

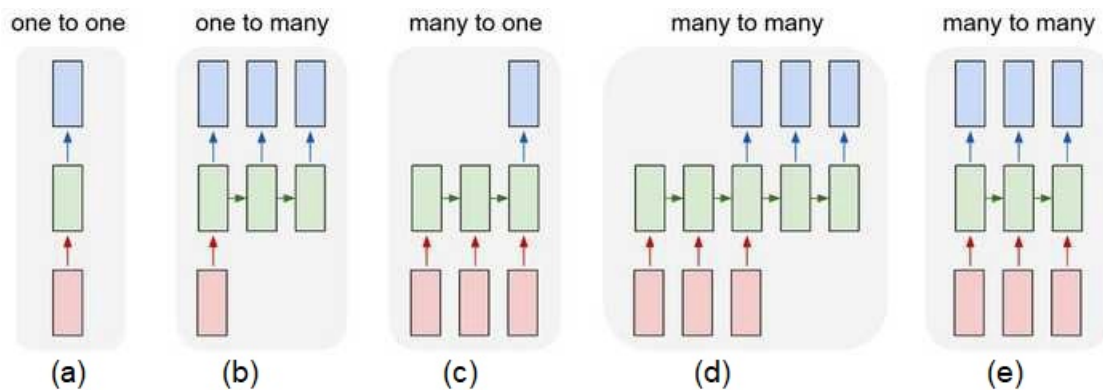
Dịch máy nơ ron - Neural Machine Translation

Người viết : Trịnh Anh Phúc

Người kiểm tra : x

1 Giới thiệu

Với mạng nơ ron truyền thống được định nghĩa với đầu vào, đầu ra kích thước *cố định* thường thích hợp dùng cho các dạng dữ liệu hình ảnh tại đó chúng được coi là độc lập xác suất, quá trình tiền xử lý cũng cần thao tác đặt biệt như xén ảnh, giãn ảnh, chuẩn hóa dài rộng của ảnh để vừa với đầu vào. Khác với mạng nơ ron truyền thống, mạng nơ ron truy hồi cho phép đầu vào là chuỗi chiều dài thay đổi, tương ứng với nó là đầu ra cũng có thể là chuỗi chiều dài thay đổi. Mạng nơ ron truy hồi đặc biệt hữu dụng trong các bài toán đòi hỏi xử lý thông tin chuỗi (xem minh họa hình 1)



Hình 1: Minh họa các dạng ứng dụng mạng nơ ron truy hồi

Danh sách các ứng dụng có thể liệt kê với chuỗi

- Sinh văn bản mô tả (description sentence from image) dạng (b) sử dụng mạng nơ ron truy hồi dạng 'one-to-many', từ một hình ảnh đầu vào sẽ sinh câu mô tả về hình ảnh đó
- Phân tích cảm xúc (sentiment analysis) dạng (c) sử dụng mạng nơ ron truy hồi dạng 'many-to-one', chuỗi văn bản tương ứng bình luận được đưa vào mạng và đầu ra sẽ phân loại bình luận tích cực hay tiêu cực.
- Nhận diện ký tự tiếp theo Character-RNN dạng (e) sử dụng mạng nơ ron truy hồi dạng 'many-to-many', theo đó kích thước chuỗi dự đoán và chuỗi đầu vào là cùng nhau
- Dịch tự động (NEURAL MACHINE TRANSLATION) dạng (d) sử dụng mạng nơ ron truy hồi dạng 'many-to-many', theo đó chuỗi văn bản ngôn ngữ gốc cần dịch, có chiều dài thay đổi, sẽ được dịch sang chuỗi văn bản ngôn ngữ đích dịch sang.

Nội dung bài đọc sẽ được chia tuần tự thành các phần con như sau

1. Dịch máy nơ ron (NEURAL MACHINE TRANSLATION - NMT)
2. Cơ chế chú ý (NMT with Attention Mechanism)
3. Lớp chuyển vị (NMT with Transformer layer)

Chú ý, bạn đọc cần có kiến thức về mạng nơ ron truy hồi như RNNs, LSTMs, GRUs

2 Bài toán dịch cùng cơ chế chú ý (NMT with Attention Mechanism)

2.1 Bài toán dịch tự động

Như đã biết, bài toán dịch tự động là ứng dụng của mô hình tuần tự (many-to-many) dạng (d) theo đó với chuỗi từ nguồn \mathbf{x} (câu ngôn ngữ cần dịch) ta sẽ thực hiện sinh (hay tìm kiếm) chuỗi từ đích \mathbf{y} (câu ngôn ngữ được dịch) với tiêu chí "đúng" nhất. Chú ý, chiều dài của chuỗi nguồn và chuỗi đích thường là không bằng nhau và được gọi **sequence-to-sequence**. Ví dụ dưới đây

Chuỗi từ nguồn \mathbf{x} (tiếng Pháp)

"Il a m'étéarte"

được dịch tự động sang chuỗi từ đích \mathbf{y} (tiếng Anh)

"he hit me with a pie"

Việc dịch sẽ tương đương tìm câu đích \mathbf{y} cực đại xác suất có điều kiện với câu nguồn \mathbf{x} đã cho, hay nói một cách khác

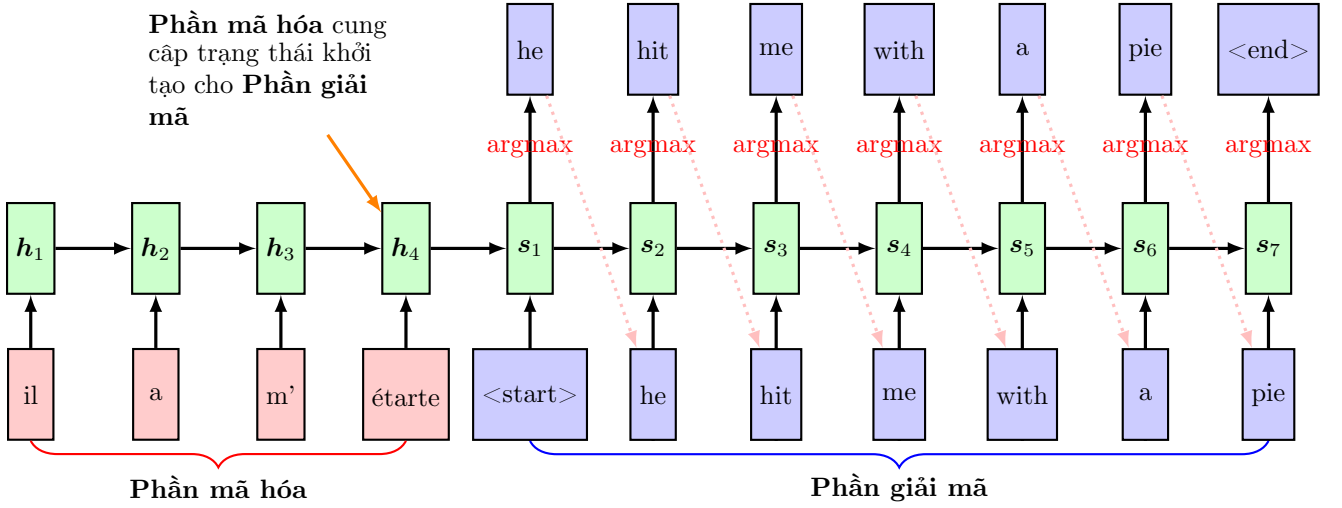
$$\mathbf{y} = \arg \max_{\mathbf{y}} p(\mathbf{y}|\mathbf{x})$$

Nghĩa là trong dịch máy nơ ron (NEURAL MACHINE TRANSLATION - NMT), ta cần tìm tham số sao cho cực đại hóa xác suất có điều kiện $p(\mathbf{y}|\mathbf{x})$, với bộ dữ liệu gồm các cặp (chuỗi từ nguồn, chuỗi từ đích). Một khi phân bố có điều kiện đã được ước lượng, tương ứng với mỗi câu \mathbf{x} đã cho, ta sẽ tìm câu \mathbf{y} theo tiêu chí trên. Do cả hai câu đều là chuỗi từ nên ta có thể ký hiệu theo chỉ số thời gian trái sang phải $\mathbf{x} = (x_1, \dots, x_{T_x})$ còn $\mathbf{y} = (y_1, \dots, y_{T_y})$. Đối với cách tiếp cận trên, NMT gồm có hai khối chính (xem minh họa Hình 2)

1. Khối mã hóa - encoder : dùng để mã hóa chuỗi nguồn \mathbf{x} , thường có chiều dài thay đổi, sang một vec tơ có chiều dài cố định (fixed-length vector) là vec tơ \mathbf{c} . Hướng tiếp cận phổ thông là dùng một mạng nơ ron hồi quy RNN sao cho

$$\mathbf{h}_t = f(x_t, \mathbf{h}_{t-1})$$

Lúc này, vec tơ khung cảnh kích thước cố định



Hình 2: Hoạt động mạng nơ ron truy hồi nhiều-nhiều dạng (d), khối bên trái gọi là khối mã hóa - encoder còn khối bên phải là khối giải mã - decoder, theo minh hoạ trạng thái khởi tạo chính tăng ẩn cuối cùng h_4 . Chú ý, minh hoạ mạng truy hồi là trong pha huấn luyện.

$$\mathbf{c} = g(\{\mathbf{h}_1, \dots, \mathbf{h}_{T_x}\})$$

trong đó, \mathbf{h}_t là tham số trạng thái ẩn \mathbb{R}^n tại thời điểm t , còn \mathbf{c} là vec tơ chiều dài cố định được sinh bởi chuỗi trạng thái ẩn với f và g là các hàm phi tuyến. Với bài báo ¹ dùng LSTMs cho hàm f còn $g(\{\mathbf{h}_1, \dots, \mathbf{h}_{T_x}\}) = \mathbf{h}_{T_x}$ giống minh hoạ Hình 2, nghĩa là vec tơ chiều dài cố định \mathbf{c} cũng thuộc \mathbb{R}^n .

2. Khối giải mã - decode : dùng để giải mã ra chuỗi đích \mathbf{y} , thường có chiều dài thay đổi, từ vec tơ \mathbf{c} có chiều dài cố định. Cách thức giải mã diễn ra theo từng thời điểm, dự đoán từ tiếp theo y_t khi đã biết vec tơ \mathbf{c} và toàn bộ các từ đã dự đoán trước đó $\{y_1, \dots, y_{t-1}\}$, tiếp tục xem minh hoạ Hình 2. Nói một cách khác, khối giải mã xác định một xác suất có điều kiện khi dịch \mathbf{y} bằng cách tính xác suất hợp của chuỗi các xác suất có điều kiện theo thứ tự thời gian

$$\mathbf{y} = \prod_t^T p(y_t | \{y_1, \dots, y_{t-1}\}, \mathbf{c})$$

trong đó $\mathbf{y} = (y_1, \dots, y_{T_y})$. Với RNN, mỗi xác suất có điều kiện được mô hình bởi

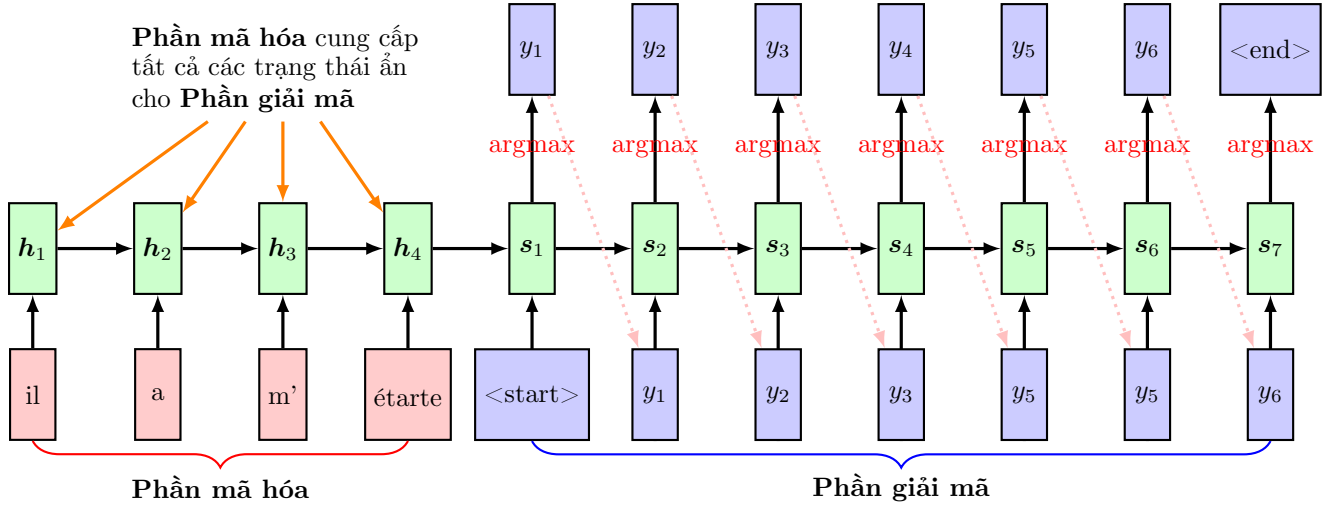
$$p(y_t | \{y_1, \dots, y_{t-1}\}, \mathbf{c}) = g(y_{t-1}, \mathbf{s}_t, \mathbf{c}) \quad (1)$$

trong đó g là hàm phi tuyến, khả năng mạng nơ ron đa tầng, xuất đầu là xác suất y_t , với \mathbf{s}_t là trạng thái ẩn của RNN.

Nhận xét chung về mô hình dịch nơ ron có tồn tại nhược điểm, khi giải mã câu đích thì ta lấy nhãn y_t có giá trị lớn nhất - argmax - tại từng thời điểm cho đến khi gặp từ kết thúc <end>, đây là cách giải mã tham lam, bởi nếu sai ở một từ tại thời điểm dự đoán nào đó, ta sẽ bị sai cả câu và không có cách nào sửa lại được câu dự đoán sai này. (Đọc thêm về beam search). Cơ chế chú ý được trình bày sau đây, cho phép quá trình giải mã dựa trên cả câu hoàn chỉnh và cho phép dóng - align - câu nguồn và câu đích bằng cách gán trọng số trên toàn bộ cặp (câu nguồn, câu đích) đảm vị trí *tương đối* giữa các cặp từ là đúng. Ví dụ như, cặp từ (il ↔ he) sẽ được dóng cùng đầu câu nguồn, câu đích.

¹Sutskever, I., Vinyals, O., and Le, Q. (2014). Sequence to sequence learning with neural networks. In Advances in Neural Information Processing Systems (NIPS 2014).

2.2 Cơ chế chú ý - Attention Mechanism



Hình 3: Hoạt động mạng nơ ron truy hồi nhiều-nhiều với cơ chế chú ý theo đó mọi thông tin của các tầng ẩn của mã hoá - encoder - sẽ được truyền làm trạng thái đầu vào cho mọi thời điểm t của phần giải mã - decoder. Theo minh hoạ ta có tất cả (h_1, \dots, h_4) đều được truyền sang phần giải mã. Minh hoạ trong pha giải mã hay dự đoán.

Theo bài báo ², việc dịch tự động được diễn ra bởi hai khối mã hóa - encoder và giải mã - decoder, đều được cài đặt bởi GRUs. *Cơ chế chú ý* thực hiện việc kết nối hoàn toàn chuỗi nguồn và chuỗi đích khi thực hiện dịch - phần giải mã - có chú ý đến vị trí t trông như sau

$$p(y_t | \{y_1, \dots, y_{t-1}\}, c) = g(y_{t-1}, s_t, c_t) \quad (2)$$

trong đó trạng thái ẩn s_t của phân giải mã

$$s_t = f(s_{t-1}, y_{t-1}, c_t)$$

Như vậy, xác suất có điều kiện sẽ phụ thuộc vào vec tơ khung cảnh từng vị trí c_t cho mỗi từ đích y_t .

Để mã hóa vec tơ khung cảnh c_t phụ thuộc vào toàn bộ chuỗi trạng thái ẩn (h_1, \dots, h_{T_x}) được mã hóa từ chuỗi nguồn $x = (x_1, \dots, x_{T_x})$. Mỗi trạng thái ẩn h_i sẽ chứa thông tin của toàn bộ chuỗi nguồn. ³ Khác với mô hình minh họa Hình 2, tại đó toàn bộ câu nguồn chỉ cung cấp trọng thái khởi tạo khi bước vào phần giải mã, ta chú trọng thêm vị trí t của vec tơ khung cảnh, thậm chí ta dùng mạng nơ ron hồi quy hai chiều (Bidirectional RNN) cho phần mã hóa để đảm bảo tính trạng thái ẩn h_i nhắm vào từ x_i phụ thuộc xung quanh cả từ phía trước lẫn từ phía sau nó trong câu nguồn. Như vậy, vec tơ khung cảnh sẽ được tính bởi tổng

$$c_t = \sum_{i=1}^{T_x} \alpha_{t,i} h_i$$

²Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural Machine Translation by Jointly Learning to Align and Translate. CoRR, abs/1409.0473.

³Chú ý, chỉ số t ta sẽ dùng cho phần giải mã còn chỉ số i ta dùng cho phần mã hóa.

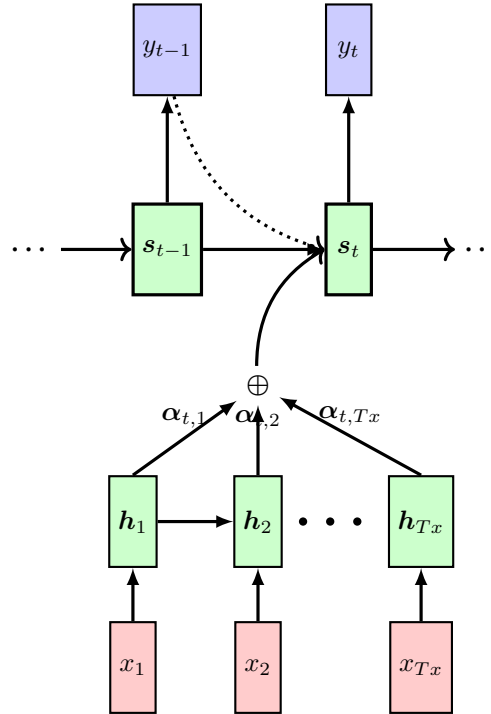
Trọng số $\alpha_{t,i}$ được tính dùng hàm *softmax* theo công thức

$$\alpha_{t,i} = \frac{\exp(e_{t,i})}{\sum_{k=1}^{Tx} \exp(e_{t,k})}$$

trong đó,

$$e_{t,i} = a(\mathbf{s}_{t-1}, \mathbf{h}_i)$$

là mô hình đóng - alignment model - cho phép đo sự tương đồng giữa từ đầu vào tại vị trí i và đầu ra dự đoán tại vị trí t . Giá trị đo này được tính dựa trên trạng thái \mathbf{s}_{t-1} và trạng thái ẩn \mathbf{h}_i tại vị trí chuỗi nguồn. Mô hình $a(\cdot)$ được tham số hóa bởi mạng truyền thẳng (trong bài báo là mạng một lớp - a single multilayer perceptron) được huấn luyện đồng thời với các thành phần khác của hệ thống. Mô hình đóng này hoàn toàn học có giám sát, thực hiện lan truyền ngược để tính toán các tham số tương ứng (xem minh họa Hình 4).



Hình 4: Cơ chế chú ý, tổng \mathbf{c}_t được biểu diễn bởi \oplus . Các trọng số $\alpha_{t,i}$ được cộng vào bởi các trạng thái ẩn \mathbf{h}_i được sinh từ chuỗi nguồn. Mô hình đóng $a(\cdot)$ chưa được minh họa ở đây.

Tóm lại, ta xác định các bước tuần tự khi giải mã dự đoán theo cơ chế chú ý tại thời điểm t như sau

1. Bắt đầu với toàn bộ ma trận trái thái ẩn của phần mã hoá $(\mathbf{h}_1, \dots, \mathbf{h}_{Tx})$ và giải mã ở thời điểm trước đó \mathbf{s}_{t-1} . Chú ý, chúng cùng kích thước.
2. Tính giá trị đo bằng mô hình đóng một đầu ra $e_{t,i} = a(\mathbf{s}_{t-1}, \mathbf{h}_i)$ với mọi $i = 1, \dots, Tx$. Mô hình đóng một đầu ra sẽ cho giá trị đo (score) nói rõ tầm quan trọng của trạng thái ẩn nào để dự đoán y_t tương ứng hay nói một cách khác là đóng trạng thái ẩn \mathbf{h}_i với nhãn dự đoán y_t tương ứng.

3. Sử dụng hàm softmax để chuẩn hoá các giá trị đo thu được này để có được $\alpha_{t,i}$ với mọi $i = 1, \dots, Tx$
4. Tính vec tơ khung cảnh theo công thức

$$\mathbf{c}_t = \sum_{i=1}^{Tx} \alpha_{t,i} \mathbf{h}_i$$

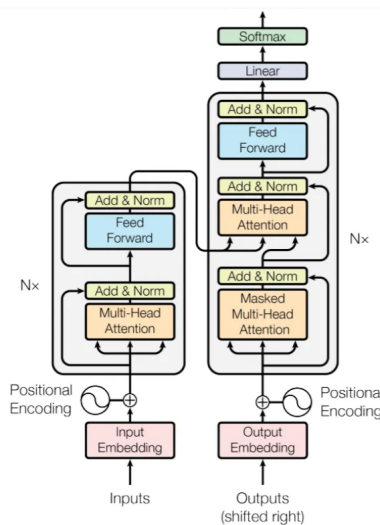
chú ý vec tơ khung cảnh này cùng kích thước với mọi trạng thái ẩn \mathbf{h}_i và \mathbf{s}_t

Chúng ta có thể hiểu là khi tổng hợp toàn bộ trạng thái ẩn $(\mathbf{h}_1, \dots, \mathbf{h}_{Tx})$, tức là ta đang tính giá trị kỳ vọng trên toàn bộ các phép đóng tương ứng với nhãn cần dự đoán y_t . Cho $\alpha_{t,i}$ là xác suất tại đó nhãn đích y_t được đóng, hay dịch từ, từ chuỗi gốc x_i . Nghĩa là, vec tơ khung cảnh \mathbf{c}_t là kỳ vọng trên toàn bộ các xác suất $\alpha_{t,i}$ trên không gian chuỗi nguồn.

Xác suất $\alpha_{t,i}$ được mô tả với giá trị độ đo $e_{t,i}$, phản ánh sự quan trọng của trạng thái ẩn \mathbf{h}_i phụ thuộc vào trạng thái \mathbf{s}_{t-1} trong quyết định trạng thái tiếp theo \mathbf{s}_t để sinh ra y_t tương ứng. Một cách ngầm định, đây là *cơ chế chú ý* - attention mechanism - theo đó phân giải mã quyết định phần nào của câu nguồn cần chú ý khi dịch, hay giải mã từ đích tại thời điểm tiếp theo t tương ứng. Bằng cách cài đặt cơ chế chú ý cho phân giải mã, ta có thể phần nào đó giảm bớt gánh nặng phải mã hóa mọi thông tin chuỗi nguồn tại phần mã hóa, vào trong một vec tơ kích thước cố định.

3 Bài toán dịch cùng lớp biến đổi (NMT with Transformer layer)

Bản chất lớn biến đổi là một lớp mạng nơ ron dùng cơ chế chú ý để **song song** hoá dữ liệu chuỗi. Chú ý, hai chuỗi đầu vào và đầu ra là không cùng kích thước và không cố định. Ý tưởng chính của mô hình sử dụng lớp biến đổi là chuỗi đầu vào có thể truyền song song vào GPU giúp tăng tốc quá trình huấn luyện. Thêm nữa, với lớp chú ý đa đầu (multi-head) cho phép giảm thiểu vấn đề suy giảm đạo hàm bậc nhất khi chiều dài của chuỗi tăng lên.

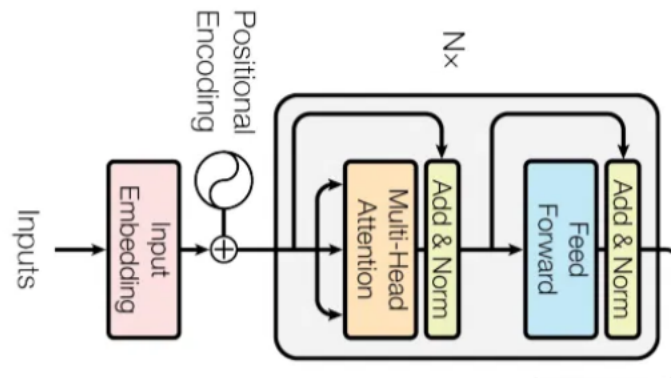


Hình 5: Kiến trúc mạng nơ ron biến đổi

Bài báo ⁴ được viết dựa trên ứng dụng dịch máy nơ ron NMT, kiến trúc mạng nơ ron minh hoạ Hình 6 hoạt động như sau ⁵

3.1 Khối mã hoá - Encoder Block

Để dễ hình dung, ta cắt và xoay hình ảnh trên thành hai phần mã hoá và giải mã như sau



Hình 6: Kiến trúc mạng nơ ron biến đổi phần mã hoá - encoder

Đầu tiên với mọi từ đầu vào ta cần chuyển sang vec tơ từ - input embedding, tuy nhiên như chúng ta đã biết mỗi từ trong một câu lại có nhiều nghĩa khác nhau nên ta cần sự trợ giúp của mô đun mã hoá vị trí - positional encoding. Nó là một vec tơ tạo khung cảnh khác nhau liên quan đến vị trí tự trong câu.

Từ \rightarrow Vec tơ từ \rightarrow Định vị vec tơ từ \rightarrow Vec tơ khung cảnh

3.1.1 Khối chú ý đa đầu - Multi-Head Attention

Điều quan trọng nhất của khối này là cơ chế tự chú ý - self attention. Cơ chế tự chú ý cho phép mạng tập trung vào các phần liên quan đến từng từ cụ thể trong câu, cụ thể nó là một vec tơ chú ý. Với mọi từ, ta sẽ có một vec tơ chú ý được sinh ra, nó nắm bắt mối quan hệ trong khung cảnh từ trong câu tương ứng.

		Ma trận vec tơ chú ý
The	\rightarrow The big red dog	$[0.71 \ 0.04 \ 0.07 \ 0.18]^T$
big	\rightarrow The big red dog	$[0.01 \ 0.84 \ 0.02 \ 0.13]^T$
red	\rightarrow The big red dog	$[0.09 \ 0.05 \ 0.62 \ 0.24]^T$
dog	\rightarrow The big red dog	$[0.03 \ 0.03 \ 0.03 \ 0.91]^T$

Hình 7: Cơ chế chú ý self-attention

Như vậy nếu câu càng dài thì vec tơ chú ý này sẽ cũng dài theo số từ được chuẩn hóa bởi softmax. Toàn bộ câu sẽ có gồm ma trận vuông các vec tơ chú ý từng từ riêng biệt. Nhắc lại, đây là dạng bài toán học

⁴Vaswani, A., Shazeer, N.M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., & Polosukhin, I. (2017). Attention is All you Need. Neural Information Processing Systems.

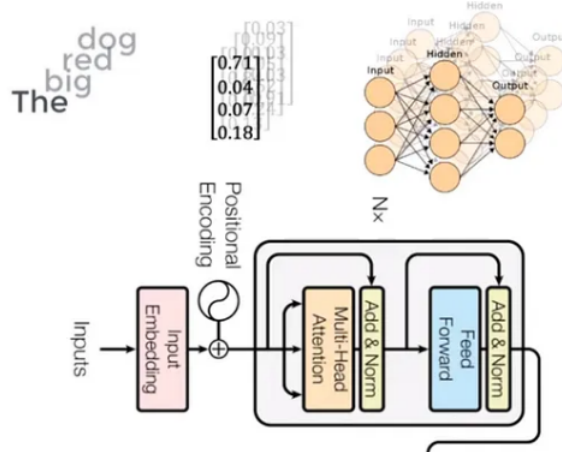
⁵<https://towardsdatascience.com/transformer-neural-network-step-by-step-breakdown-of-the-beast-b3e096dc857f>

không giám sát giữa các vec tơ từ Word2Vec trong cùng ngôn ngữ. Cơ chế tự chú ý trên riêng chuỗi ngôn ngữ nguồn hay đích đều có tác dụng thể hiện sự liên quan vị trí giữa các từ trong câu với văn phạm ngôn ngữ nguồn hoặc đích. Cơ chế chú ý phần trước là dùng *dòng* các từ hai ngôn ngữ khác nhau. Trong cơ chế tự chú ý, các vec tơ từ Word2Vec thay thế cho các trạng thái ẩn của từ trong phần mã hóa -encoder.

Vấn đề ở chỗ là trọng số đánh cho mọi từ của một câu có xu hướng cao hơn các từ đó trong câu khác. Chính bởi vậy, các tác giả đề nghị thực hiện cơ chế nhiều lần tương ứng với một từ và lấy giá trị trung bình để có được vec tơ chú ý cuối cùng. Cách này đảm bảo sự tương tác giữa các từ trong câu. Khi đó ta gọi chúng là khối chú ý đa đầu Multi-Head Attention.

3.1.2 Khối nơ ron truyền thẳng - FeedForward Network

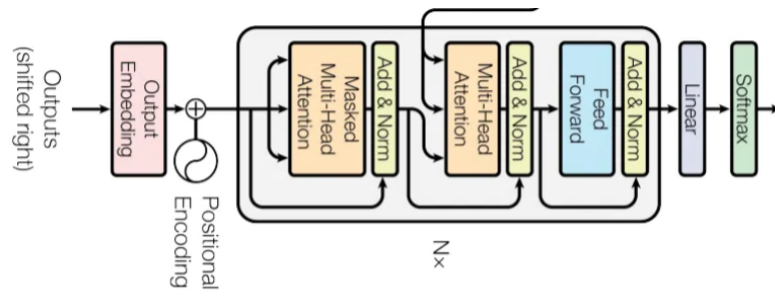
Khối tiếp theo là mạng nơ ron kết nối đầy đủ, được sử dụng cho mọi vec tơ chú ý. Nhiệm vụ chính của khối này là biến đổi - transform - vec tơ chú ý sang dạng mới mà phần giải mã sử dụng. Điều quan trọng nhất là các vec tơ chú ý là độc lập với nhau nên ta có thể song song hóa làm tăng tốc quá trình huấn luyện. Điều này khác với mạng RNN truyền thống là ta phải chờ đủ (h_1, \dots, h_{T_x}) chuỗi trạng thái ẩn mới quay lại - backward - được. Bây giờ, ta có thể đưa mọi từ đồng thời qua bộ mã hóa và có được tập các vec tơ được mã hóa của mọi từ lập tức.



Hình 8: Mạng nơ ron học song song trên các vec tơ chú ý độc lập

3.2 Khối giải mã - Decoder block

Như vậy, thông tin đầu ra của khối mã hóa sẽ làm thông tin đầu vào khối giải mã. Với NMT, ta lại dùng tiếp ví dụ chuỗi nguồn là ngôn ngữ tiếng Anh và chuỗi đích ra khỏi khối giải mã là ngôn ngữ tiếng Pháp. Đầu tiên hai khối chức năng nhúng và định vị làm nhiệm vụ giống với phần mã hóa trước đó. Tiếp đó khối tự chú ý đa đầu sẽ sinh các vec tơ chú ý trong ngôn ngữ tiếng Pháp, khối này hoạt động cũng tương tự như phần mã hóa. Tuy nhiên khối này cần có mặt nạ (Masked Multi-Head Attention), lý do là đây là phần pha dự đoán tại đó ta sẽ không biết là từ nào tiếng Pháp sẽ đưa vào kết quả ra. Để dễ hình dung, ta cắt và xoay hình ảnh trên thành hai phần mã hoá và giải mã như sau



Hình 9: Kiến trúc mạng nơ ron biến đổi phần giải mã - decoder

3.2.1 Khối chú ý đa đầu mặt nạ - Masked Multi-Head Attention

Để giải thích cơ chế học của bài toán dịch trong khối tự chú ý này, ta bắt đầu với từ tiếng Anh được nạp vào đầu vào của phần giải mã. Phần giải mã sẽ dịch sang từ tương ứng tiếng Pháp bằng cơ chế tự chú ý, sau đó kết quả này sẽ được so sánh với dữ liệu dịch thực tế bởi câu tiếng Pháp (chúng được cung cấp trong pha huấn luyện). Sau khi so sánh, các vec tơ chú ý này sẽ được cập nhật lại sao cho đúng trong ngôn ngữ đích qua các vòng lặp.

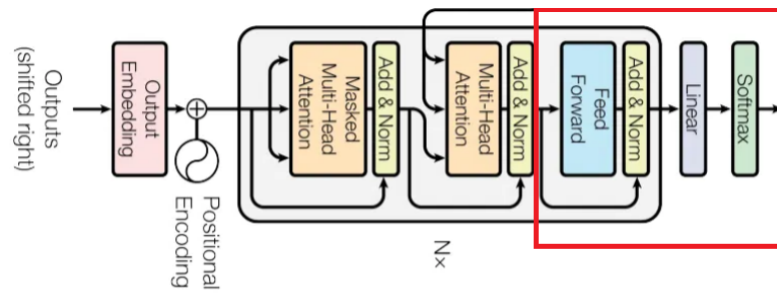
Ta quan sát được là cần phải dấu từ tiếng Pháp tiếp theo, theo đó nó cần dự đoán trước từ tiếp theo, không biết nhãn đúng thực sự, rồi mới so sánh được kết quả dự đoán và nhãn để học. Trong quá trình học, nó không có nghĩa gì lắm vì ta đã biết trước từ tiếng Pháp tiếp theo.



Hình 10: Kiến trúc mạng nơ ron biến đổi phần giải mã cơ chế tự chú ý đa đầu mặt nạ

Về cơ bản ta có thể lấy mọi từ từ câu nguồn tiếng Anh nhưng chỉ có thể lấy từ trước đó tại câu đích tiếng Pháp để học. Vì vậy, khi thực hiện song song hóa khi huấn luyện, ta cần **mặt nạ** các vị trí sau vị trí hiện tại bởi 0 như vậy cơ chế tự chú ý đa đầu sẽ không dùng được chúng khi huấn luyện giống như khi dự đoán.

Như vậy, hai kết quả các vec tơ chú ý với mọi từ trong câu nguồn tiếng Anh và câu đích tiếng Pháp sẽ được dẫn đến khối chú ý đa đầu **Multi-Head Attention Block**. Khối này học ánh xạ mọi từ tiếng Anh sang mọi từ tiếng Pháp trong chuỗi nguồn sang chuỗi đích và tìm mối quan hệ giữa chúng. Phần này là phần học ánh xạ tiếng Anh và tiếng Pháp. Đầu ra của khối này là các vec tơ chú ý của mọi từ câu nguồn tiếng Anh và câu đích tiếng Pháp. *Mỗi vec tơ chỉ ra mối quan hệ của nó với từ trong cả hai ngôn ngữ tiếng Anh và tiếng Pháp.*



Hình 11: Kiến trúc mạng nơ ron biến đổi phần giải mã với đầu ra dự đoán từ tiếng Pháp

3.2.2 Khối truyền thẳng đầu ra

Toàn bộ vec tơ chú ý sẽ được đưa qua khối cuối cùng để chuyển sang không gian ngôn ngữ đích, tiếng Pháp. Một mạng nơ ron truyền thẳng tuyến tính sẽ tăng kích thước bằng kích thước từ điển tiếng Pháp. Sau đó với lớp chuẩn hóa xác suất, sẽ giúp ta xác định được từ dự đoán trong ngôn ngữ đích (Xem minh họa Hình 11). Chú ý, khối tuyến tính Linear cần có kích thước bằng từ vựng tiếng Pháp trong tình huống này.

4 Tổng kết

Ứng dụng nơ ron dịch tự động là một trong ứng dụng khá phổ biến trên nhiều nền tảng trực tuyến hiện nay. Ứng dụng tạo nên cuộc cách mạng về dịch tự động không những dữ liệu văn bản và hình ảnh. Gần hiện nay đã các dịch đôi một nhiều ngôn ngữ phổ biến Anh-Pháp, Pháp-Đức, ...