняются от 0 до 1 и последующая операция поэлементного умножения. Передаваемое значение может быть ослаблено, если значение на выходе сигмоида колеблется от 0 до 1.

Пошаговая структура

1. Первым идет вентиль забывания, так как он определяет, что нам нужно стереть из ячейки. Он анализирует текущие данные, которые поступили на вход нейронной сети и выходные значения сети, которые были получены на предыдущем шаге. Этот слой с функцией активации сигмоид, если какой-то нейрон выдает значение 0, то соответствующее значение в ячейке состояния стирается.



1. Следующим идет входной вентиль. Он определяет какие данные нужно поместить в состояние. Для этого сначала производится расчет таких данных (tanh) с помощью нейронного слоя с функцией активации гиперболический тангенс, этот слой также анализирует данные, которые поступили на вход нейронной сети и поступили с предыдущего состояния НС. Второй нейронный слой (ф-я активации сигмоид) является вентилем, который задает, какие именно данные нужно записать в ячейку памяти.
2. *Как происходит обновление состояния*. На первом этапе мы определяем какие именно данные нужно стереть. Для этого мы получаем данные с выхода вентиля забвения и выполняем операцию поэлементного умножения с теми данными, которые пришли с предыдущего этапа работы нейронной сети. После этого выполняем операцию поэлементного сложения, с теми данными, которые получили на выходе из входного вентиля. И это значение является новым значением в ячейке состояния, и именно оно подается на вход сети в следующий момент времени.
3. *Выход нейронной сети*. Нужно посчитать значение, которое выдаст нейронная сеть. Есть нейронный слой с функцией активации гиперболический тангенс, который определяет какой сигнал должен быть выдан на вход, и есть выходной вентиль, там есть нейронный слой с ф-ей активации сигмоид и если на выходе 0, то сигнал не проходит (1 --- сигнал проходит без изменений).

Изображение выглядит как текст, часы

Автоматически созданное описание

Особенность такой нейронной сети в том, что она умеет выдавать выходные сигналы и одновременно формировать сигналы вентилей входных, выходных и вентилей забвения, для того, чтобы на выходе получались правильные сигналы.

Сеть LSTM является сейчас достаточно популярной, однако ее недостаток в том, что в ней много элементов, поэтому, чтобы обучить такую сеть нужны большие вычислительные ресурсы. Было предпринято много попыток упростить сети LSTM и даже усложнить архитектуру, для того чтобы они работали лучше и обучались проще. И один из упрощенных вариантов архитектуры LSTM являются сети GRU. Но это уже выходит за рамки моего доклада, поэтому прошу ознакомиться с сетями GRU самостоятельно.