1. Token hoá và phân loại văn bản (Tokenization and Text Classification)

Token hoá (Tokenization) là quá trình biến mọi thứ thành tài sản kỹ thuật số, bao gồm chia nhỏ các từ thành từng phần (hoặc mã token) mà máy móc có thể hiểu được. Các tài liệu, văn bản bằng tiếng Anh rất dễ token hóa vì chúng có khoảng cách rõ ràng (clear spaces) giữa các từ và đoạn văn. Tuy nhiên, hầu hết các ngôn ngữ khác đều là những thử thách hoàn toàn mới. Ví dụ, các ngôn ngữ logic như tiếng Quảng Đông, tiếng Quan Thoại và tiếng Kanji của Nhật Bản có thể là những thách thức, khó khăn vì chúng không có khoảng cách giữa các từ hoặc thậm chí là câu.

Nhưng tất cả các ngôn ngữ tuân theo các quy tắc và mẫu nhất định. Thông qua học tập sâu, chúng ta có thể đào tạo các mô hình để thực hiện token hoá. Do đó, hầu hết các khóa học về AI và DL khuyến khích các chuyên gia DL thử nghiệm các mô hình đào tạo DL để xác định và hiểu các mẫu và văn bản này.

Ngoài ra, các mô hình DL có thể phân loại và dự đoán chủ đề của văn bản, đoạn văn đó đang hướng tới. Ví dụ, mạng nơ ron tích chập CNN (Convolutional Neural Networks) và mạng nơ ron hồi quy RNN (Recurrent Neural Networks) có thể tự động phân loại tone và sắc thái của văn bản bằng cách sử dụng word embeddings giúp tìm ra mô hình không gian vector cho các từ. Hầu hết các nền tảng social media đều triển khai các hệ thống phân tích dựa trên CNN và RNN để gắn cờ và xác định nội dung spam trên nền tảng của họ. Phân loại văn bản cũng được áp dụng trong tìm kiếm trên web, nhận dạng ngôn ngữ và đánh giá khả năng đọc.

2. Tạo chú thích cho hình ảnh (Generating Captions for Images)

Tự động mô tả nội dung của một hình ảnh bằng cách sử dụng các câu tự nhiên là một nhiệm vụ đầy thách thức. Chú thích của hình ảnh không chỉ để nhận ra các đối tượng trong ảnh mà còn thể hiện cách chúng có liên quan tới nhau hay không cùng với các thuộc tính của chúng (mô hình nhận dạng hình ảnh). Ngoài mô hình nhận dạng ảnh ra, kiến thức ngữ nghĩa phải được thể hiện bằng ngôn ngữ tự nhiên cũng đòi hỏi một mô hình ngôn ngữ nữa.

Căn chỉnh các yếu tố hình ảnh và ngữ nghĩa là cốt lõi để tạo chú thích hình ảnh một cách hoàn hảo. Các mô hình DL có thể giúp tự động mô tả nội dung của hình ảnh bằng cách sử dụng các câu tiếng Anh chính xác. Điều này có thể giúp những người khiếm thị dễ dàng truy cập nội dung trực tuyến.

Trình tạo chú thích hình ảnh nơ ron của Google (NIC) dựa trên một network bao gồm CNN (ứng dụng trong Computer Vision), theo sao là RNN (ứng dụng trong xử lý ngôn ngữ) . Mô hình này sẽ tự động xem hình ảnh và tạo bản mô tả bằng tiếng Anh.

3. Nhận dạng giọng nói (Speech Recognition)

DL đang ngày càng được sử dụng để xây dựng và huấn luyện các mạng thần kinh để chuyển dịch các đầu vào là âm thanh (audio inputs) và thực hiện các nhiệm vụ nhận dạng và tách từ vựng phức tạp. Trên thực tế, các mô hình và phương pháp này được sử dụng trong xử lý tín hiệu, ngữ âm và nhận dạng từ, các lĩnh vực cốt lõi của nhận dạng giọng nói.

Ví dụ, các mô hình DL có thể được đào tạo để xác định từng giọng nói cho người nói tương ứng và trả lời riêng từng người nói. Hơn nữa, các hệ thống nhận dạng giọng nói dựa trên CNN có thể dịch lời nói “raw” thành tin nhắn văn bản, điều này giúo cung cấp những insight thú vị liên quan đến người nói.

4. Dịch máy (Machine Translation)

Dịch máy (MT) là một nhiệm vụ cốt lõi trong xử lý ngôn ngữ tự nhiên, có thể điều tra việc sử dụng máy tính để dịch ngôn ngữ mà không cần sự can thiệp của con người. Gần đây, chỉ có các mô hình học sâu mới được sử dụng cho dịch máy bằng nơ ron. Không giống như dịch máy truyền thống, các mạng nơ ron sâu (DNN) cung cấp bản dịch chính xác và hiệu suất tốt hơn. Mạng nơ ron tích chập (RNN), mạng nơ ron truyền ngược (FNN), bộ mã hóa tự động đệ quy (RAE) và bộ nhớ dài ngắn hạn (LSTM) được sử dụng để huấn luyện máy chuyển đổi câu từ ngôn ngữ gốc của văn bản đó sang ngôn ngữ muốn chuyển đổi một cách chính xác.

Các giải pháp DNN phù hợp được sử dụng cho các quy trình, chẳng hạn như căn chỉnh từ, quy tắc sắp xếp lại câu, xây dựng mô hình ngôn ngữ và tham gia dự đoán dịch để dịch câu mà không cần sử dụng một lượng lớn cơ sở dữ liệu các nguyên tắc.

5. Trả lời câu hỏi (Question Answering)

Các hệ thống QA này cố gắng trả lời một thắc mắc được đặt dưới dạng câu hỏi. Vì vậy, câu hỏi định nghĩa, câu hỏi tiểu sử và câu hỏi đa ngôn ngữ trong số các loại câu hỏi khác mà được hỏi bằng ngôn ngữ tự nhiên thì được trả lời bởi các hệ thống như vậy.

Tạo ra một hệ thống trả lời câu hỏi đầy đủ chức năng là một trong những thách thức phổ biến mà các nhà nghiên cứu phải đối mặt trong phân khúc DL. Mặc dù các thuật toán học sâu đã đạt được tiến bộ đáng kể trong phân loại văn bản và hình ảnh trong quá khứ, nhưng không thể giải quyết các tác vụ liên quan đến lý luận logic (như vấn đề trả lời câu hỏi). Tuy nhiên, trong thời gian gần đây, các mô hình học sâu đang cải thiện hiệu suất và độ chính xác của các hệ thống QA này.

Ví dụ, các mô hình mạng nơ ron tích chập có thể trả lời một cách chính xác các câu hỏi dài trong khi đó các cách tiếp cận truyền thống hồi xưa đã từng thất bại. Quan trọng hơn, mô hình DL được đào tạo theo cách mà không cần phải xây dựng hệ thống bằng kiến ​​thức ngôn ngữ như tạo một trình phân tách ngữ nghĩa

6. Tóm tắt văn bản (Document Summarization)

Việc tóm tắt văn bản đang đóng vai trò cực kỳ quan trọng khi ngày nay càng khối lượng dữ liệu (data) ngày càng gia tăng. Những tiến bộ mới nhất trong các mô hình sequence-to-sequence đã giúp các chuyên gia DL dễ dàng phát triển các mô hình tóm tắt văn bản tốt hơn. Hai loại tóm tắt văn bản cụ thể là

Tóm tắt rút trích (Extract): là một bản tóm tắt bao gồm các nội dung được rút trích từ văn bản gốc.

Tóm tắt tóm lược (Abstract): là một bản tóm tắt có chứa các nội dung không được thể hiện trong văn bản gốc.

Trong mô hình sequence-to-sequence (seq2seq), kỹ thuật attention là một kỹ thuật cho phép có thể học hiệu quả mô hình sinh cho những chuỗi có độ dài lớn và đã dành được sự quan tâm lớn của cộng đồng nghiên cứu.. Tham khảo sơ đồ bên dưới từ blog Pointer Generator của Abigail See.

Mô hình seq2seq cơ bản bao gồm hai mạng neural thành phần được gọi là mạng mã hóa (encoder) RNN và mạng giải mã (decoder) RNN để sinh ra chuỗi đầu ra t1:m từ một chuỗi đầu vào x1:n. Mạng neural encoder mã hóa chuỗi đầu vào thành một vector c có độ dài cố định. Mạng neural decoder sẽ lần lượt sinh từng từ trong chuỗi đầu ra dựa trên vector c và những từ được dự đoán trước đó cho tới khi gặp từ kết thúc câu. Bởi vì mạng giải mã decoder có thể tự do tạo ra các từ theo bất kỳ thứ tự nào, nên mô hình seq2seq là một giải pháp vững mạnh cho tóm tắt tóm lược văn bản.