

LSTM

Job van der Laken

December 2022

1 Algemeen

LSTM staat voor Long Short Term Memory en is een uitbreiding van de RNN's, ofwel Recurrent Neural Networks. Deze neurale netwerken worden onder andere gebruikt om voorspelling te maken met sequentiële data of tijdreeksdata. LSTM's en RNN's werken goed op dit soort data, omdat er tijdens het maken van een voorspelling rekening wordt gehouden met voorgaande inputs, i.p.v. het individueel behandelen van elke input. In dit verslag wordt er vanuit gegaan dat de lezer al enige kennis heeft over de basis van neurale netwerken en back-propagation.

2 RNN

Als eerst is een recurrent neural network niet hetzelfde als een traditioneel feed-forward neuraal netwerk. Zo heeft een RNN nog een extra connectie met de hidden layer, genaamd de feedback loop. Deze feedback loop zorgt ervoor dat de uitkomsten van voorgaande outputs ook een rol spelen in de huidige input en output. Dit is te zien in de figuur hieronder, aan de rechterkant van het gelijkheidsteken. Wanneer de sequentiële data waar je op traint wat langer is, kan er bij RNN's een probleem ontstaan waarbij er door te veel stappen een exploding gradient problem of vanishing gradient problem kan ontstaan.

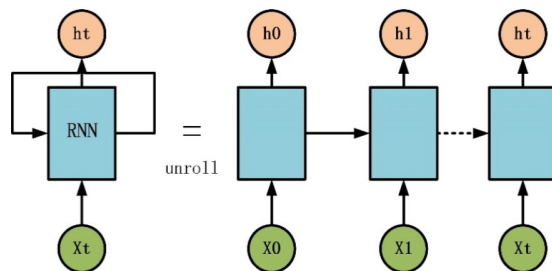


Figure 1: Unrolled RNN

2.1 Exploding and Vanishing gradient problem

Bij het exploding gradient problem, komen er na veel RNN's ontzettend grote getallen voor bij de gradiënten die de gewichten van het neurale netwerk aanpassen. Op het moment dat deze heel groot zijn, wordt het heel lastig om met backpropagation de meest optimale gewichten te vinden. Dit komt omdat door de grote getallen bij gradient descent steeds over de optimale oplossing heen wordt gesprongen.

Bij het vanishing gradient problem, gebeurt eigenlijk het omgekeerde. Hierbij nemen de getallen bij veel stappen juist exponentieel af, waardoor er uiteindelijk geen verandering meer zit in het aanpassen van de gewichten in het recurrent neural network er er geen optimale gewichten worden gevonden.

3 LSTM

Om wel te kunnen werken met langere sequentiële data bestaan er LSTM's. Net als bij RNN's, wordt er gebruik gemaakt van voorgaande data in de huidige input en output door middel van een feedback loop. In figuur 2 is een LSTM diagram te zien. Deze bestaat uit drie delen: de forgetgate, Inputgate en Output gate. Deze gates zijn op zichzelf neurale netwerken, die dienen als een soort filter van de data.

Als eerst begint het netwerk met de Forgetgate, waarin bepaald wordt welk percentage van de huidige cell state(long term memory) belangrijk is om te onthouden. Dit wordt gedaan met de hidden state(short term memory) en de nieuwe input data. Deze worden dan verwerkt in het neurale netwerk van de forgetgate door een sigmoid activatiefunctie, die een vector teruggeeft waarvan elk element in het interval $[0, 1]$ ligt. Op het moment dat de uitkomst dichtbij nul is, wordt er klein percentage onthouden van de huidige cell state en wanneer de uitkomst dichtbij 1 is, wordt er een groot percentage meegenomen.

De volgende gate is de input gate, waarin bepaald wordt wat de nieuwe long term memory gaat zijn volgens de huidige hidden state en de nieuwe input data. Dit zijn dezelfde twee inputs als bij de forgetgate en de output gate. Wat opvalt, is dat er bij de input gate 2 activatiefuncties staan. Als eerste wordt de tanh functie gebruikt om te bepalen in hoeverre elk element van de long term memory veranderd moet worden. Met de sigmoid functie wordt dan bepaald, hoe belangrijk de veranderingen van de tanh functie zijn voor de nieuwe long term memory. Net als bij de forgetgate betekent een waarde dichtbij nul niet belangrijk en een waarde dichtbij 1 wel belangrijk. De uitkomsten van deze twee functies worden dan met elkaar vermenigvuldigd en opgeteld bij de huidige cell state om de nieuwe long term memory te creëren.

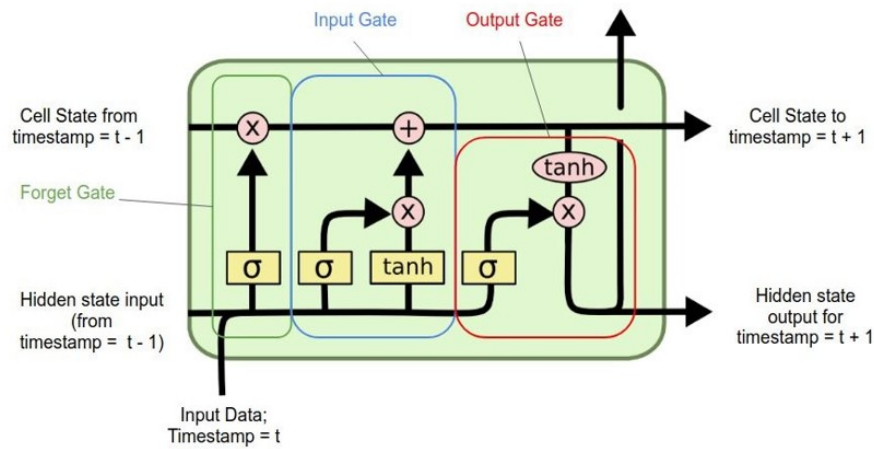


Figure 2: LSTM diagram

Het laatste deel gaat over de output gate, waar de nieuwe hidden state wordt gecreëerd. Als eerst gaat de vector van de nieuwe cell state door de tanh activatiefunctie. Hieruit wordt bepaald, soortgelijk als bij de input gate, in hoeverre de huidige hidden state moet worden aangepast. Het kan omschreven worden als de potentiële nieuwe short term memory. Met de sigmoid activatiefunctie wordt dan bepaald hoeveel procent van de potentiële short term memory onthouden moet worden voor de nieuwe hidden state. Om uiteindelijk een voorspelling te maken moet er nog een lineaire transformatie plaatsvinden over de hidden state. Hoe dit verder in zijn werk gaat valt buiten de scope.

Staudemeyer, R. C., Morris, E. R. (2019). Understanding LSTM—a tutorial into long short-term memory recurrent neural networks. arXiv preprint arXiv:1909.09586.