

Nhận diện và trích xuất cảm xúc trong đoạn văn bản Tiếng Việt

Quy Hoang Nguyen, Van T. M. Nguyen, Tien Minh Nguyen, Khanh P. B. Truong, Kiet Van Nguyen
*Khoa Khoa Học và Kỹ Thuật Thông Tin, Trường Đại Học Công Nghệ Thông Tin,
 Đại Học Quốc Gia thành phố Hồ Chí Minh, Việt Nam*

Tóm tắt nội dung—Phân tích cảm xúc dựa trên khía cạnh là một tác vụ phức tạp đóng vai trò quan trọng trong xử lý ngôn ngữ tự nhiên (NLP). Tuy nhiên, các nghiên cứu gần đây chỉ tập trung vào việc phát hiện khía cạnh và phân loại cảm xúc mà bỏ qua công việc phụ của việc phát hiện ý kiến người dùng, mà có tiềm năng lớn trong các ứng dụng thực tế. Trong bài báo này, chúng tôi sẽ thực hiện trên bộ dữ liệu UIT-VISD4SA. Chúng tôi đề xuất sử dụng mô hình PhoBERT để trích xuất ngữ cảnh và kết hợp với mô hình BiLSTM-CRF cho tác vụ nhận diện và trích xuất cảm xúc. Mô hình PhoBERTv2 sử dụng tách từ đạt F1score cao nhất là 51%.

Index Terms—Phân tích cảm xúc dựa trên khía cạnh, xử lý ngôn ngữ tự nhiên, phát hiện ý kiến người dùng, BiLSTM-CRF, XLM-RoBERTa, PhoBERT

1 GIỚI THIỆU

Trước khi mua một sản phẩm hoặc quyết định sử dụng một dịch vụ, người dùng thường tìm kiếm lời khuyên từ những người đã sử dụng sản phẩm hoặc dịch vụ trước đó. Với sự bùng nổ của thương mại điện tử, ngày càng có nhiều người tìm kiếm những đánh giá từ các trang web, cửa hàng trực tuyến, diễn đàn hoặc kênh đánh giá sản phẩm. Do đó, số lượng đánh giá đang tăng lên và trở thành một nguồn tài nguyên quý giá cho khách hàng và doanh nghiệp. Đối với khách hàng, nguồn dữ liệu này cung cấp thông tin về sản phẩm và lời khuyên hữu ích giúp họ tránh việc mua sản phẩm hoặc sử dụng dịch vụ không phù hợp với nhu cầu cá nhân. Trên mặt khác, đánh giá của người dùng cũng là thông tin quý giá cho doanh nghiệp, và nếu được sử dụng một cách chính xác và hiệu quả, dữ liệu này có thể giúp doanh nghiệp cải thiện chất lượng sản phẩm, xác định đúng khách hàng mục tiêu cho từng đối tượng.

Phân tích cảm xúc (SA) là một trong những tác vụ quan trọng và được chú trọng nghiên cứu trong lĩnh vực xử lý ngôn ngữ tự nhiên. Tầm quan trọng của nó không chỉ được chú ý trong giới học thuật mà còn trong nhiều ngành công nghiệp khác. Ví dụ như phân tích cảm xúc từ bình luận của khách hàng trên các diễn đàn thương mại điện tử có thể giúp doanh nghiệp biết được sản phẩm, dịch vụ này có tốt hay không và sẽ đưa ra các chiến dịch tương ứng. Tuy nhiên phân tích cảm xúc thông thường chỉ đưa ra một kết quả tương ứng với mỗi đoạn bình luận, điều này đôi khi không phù hợp trong thực tế. Trong nhiều trường hợp ta cần nhận diện được cảm xúc trên các khía cạnh chi tiết hơn. Do đó phân tích cảm xúc trên khía cạnh (ABSA) bắt đầu nhận được nhiều sự chú ý.

Phân tích cảm xúc dựa trên khía cạnh (ABSA) nhằm mục đích phân tích những câu, đoạn văn có nhiều khía cạnh khác nhau. ABSA sẽ nhận diện các khía cạnh được đề cập tới sau đó sẽ phân tích cảm xúc của từng khía cạnh đã được nhận diện. Trong bài báo này, chúng tôi sẽ tập trung tiến hành

nhận diện các khoảng cho ABSA (span detection). Cụ thể chúng tôi sẽ xác định các khoảng hay có thể gọi là vị trí mà các khía cạnh đó được đề cập. Ví dụ tới một bình luận như sau “Hiệu năng máy khá tốt nhưng pin rất mau hết”, nhận diện khoảng cho ABSA sẽ xác định được “hiệu năng máy khá tốt”, “pin rất mau hết” và sau đó nhận diện khía cạnh cũng như cảm xúc của chúng.

Trong bài báo này, chúng tôi sẽ thực hiện nhận diện và trích xuất cảm xúc được đề cập sử dụng mô hình PhoBERT cùng với mô hình BiLSTM-CRF.

2 CÔNG TRÌNH LIÊN QUAN

So với các ngôn ngữ giàu tài nguyên như tiếng Anh, số lượng bộ dữ liệu tiếng Việt vẫn còn khá hạn chế. Mai [?] và các cộng sự đã thu thập các bài đánh giá về điện thoại và gán nhãn chỉ bao gồm 2.098 câu cho hai tác vụ: OTE và phân loại cảm xúc. Đồng thời, họ cũng giới thiệu một mô hình đa tác vụ sử dụng Bi-LSTM và CRF. Cộng đồng VLSP đã cung cấp một bộ dữ liệu bao gồm các đánh giá về khách sạn và nhà hàng được thu thập từ nền tảng đánh giá trực tuyến. Tuy nhiên, bộ dữ liệu từ VLSP chỉ gán nhãn cho khía cạnh ASPECT#ENTITY và cảm xúc tương ứng của nó mà bỏ qua tác vụ OTE. Nguyen [?] và các cộng sự đã giới thiệu một bộ dữ liệu giống như bộ dữ liệu VLSP, bao gồm 7.828 đánh giá tài liệu với bảy loại khía cạnh và năm loại cảm xúc, với nhiệm vụ giống như công việc trước: trích xuất khía cạnh và phân loại cảm xúc. Dang [?] và các cộng sự cũng đã xây dựng một bộ dữ liệu ở cấp độ câu cùng miền như hai công trình trước nhưng có độ đồng thuận cao giữa các người gán nhãn.

Để nhận diện và trích xuất cảm xúc từ đoạn văn, Hu [?] và các cộng sự đã đề xuất mô hình có thể trích xuất sau đó nhận diện cảm xúc. Mô hình này sẽ trích xuất trực tiếp các ý kiến của người dùng, tiếp theo sẽ phân loại cảm xúc dựa vào ý kiến đó. Mô hình được xây dựng dựa trên những tiền bộ của đọc hiểu văn bản và hỏi đáp tự động, hai mô hình

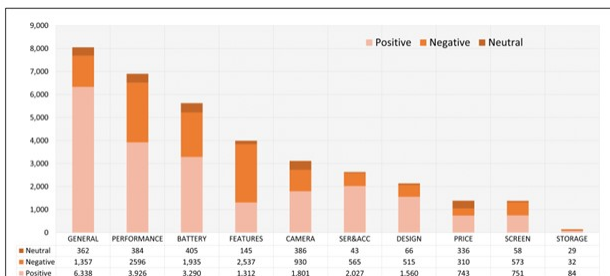
này sẽ trích xuất trực tiếp các câu từ đoạn văn để trả lời cho câu hỏi được đề cập. Xu [?] và các cộng sự đã chứng minh được tính hiệu quả của mô hình trong việc sử dụng CRF để trích xuất ý kiến dựa trên khía cạnh. Kim [?] và các cộng sự cũng đã chứng minh được sự hiệu quả trong việc sử dụng mô hình BiLSTM kết hợp với CRF. Tuy nhiên, mô hình này có dữ liệu đầu vào chưa tách từ và chưa sử dụng mô hình pretrained cho Tiếng Việt là PhoBERT.

3 BỘ DỮ LIỆU

Bộ dữ liệu được chúng tôi sử dụng là UIT-ViSD4SA. Đây là một bộ dữ liệu gồm các bình luận, phản hồi của người dùng về sản phẩm điện thoại thông minh từ một website thương mại điện tử ở Việt Nam.

Bộ dữ liệu bao gồm 35396 span trên tổng số 11122 bình luận của người dùng. Mỗi span đều được gán nhãn khía cạnh và cảm xúc tương ứng của nó.

Hình 1 cho thấy phân phối của các nhãn khía cạnh trong bộ dữ liệu. Nhãn GENERAL chiếm số lượng nhiều nhất cho thấy thiên hướng bình luận của người dùng thiên về tổng quan của điện thoại. Các khía cạnh phổ biến khác như PERFORMANCE, BATTERY, FEATURES, và CAMERA cũng nhận được sự quan tâm trong khi STORAGE chiếm số lượng rất ít. Trong nội bộ của từng khía cạnh, ngoại trừ nhãn FEATURES thì còn lại đều có đặc điểm chung là nhãn Positive chiếm số lượng lớn nhất, kể đến là Negative và thấp nhất là Neutral.



Hình 1: Phân phối của 10 nhãn khía cạnh và cảm xúc của chúng

4 PHƯƠNG PHÁP THỰC HIỆN

4.1 Tiền xử lý

Trong xử lý ngôn ngữ tự nhiên, tiền xử lý dữ liệu là một quá trình không thể thiếu. Đặc biệt là trong ngôn ngữ như tiếng Việt do đôi khi cùng một từ nhưng có nhiều cách viết khác nhau, đôi khi do người dùng viết sai chính tả, viết tắt.

Nếu tiền xử lý tốt có thể cải thiện đáng kể hiệu suất của mô hình. Đối với bài toán này, kết quả của quá trình tiền xử lý phải đảm bảo đúng offset đối với từng nhãn từ loại, do đó quá trình tiền xử lý của chúng tôi được mô tả ở hình 2.

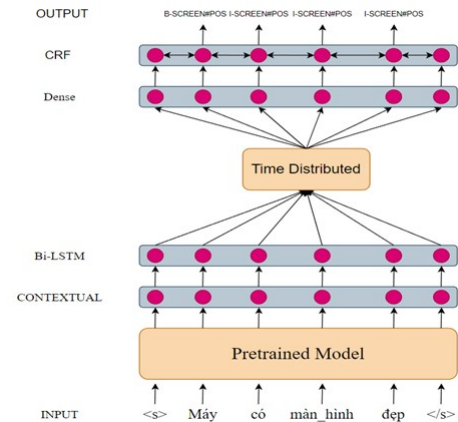


Hình 2: Các bước tiền xử lý dữ liệu

Đầu tiên, chúng tôi sẽ tiến hành kiểm tra xem từ đó có phải từ Tiếng Việt hay không. Nếu đó là từ Tiếng Việt thì chúng tôi sẽ chuẩn hóa các thanh điệu về dạng chuẩn theo các quy tắc về nguyên âm và phụ âm của Tiếng Việt (vd: “máy” => “máý”). Tiếp theo sẽ chuyển các câu bình luận sang kiểu viết thường, rút gọn các từ kéo dài (“xìnnnn” => “xìn”), sau đó sẽ chuẩn hóa các từ viết tắt thông dụng. Cuối cùng, chúng tôi sẽ thực hiện tách từ để làm đầu vào cho các mô hình Pretrained trên Tiếng Việt. Input và output của bước tiền xử lý được mô tả ở bảng 2

4.2 Hướng tiếp cận

Bài toán này có hướng giải quyết giống tác vụ gán nhãn từ loại (Sequence Labeling) ở mức độ âm tiết (syllable). Để thực hiện bài toán, chúng tôi xây dựng nên mô hình với sự kết hợp giữa pretrained model và các mạng truyền thống. Mô hình này gồm 4 Layers: Contextual Layer là output của mô hình Pretrained giúp mô hình có thể hiểu được vector nghĩa của từ trong câu, sau đó sẽ được chuyển qua các layer còn lại được mô tả ở hình 3.



Hình 3: Tổng quan về mô hình sử dụng

4.2.1 Contextual Embedding

Vector biểu diễn từ sử dụng Contextual Embedding có thể biểu diễn được cú pháp, ngữ nghĩa của một từ trong ngữ cảnh. Chúng tôi sẽ trích xuất vector biểu diễn từ của mô hình các mô hình pretrained để thực hiện bài toán.

4.2.2 Bidirectional Long Short-Term Memory Long Short-Term Memory (LSTM)

Đây là một kiến trúc đặc biệt của mạng RNN có khả năng học được sự phụ thuộc dài hạn. LSTM đã khắc phục các vấn đề của mạng RNN trước đây là Vanishing Gradient và Exploding Gradient. Tuy nhiên LSTM có cấu trúc phức tạp hơn mặc dù vẫn giữ được ý tưởng chính của RNN là sao chép các kiến trúc theo dạng chuỗi.

Dù cho LSTM có thể giải quyết bài toán phụ thuộc dài hạn, nó vẫn mất một vài thông tin ngữ nghĩa trong quá trình truyền đi. Do đó, việc sử dụng mô hình Bi-LSTM là cần thiết vì mô hình này có thể biểu diễn 2 chiều thông tin ngữ cảnh của từ.

Bảng 1: Thống kê tổng quan của bộ dữ liệu UIT-ViSD4SA

Set	Comment	Average aspect per comment	Average span length	Positive	Negative	Neutral	Total span
Train	7,784	3.2	32.6	15,356	7,793	1,560	35,396
Dev	1,113	3.1	32.4	2,110	1,144	241	
Test	2,225	3.2	32.5	4,266	2,269	413	

Bảng 2: Ví dụ về tiền xử lí

Input	Output
Mua được 1 tuần thấy máy quá OK Pin trâu , máy mượt , cảm biến vân tay nhạy , k nóng máy . Nhu cầu chơi game nên k cần gì hơn , liên quân vô nhanh gần ngang ip . Vs mức giá này mua con máy sài thấy quá hời luôn	mua được 1 tuần thấy máy quá ok pin trâu , máy mượt , cảm biến vân tay nhạy , k nóng máy . nhu cầu chơi game nên k cần gì hơn , liên quân vô nhanh gần ngang ip . vs mức giá này mua con máy sài thấy quá hời luôn

4.2.3 Conditional Random Field Conditional Random Field (CRF)

CRF là một mô hình trình tự có được tất cả các ưu điểm của mô hình (MEMMS) đồng thời giải quyết được vấn đề sai lệch nhãn (label bias) bằng việc sử dụng Global Normalizer. Không giống như LSTM và Bi-LSTM nơi các cell state/recurrent được sử dụng, mô hình CRF kết nối trực tiếp đầu vào và đầu ra. Cho một tập dữ liệu huấn luyện $D = (x^1, y^1), \dots, (x^N, y^N)$ của N chuỗi dữ liệu có nhãn dự đoán là x^i và nhãn thực tương ứng là y^i , CRF tối đa hóa log-likelihood xác suất có điều kiện của chuỗi nhãn dựa trên chuỗi dữ liệu như sau:

$$L = \sum_{i=1}^N \log(P(y^i|x^i)) - \sum_{k=1}^K \frac{\lambda_k}{2\sigma^2} \quad (1)$$

5 KẾT QUẢ THỰC NGHIỆM

5.1 Thiết kế

Chúng tôi thực hiện nhận diện và trích xuất cảm xúc trong đoạn văn dựa theo định dạng của IOB (inside, outside, beginning).

Chúng tôi sẽ nối vector ngữ cảnh của câu tại 2 layer cuối cùng từ các mô hình pretrained là XLM-RoBERTa, PhoBERTv1 và PhoBERTv2. Sau đó sẽ triển khai lớp Bi-LSTM với số Hidden Layers bằng với số chiều của vector ngữ cảnh, tỉ lệ Dropout là 0.1. Chúng tôi sẽ so sánh kết quả thử nghiệm với các mô hình pretrained để chọn ra mô hình phù hợp với bài toán.

Để đánh giá mô hình, chúng tôi sử dụng 3 thang đo là: Precision, Recall, F1-score. Nhận được dự đoán chính xác khi và chỉ khi nhãn này đúng bằng nhãn thực.

5.2 Kết quả

Bảng 3 cho thấy kết quả so sánh của các mô hình pretrained đối với tác vụ phát hiện Aspect#Polarity. Đồng thời chúng tôi cũng so sánh với mô hình Base line của tác giả trước đó là BiLSTM-CRF.

Dựa vào kết quả từ bảng 3 có thể thấy việc tách từ và sử dụng các mô hình pretrained trên Tiếng Việt có kết quả tốt hơn hẳn. Cụ thể là, mô hình PhoBERT và PhoBERTv2

đều có kết quả tốt hơn. Mô hình PhoBERTv2 thu được kết quả cao nhất trên cả 3 thang đo Precision, Recall, F1-score lần lượt là 45.9%, 57.7%, 51.1% đối với tác vụ trích xuất ASPECT#POLARITY. Trái lại, mô hình không sử dụng tách từ là BiLSTM-CRF có kết quả khá thấp, F1-score chỉ đạt được 46%.

Kết quả chi tiết của từng ASPECT#POLARITY của mô hình PhoBERTv2 được thể hiện ở bảng 4. Trong tác vụ nhận diện khía cạnh cảm xúc, F1-scores có kết quả không đồng đều nhau. Đối với nhãn cảm xúc Positive, F1-score cao nhất là 79% thuộc về CAMERA, thấp nhất là 9% thuộc về PRICE. Ngược lại, nhãn cảm xúc NEGATIVE và NEUTRAL có kết quả khá thấp, F1-score cao nhất lần lượt là 56% và 42% đều thuộc về nhãn CAMERA và thấp nhất đều là 0%. Nguyên nhân của sự chênh lệch này là do sự bất cân bằng nhãn trong bộ dữ liệu.

Bảng 3: Kết quả của từng mô hình

	Precision	Recall	F1-score
BiLSTM-CRF	0.468	0.454	0.457
PhoBERTv1	0.413	0.54	0.468
PhoBERTv2	0.459	0.577	0.511

Bảng 4: Kết quả chi tiết từng ASPECT#POLARITY của mô hình PhoBERTv2

Aspect	NEGATIVE	NEUTRAL	POSITIVE
BATTERY	0.54	0.07	0.73
CAMERA	0.56	0.42	0.79
DESIGN	0.16	0.06	0.67
FEATURES	0.41	0.24	0.65
GENERAL	0.13	0.27	0.56
PERFORMANCE	0.32	0.22	0.65
PRICE	0.34	0.10	0.09
SCREEN	0.45	0.00	0.71
SER&ACC	0.12	0.14	0.64
STORAGE	0.00	0.14	0.56

5.3 Phân tích lỗi

Trong quá trình thử nghiệm, chúng tôi nhận thấy mô hình gồm 3 lỗi chính sau: nhận diện sai khía cạnh cảm xúc và nhận diện sai khoảng ý kiến. Các lỗi sai chủ yếu được mô tả trong bảng 5.

Mô hình PhoBERTv2 được chúng tôi huấn luyện chưa bao quát hết được ý của câu. Ở trường hợp Nhận diện sai khoảng, nhãn dự đoán của mô hình cũng có ý nghĩa đúng nhưng chưa bao hàm đủ ý. Đối với trường hợp 2, mô hình gặp khó khăn đối với các câu so sánh, để hiểu rõ được pin 'điện thoại j7pro' thì yêu cầu mô hình phải nắm bắt được thông tin về chính điện thoại đó.

Bảng 5: Các lỗi chủ yếu của mô hình

	Actual	Predict
Nhận diện sai khoảng	sao cam bị hột quá sao thời gian sử dụng lúc mua về bt sử dụng dc khoảng 1tuần camre bị vậy CAMERA#NEGATIVE	sao cam bị hột quá CAMERA#NEGATIVE sao thời gian sử dụng lúc mua về bt sử dụng dc khoảng 1tuần camre bị vậy
Nhận diện sai nhân cảm xúc	mình mới mua thấy sai cũng tốt GENERAL#POSITIVE pin thì dùng chơi game liên_tục cũng được 9 h.nói chung pin k bang con j7pro. BATTERY#NEGATIVE	mình mới mua thấy sai cũng tốt GENERAL#POSITIVE pin thì dùng chơi game liên_tục cũng được 9 h.nói chung pin k bang con j7pro. BATTERY#POSITIVE

Có sự xuất hiện của các lỗi sai được đề cập ở trên có thể là do dữ liệu huấn luyện chưa được nhiều, các nhân bị mất cân bằng với nhau và mô hình chưa hiểu rõ ngữ nghĩa của câu.

6 KẾT LUẬN

Trong bài toán này chúng tôi đã tiến hành trích xuất ý kiến của người dùng trong câu bình luận. Chúng tôi đã xây dựng mô hình pretrained với sự kết hợp của Bi-LSTM-CRF. Kết quả thử nghiệm cho thấy việc sử dụng tách từ cùng với mô hình PhoBERTv2 cho kết quả dự đoán tốt nhất trên với F1-score là 51%. Qua kết quả trên, chúng tôi thấy kết quả mô hình còn khá thấp. Nguyên nhân là do nhân trong bộ dữ liệu không được cân bằng, dữ liệu còn hạn chế. Trong tương lai, chúng tôi sẽ áp dụng thêm các kỹ thuật tiền xử lý mới để tiếp tục thực hiện bài toán này.

TÀI LIỆU

- [1] Mai, L., Le, B: Aspect-based sentiment analysis of vietnamese texts with deep learning. In: Asian Conference on Intelligent Information and Database Systems. pp. 149–158. Springer (2018)
- [2] Nguyen, N.T.H., Phan, P.H.D., Nguyen, L.T., Van Nguyen, K., Nguyen, N.L.T: Vietnamese open-domain complaint detection in e-commerce websites. In: arXiv preprint arXiv:2104.11969 (2021)
- [3] Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S., Androutsopoulos, I.: Semeval-2015 task 12: Aspect based sentiment analysis. In: Proceedings of SemEval 2015
- [4] Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., Manandhar, S.: S.: SemEval-2014 task 4: Aspect based sentiment analysis. In: Proceedings of SemEval 2014. pp. 27–35. Association for Computational Linguistics, Dublin, Ireland (2014)
- [5] Van Huynh, T., Nguyen, V.D., Van Nguyen, K., Nguyen, N.L.T., Nguyen, A.G.T.: Hate speech detection on vietnamese social media text using the bi-gru-lstm-cnn model. In: arXiv preprint arXiv:1911.03644 (2019)
- [6] Luong Luc Phan, Phuc Huynh Pham, Kim Thi-Thanh Nguyen: SA2SL: From Aspect-based Sentiment Analysis to Social Listening System for Business Intelligence
- [7] Minghao Hu, Yuxing Peng, Zhen Huang, Dongsheng Li, Yiwei Lv: Open-domain Targeted Sentiment Analysis via Span-based Extraction and Classification (2019)
- [8] Hu Xu, Bing Liu, Lei Shu, S Yu Philip: Double embeddings and CNN-based sequence labeling for aspect extraction (2018)
- [9] Wenxuan Zhang, Xin Li, Yang Deng, Lidong Bing, Wai La: A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges (2022)
- [10] Gustavo Aguilar, Suraj Maharjan, Adrian Pastor López-Monroy, Tamar Solorio: A Multi-task Approach for Named Entity Recognition in Social Media Data. In: Proceedings of the 3rd Workshop on Noisy User-generated Text, pages 148–153, Copenhagen, Denmark. Association for Computational Linguistics.
- [11] Dat Quoc Nguyen and Anh Tuan Nguyen. 2020. PhoBERT: Pre-trained language models for Vietnamese. In Findings of the Association for Computational Linguistics: EMNLP 2020, pages 1037–1042, Online. Association for Computational Linguistics.
- [12] Dang Van Thin, Ngan Luu-Thuy Nguyen, Tri Minh Truong, Lac Si Le, and Duy Tin Vo. 2021. Two new large corpora for vietnamese aspect-based sentiment analysis at sentence level. ACM Trans. Asian LowResour. Lang. Inf. Process., 20(4).
- [13] Hu Xu, Bing Liu, Lei Shu, and S Yu Philip. 2019. Bert post-training for review reading comprehension and aspect-based sentiment analysis. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 2324–2335.
- [14] Lu Xu, Lidong Bing, Wei Lu, and Fei Huang. 2020. Aspect Sentiment Classification with Aspect-Specific Opinion Spans. In Proceedings of EMNLP, pages 3561–3567. Association for Computational Linguistics
- [15] Adwait Ratnaparkhi. 1996. A maximum entropy model for part-of-speech tagging. In Conference on Empirical Methods in Natural Language Processing.
- [16] Minjoon Seo, Aniruddha Kembhavi, Ali Farhadi, and Hannaneh Hajishirzi. 2018. Bidirectional attention flow for machine comprehension.