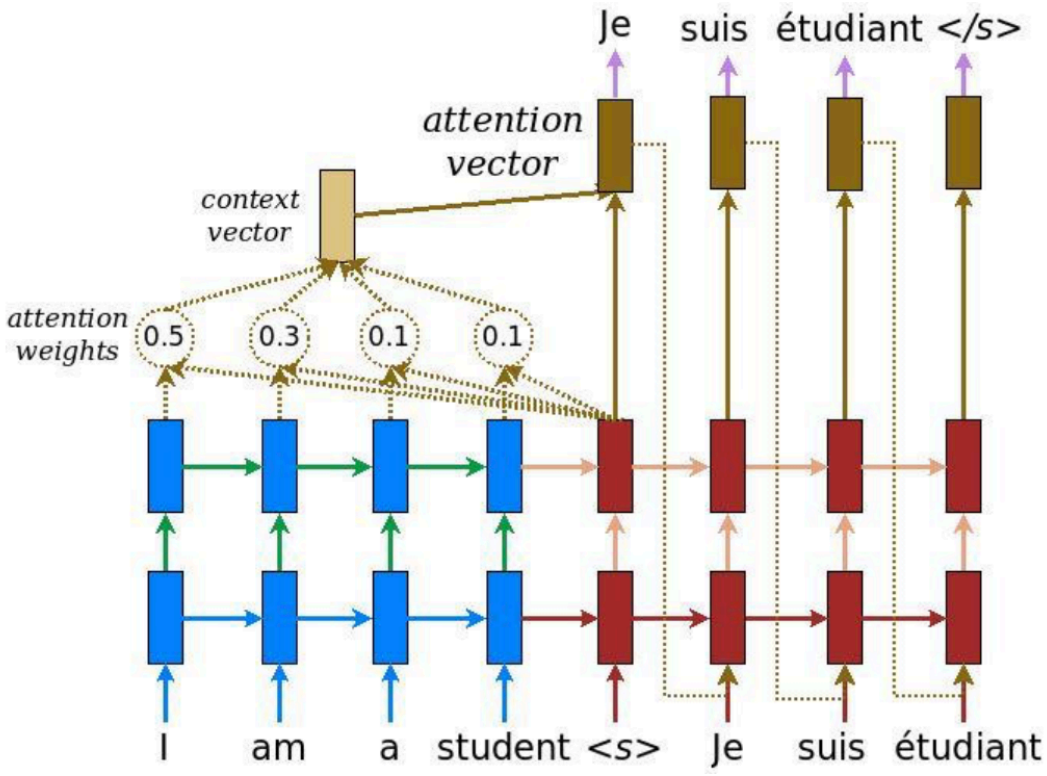
CHƯƠNG 3: GIẢI PHÁP ĐỀ TÀI

3.1 TỔNG QUAN GIẢI PHÁP KIẾN TRÚC MÔ HÌNH

Nhóm sinh viên sử dụng mô hình kiến trúc đầu cuối (end-to-end hay còn được gọi với Sequence to Sequence Model – seq2seq) có ý tưởng từ bài báo *Sequence to Sequence Learning with Neural Networks* do nhóm tác giả đến từ google được ông bố vào năm 2014 tại Silicon Valley AI Lab đã trình bày ý tưởng cụ thể để xây dựng một mô hình mạng nơ-ron hồi quy tối ưu với hướng đi mới so với các hệ thống dịch máy truyền thống kết hợp cùng với cơ chế chú ý (Attention mechanism) từ bài báo *Effective Approaches to Attention-based Neural Machine* được thực hiện bởi nhóm tác giả đến từ đại học Stanford vào năm 2015. Từ đó nhóm sinh viên xây dựng một mô hình mạng nơ-ron hồi quy (Recurrent neural network) dùng để xây dựng một hệ thống dịch máy tiên tiến.

Mô hình seq2seq hoạt động dựa 2 mạng RNN kết hợp lại với một mạng RNN nhận nhiệm vụ mã hoá (encoder) câu đầu vào tiếng Anh thành một vector biểu diễn câu đầu vào và một mạng RNN giải mã (decoder) có nhiệm vụ giải mã vector biểu diễn câu đầu vào và kết hợp với cơ chế chú ý (Attention mechanism) để giải mã câu đầu vào và cho ra kết quả câu đầu ra tiếng Việt tương ứng.

Kiến trúc tổng thể cho việc kết hợp 2 kiến trúc trên để xây dựng một mô hình dịch máy được minh hoạ cụ thể ở hình 3.1.



Hình 3.1: Torng quan kiến trúc mô hình dịch máy

3.2 GIẢI PHÁP BIỂU DIỄN TỪ

3.2.1 Tổng quan về giải pháp

Để có thể sử dụng các mô hình Deep Learning (học sâu) phục vụ cho việc dịch máy, chúng ta cần biểu diễn các từ thành các số vì các mô hình chỉ làm việc với dữ liệu số. Vì thế dựa trên các kết quả tìm kiếm và thực nghiệm [], nhóm sinh viên đề xuất sử dụng Word Embedding (nhúng từ) dùng để biểu diễn các từ thành các vector số thực. Mô hình mà nhóm chọn là Word2vec với mục đích biểu diễn các từ tiếng Anh và tiếng Việt thành các vector số thực n chiều bằng nhau (mỗi chiều là một giá trị số thực) để phục vụ cho quá trình huấn luyện.

Word2vec là một mô hình học không giám sát (model unsupervised learning) nó dùng để thể hiện mỗi quan hệ giữa các từ, nó được kết hợp từ hai thuật toán Skip-gram và Continuous bag of words (CBOW). Ở đây nhóm sinh viên đề xuất sử dụng mô hình skip-gram cho biểu diễn từ. Với skip-gram, kích thước biểu diễn từ giảm từ kích thước bằng số từ trong bộ từ vựng xuống bằng chiều dài lớp ẩn. Hơn nữa các vector có ý nghĩa nhiều hơn về mặt mô tả mối quan hệ giữa các từ. Chi tiết mô hình đã được mô tả ở Chương 2.

3.2.2 Chi tiết giải pháp

Nhóm sinh viên sử dụng đầu vào là tập dữ liệu được chia làm 2 tập tin chính chia làm 2 ngôn ngữ tiếng Anh và tiếng Việt.

Với bộ dữ liệu “IWSLT’15 English-Vietnamese data” với khoảng 100.000 câu song ngữ English-Vietnamese với dữ liệu thô chia làm hai tập tin tiếng English-Vietnamese như sau:

```
18  It &apos;s a huge amount of stuff . It &apos;s equal to the weight of methane .
19  And because it &apos;s so much stuff , it &apos;s really important for the atmospheric system .
20  Because it &apos;s important to the atmospheric system , we go to all lengths to study this thing .
21  We blow it up and look at the pieces .
22  This is the EUPHORE Smog Chamber in Spain .
23  Atmospheric explosions , or full combustion , takes about 15,000 times longer than what happens in your car .
24  But still , we look at the pieces .
25  We run enormous models on supercomputers ; this is what I happen to do .
26  Our models have hundreds of thousands of grid boxes calculating hundreds of variables each , on minute timescales .
27  And it takes weeks to perform our integrations .
28  And we perform dozens of integrations in order to understand what &apos;s happening .
29  We also fly all over the world looking for this thing .
30  I recently joined a field campaign in Malaysia . There are others .
```

```
18  Đó là một lượng khí thải khổng lồ , bằng tổng trọng lượng của mêtan .
19  Chính vì lượng khí thải rất lớn , nó có ý nghĩa quan trọng với hệ thống khí quyển .
20  Chính vì nó có ý nghĩa quan trọng với hệ thống khí quyển , giá nào chúng tôi cũng theo đuổi nghiên cứu này đến cùng .
21  Chúng tôi cho nó nổ và xem xét từng mảnh nhỏ .
22  Đây là Phòng nghiên cứu khói bụi EUPHORE ở Tây Ban Nha .
23  Nổ trong không khí hay cháy hoàn toàn diễn ra chậm hơn 15,000 lần so với những phản ứng trong động cơ xe .
24  Dù vậy , chúng tôi vẫn xem xét từng mảnh nhỏ .
25  Chúng tôi chạy những mô hình khổng lồ trên siêu máy tính ; đây là công việc của tôi .
26  Mô hình của chúng tôi gồm hàng trăm ngàn thùng xếp chồng tính toán với hàng trăm biến số trong thời gian cực ngắn .
27  Mà vẫn cần hàng tuần mới thực hiện xong các phép tích phân .
28  Chúng tôi cần làm hàng tá phép tính như thế để hiểu được những gì đang xảy ra .
29  Chúng tôi còn bay khắp thế giới để tìm phân tử này .
30  Gần đây tôi tham gia một cuộc khảo sát thực địa ở Malaysia . Còn nhiều chuyến khác nữa .
```

❖ Bước 1

Bước đầu tiên chúng ta thực hiện xử lí các câu dữ liệu như: xoá dấu “?”, “.”, xoá dấu khoảng trắng thừa và một số thứ khác.

Nhóm sinh viên thực hiện loại bỏ các câu có độ dài hơn 100.

❖ Bước 2

Nhóm sinh viên thực hiện tạo các từ điển word2int và int2word cho cả hai ngôn ngữ English-Vietnamese bằng cách sử dụng word_tokenize và ta được kết quả như sau:

```
The word2index:
 {'<pad>': 0, '<unk>': 1, '<s>': 2, '</s>': 3, '.': 4, ',': 5, 'tôi': 6, 'là': 7, 'và': 8, 'có': 9, 'một': 10

The int2word:
 {0: '<pad>', 1: '<unk>', 2: '<s>', 3: '</s>', 4: '.', 5: ',', 6: 'tôi', 7: 'là', 8: 'và', 9: 'có', 10: 'một'
```

❖ Bước 3

Tiếp theo, nhóm sinh viên thực hiện chuyển từng câu song ngữ sang từng vector với từng từ ứng với vị trí của từ trong từ điển.

Với câu đầu vào tiếng Anh ta thực hiện thêm (padding) với những câu có độ dài bé hơn 100.

Với câu đầu vào tiếng Việt ta sẽ thực hiện thêm (padding) trong quá trình huấn luyện vì ta sẽ sử dụng độ dài thật của câu song ngữ để huấn luyện nhanh hơn. Và ta được kết quả như sau:

```
icle:  the science behind a climate headline
icle(int seq):  [6, 310, 573, 13, 749, 4626, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,

le:  khoa học đằng sau một tiêu đề về khí hậu
le(int seq):  [326, 75, 1083, 116, 10, 372, 117, 41, 411, 743]
```

❖ Bước 4

Tại đây ta thực hiện lấy nhúng từ (word embedding) tất cả các từ có trong từ điển English-Vietnamese.

```
The first 5 word embedding:
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0.]
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0.]
[ 0.7667063  -0.09010636  0.08270185  1.25561469 -0.38717267  0.23894839
  0.07759737  0.12549625 -0.24102556  0.75143802  0.7187233  -1.04103127
 -0.47745857 -0.29726038 -1.06583469 -0.46455254  0.57115822 -0.32603807
  2.52081638 -0.95636765  1.01242033 -0.58393729  1.04310236  0.16477381
  1.39435824 -0.60122542 -1.3773635  -0.56987472 -1.54319531  1.09977543
  0.96689789 -0.1567848   0.34997361  0.96210557  0.48756319  1.05396065
 -0.51731518  1.63882088 -1.91224974 -0.09610153  0.65105395 -1.31423126
 -0.85510288 -0.81536141 -0.27825943 -0.94591077  1.1795865  -2.39665827
 -0.30680849 -0.46162846 -0.25772348  0.07643085 -1.32485165 -0.77122291
  1.23029739 -0.85100524  0.28650156  2.17254923  0.40640529 -0.57777529
  1.39834145  0.43437027 -1.27729737 -0.67731375 -1.65855753 -0.50178174
 -0.3156953  -0.09693184 -1.16374096  1.02022349  0.33924457  0.01924257
  1.23707084  1.55508455  0.03988794 -0.06625016  0.32820739 -1.35715819
  1.26497508  0.48284226 -1.11201119 -0.14896187  0.25508656 -1.76616043
 -1.72591984  1.51697642 -0.62773576 -0.34362389  0.21022845  0.22799106
 -0.18746373  0.80275874  1.51274783 -0.01218051 -0.85284633  0.11408914
 -0.34459046  1.72416776  2.19514605 -1.11158781]
```

Sau đó quá trình chuyển từ câu song ngữ (đã được chuyển thành vector số ứng với vị trí từng từ trong từ điển) thành nhúng từ sẽ được thực hiện trực tiếp trong mô hình.

3.3 GIẢI PHÁP XÂY DỰNG MÔ HÌNH DỊCH MÁY

3.3.1 Tổng quan về giải pháp

Dựa trên các đánh gia thực tế và điều kiện phần cứng lẫn lượng dữ liệu (data) cho phép, nhóm sinh viên lựa chọn phương pháp học sâu (deep

learning) để xây dựng mô hình mạng nơ-ron hồi quy (Recurrent neural netword) trong mô hình dịch máy (machine neural translation). Mô hình được đào tạo từ đầu đến cuối từ những câu đã được biểu diễn dưới các nhúng từ (word embedding) để tạo ra các chuỗi đầu vào bộ mã hoá (encoder) và bộ giải mã (decoder). Do đó với lượng dữ liệu đủ lớn và khả năng tính toán, mô hình có thể tự học một cách chính xác để thực hiện việc dịch một câu từ tiếng Anh sang tiếng Việt.

3.3.2 Mô hình mạng nơ-ron hồi quy (RNN) và khung huấn luyện

Cốt lõi của quá trình đào tạo một mô hình RNN là để nhận vào một văn bản tiếng Anh và tạo ra một văn bản tiếng Việt tương ứng. Để dễ hình dung ta có ví dụ một tập huấn luyện X = {($x^{(1)}$,$y^{(1)}$),($x^{(2)}$,$y^{(2)}$),…} với x là một vector các nhúng từ (word embedding) tương ứng với câu tiếng Anh đầu vào và y là một nhãn tức là một vector các nhúng từ (word embedding) tương tứng với câu tiếng Việt ở đầu ra. Mỗi câu tiếng Anh $x^{(i)}$ là một chuỗi thời gian có độ dài $T^{(i)}$, trong đó mỗi đoạn thời gian nhất định là một vector nhúng từ (word embedding) $x_t^{(i)}$, t = 1,2,…, $T^{(i)}$. Mục tiêu của RNN là chuyển đổi đầu vào x thành một chuỗi xác suất ký tự cho nhãn y, với $\hat{y}_t$ = P(w_t |x), trong đó w_t thuộc các từ trong từ điển tiếng Việt và một vài ký tự đặc biệt khác.

Mô hình RNN được nhóm sinh viên chọn sử dụng là mô hình hồi quy với 3 thành phần chính. Thành phần đầu tiên là lớp nhúng từ (embedding), lớp (layer) này có nhiệm vụ chuyển các đầu vào của bộ mã hoá (encoder) và

bộ giải mã (decoder) từ dạng int sang dạng nhúng từ (word embedding) để phục vụ cho công việc tính toán phía sau.

Thành phần thứ hai là bộ mã hoá (encoder), với bộ mã hoá chúng ta sử dụng Multi layer Bi-directional LSTM với số lượng layer và số lượng hidden units của LSTM cell được thiết lập trong param. Ngoài ra nhóm sinh viên còn sử dụng DropoutWrapper để thiết lập giá trị Drop Out cho các LSTM cell để tránh hiện tượng quá khớp (over-fitting) với dữ liệu huấn luyện.

Thành phần thứ ba là bộ giải mã (decoder). Đối với bộ giải mã, chúng em chia thành hai trường hợp riêng biệt là huấn luyện mô hình(training) và dự đoán (inference). Trong quá trình huấn luyện chúng em sử dụng TrainingHelper còn khi dự đoán, chúng em sử dụng BasicDecoder với BeamSearchDecoder.

❖ Huấn luyện: chúng em sử dụng BahdanauAttention và TrainingHelper để huấn luyện mô hình. Chúng em còn sử dụng AdamOptimizer để cập nhật tham số cho mô hình và còn sử dụng Gradient Clipping để tránh mô hình bị bùng nổ độ dốc (exploding gradients).

❖ Dự đoán: sau khi huấn luyện xong mô hình và sử dụng mô hình này để dự đoán kết quả. Tuy nhiên do chúng ta không biết kết quả thực tế như trong quá trình huấn luyện, nên ta cần sử dụng các thuật toàn tìm kiếm để cho ra kết quả phù hợp nhất và chúng em chọn sử dụng thuật toán tìm kiếm chùm tia (Beam Search) với beam-width = 10.

3.5 GIẢI PHÁP XÂY DỰNG MÁY CHỦ

Máy chủ (server) được nhóm sinh viên chọn Amazon EC2 làm máy chủ với mục đích tạo ra một cầu nối giữa mô hình đã được huấn luyện (model) và phía ứng dụng sản phẩm (client) – được xây dựng với React Native. Vì vậy trong giới hạn của khoá luận, máy chủ chỉ cung cấp duy nhất một giao diện lập trình (API) với chức năng chuyển đổi từ một văn bản (text) tiếng Anh thành một văn bản (text) tiếng Việt tương ứng.

3.6 GIẢI PHÁP XÂY DỰNG ỨNG DỤNG

Để ng dụng hoá hệ thống dịch máy từ tiếng Anh sang tiếng Việt, nhóm sinh viên quyết định xây dựng web để ứng dụng kết quả của hệ thống vào một tình huống cụ thể có thể ứng dụng và thương mại hoá tốt.

Ứng dụng web do nhóm sinh viên xây dựng có chức năng chính là chuyển đổi văn bản tiếng Anh do người dụng nhập vào và đưa ra văn bản tiếng Việt tương ứng .

3.6.1 Thiết kế giao diện ứng dụng

Giao diện ứng dụng chỉ có một màn hình với chức năng chính là chuyển đổi một văn bản tiếng Anh thành một văn bản tiếng Việt tương ứng.

Hình 3.2 Màn hình chính của ứng dụng

Để sử dụng, người dùng nhập văn bản tiếng Anh vào ô tiếng Anh tương ứng và nhập vào nút dịch. Kết quả sẽ được hiển thị tại ô tiếng Việt.

3.6.2 Thiết kế kiến trúc ứng dụng

3.7 TỔNG KẾT

Thông qua chương 3, sinh viên đã làm rõ được các giải pháp cụ thể cho từng phần trong hệ thống dịch máy từ tiếng Anh sang tiếng Việt, hướng xây dựng máy chủ và cả ứng dụng trên nền tảng web.

Nhóm sinh viên đã trình bày một hệ thống dịch máy từ tiếng Anh sang tiếng Việt dựa trên việc học sâu (deep learning) từ đầu đến cuối có khả năng vượt trội và hiện đại trong hiện đại. Nhóm sinh viên tin rằng phương pháp này sẽ tiếp tục được cải thiện với các mô hình mới hơn, đơn giải hoặc phức tạp hơn khi tận dụng được sức mạnh tính toán phần cứng và kích thước dữ liệu được tăng thêm trong tương lai.

Chương kế tiếp nhóm sinh viên sẽ trình bày về các thư viện, công cụ và những khó khăn cụ thể nếu có cho các giải pháp đã trình bày ở chương này.