

인터랙션 기반 추천 시스템 개발을 위한 데이터셋 연구¹

정의석^{0*}, 김현우^{**}, 오효정^{**}, 송화전^{*}

한국전자통신연구원^{*}, 전북대학교^{**}

{eschung, kimhw, songhj}@etri.re.kr^{*}, ohj@jnbu.re.kr^{**}

Dataset for Interactive Recommendation System

Euisok Chung^{0*}, Hyun Woo Kim^{*}, Hyo-Jung Oh^{**}, Hwa Jeon Song^{*}
ETRI^{*}, Jeonbuk National University^{**}

요약

AI와 사용자간의 대화를 통해 사용자의 요구사항을 파악하고, 해당 요구사항에 적합한 상품을 추천하는 형상을 인터랙션 기반 추천 시스템의 한 예로 볼 수 있다. 우리는 해당 시스템 개발을 위하여 의상 코디셋 추천을 위한 대화 기반 데이터셋을 구축하였다. 데이터셋은 대화와 의상 추천 절차를 반복하여 사용자가 원하는 의상셋을 찾아가는 내용으로 구성된다. 그리고, AI의 코디셋 추천 기술 검증에 위해 두가지 의상 추천 평가셋을 제안한다. 본 논문은 대화 데이터셋 및 관련 평가셋의 개발 절차 및 구성에 대하여 기술하고, 관련된 실험 결과 일부를 보여준다.

주제어: 인터랙션, 의상추천, 평가셋

1. 서론

최근 언어이해 기술과 언어생성 기술은 대용량 신경망 접근 방법을 통해 다양한 응용 영역 활용을 가능하게 하였다. 하나 이상의 모달리티에 기반하는 복합지능 영역의 연구를 더 쉽게 하였고 [1], AI와 사용자간의 대화를 통한 AI 지식의 자율 성장 시도도 가능하게 하였다 [2]. 일반적인 기계학습을 통한 AI의 개발은 학습에 이용되는 데이터셋 형상의 의도에서 크게 벗어나지 않는다. 기계번역을 위해서는 다국어의 병렬 코퍼스가 필요하고, 특수한 목적을 지닌 분류기는 해당 분류를 위한 학습 데이터가 요구된다. 따라서, 복합지능과 인터랙션을 통한 지식성장형 AI를 개발하기 위해서는 해당 형상에 부합되는 학습 데이터셋의 설계와 구축이 요구된다.

AI는 사용자와 인터랙션을 통해 사용자의 요구사항을 파악할 수 있다. 그리고 사용자의 요구사항에 적합한 피드백을 줄 수 있는 능력을 지속적으로 성장 시킨다면, 전술한 인터랙션을 통한 지식성장의 개념과 유사해진다. 본 논문은 인터랙션 기반 추천 시스템으로 해당 개념을 형상화 하고, 추천 상품 도메인을 의상 영역으로 한정하여 복합지능 특징을 보장한다. 그리고, 단순 의상 추천이 아니라 의상 코디셋을 추천하는 접근 방법을 통해 지식성장 개념의 구체화를 시도한다. 인터랙션 개념은 대화셋으로 접근하고, 복합지능은 의상의 이미지와 특징들을 언어로 기술한 메타정보로 접근한다.

일반적인 AI연구를 위한 데이터셋은 평가셋 자체를 포함한다. 일부는 학습용, 일부는 평가용으로 분리하여 접근한다. 그러나 대화 영역 AI와 같이 학습 데이터를 그대로 평가셋으로 활용하지 못하는 경우도 있다. 이 경우

는 응답 생성 문장의 언어모델 관점에서의 평가 척도를 활용하거나, 응답셋을 한정하여 분류기 형태로 평가하는 접근 방법이 가능하리라 본다. 본 논문이 의도하는 인터랙션 기반 추천 시스템의 경우는 의상셋 추천 능력을 학습 데이터셋을 이용하여 그대로 평가에 적용하기에는 부적절하다. 사용자의 요구에 부합되는 의상셋은 하나만의 정답으로 한정될 수 없기 때문이다. 따라서, 본 논문은 의상셋 추천 능력을 평가하기 위한 별도의 평가셋을 제시한다. 첫번째는 대화 컨텍스트를 입력으로 몇가지 의상 코디셋의 선호도에 따른 랭킹을 결정하는 평가셋과, 두번째는 대화 컨텍스트와 의상 코디셋 쌍에 대한 관계 척도를 수치화하여 다수의 대화 컨텍스트, 코디셋 쌍을 정렬하는 능력에 대한 평가셋이다.

본 논문이 제안하는 데이터셋은 FASCODE (FASHion COordination DatasEt / FASHion CODE)²로 칭한다. FASCODE는 의상 코디 추천 대화셋, 의상 이미지, 의상 이미지 메타정보, 평가셋1, 평가셋2로 구성되어 있다. 본 논문은 평가셋1에 대한 실험결과 일부를 포함하고 있다. 아래의 URL에 검증이 진행 중인 평가셋2를 제외한 대부분의 데이터가 공개되어 있다.

2. 관련 연구

본 논문과 유사한 선행 연구로 Fashion-IQ [3]가 있다. 하나의 패션 이미지와 해당 이미지의 수정을 요구하는 질의를 입력으로 하고, 해당 입력에 부합하는 의상을 추천하는 정확도를 평가하는 접근 방법을 취했다. 이를 위한 데이터셋과 챗봇지 대화를 통해 관련 접근방법을 검증하였다. Fashion-IQ는 대화가 아닌 질의 응답 형식과

¹ + 동일 기여도 저자

² <https://fashion-how.org/ETRI/board.html>

의상 코디셋이 아닌 단일 의상 추천을 평가 대상으로 하는 점이 본 논문과의 차이점이다.

대화 기반의 인터랙티브 이미지 검색 연구 [4]는 Fashion-IQ와 유사한 연구로 볼 수 있다. 여기서는 후보 이미지와 이에 대한 수정을 요구하는 자연어 질의를 반복적으로 진행하여 원하는 상품 이미지를 검색하는 end-to-end 프레임워크를 제안하였다. 해당 프레임워크에 기반하여 본 논문이 제시하는 데이터셋 형상으로의 접근이 가능하다고 본다. 그러나 FASCODE는 단일 의상 검색을 포함하여 상의, 하의, 겹옷, 신발로 구성된 의상 코디셋의 생성 능력을 학습 및 평가 대상으로 한다.

3. FASCODE

1. 의상 코디 추천 대화셋

화자	발화	발화 유형
<CO>	안녕하세요 코디봇입니다. 무엇을 도와드릴까요?	INTRO
<US>	오빠가 결혼해요.	
<US>	차분하면서 단정한 코디 보여주세요.	
<CO>	네. 잠시만 기다려주세요.	WAIT
<CO>	차분하고 고급스러운 블라우스와 바지 코디입니다.	EXP_RES_DESCRIPTION
<AC>	JP-076 BL-027 PT-027 SE-004	
<US>	겹옷은 자켓으로 바꿔 주시구요.	USER_FAIL
<US>	신발도 단정한 구두였으면 좋겠습니다.	USER_FAIL
<CO>	네. 잠시만 기다려주세요.	WAIT
<CO>	단정한 느낌의 자켓과 구두로 바꾸어 보았습니다.	EXP_RES_DESCRIPTION
<AC>	JK-013 BL-027 PT-027 SE-011	
<US>	네, 좋습니다.	USER_SUCCESS
<CO>	마음에 드신다니 정말 다행입니다.	SUCCESS
<CO>	이용해주셔서 감사합니다.	CLOSING

표 1. 대화셋 예제

대화셋 예제는 표 1에 기술되어 있다. 화자는 코디 발화 “<CO>”와 사용자 발화 “<US>”로 기술되어 있다. 여기서 “<AC>”는 의상 ID정보로 구성된 코디셋을 나타낸다. 코디 발화와 사용자 발화는 대화형의 자연어 문장으로 기술되어 있다. 발화 유형은 대화의 시작과 종료 그리고 의상 추천에 대한 추가적인 이유 기술, 추천 의상에 대한 사용자의 피드백 등의 의미를 기술한다. 표 2는 발화 유형을 기술한다.

하나의 대화셋을 구축하기 위해 처음 결정해야 하는 것은 TPO (Time, Place, Occasion)이다. 그리고, 사용자 프로파일을 선택한다. 여기서 TPO와 프로파일은 미리 정해진 셋을 이용한다. 이후 코디 역할을 담당한 작업자와 사용자 역할을 담당한 작업자는 채팅 도구를 이용하여 대화셋을 구축한다. 해당 대화셋을 위해선 모든 의상 이미지와 해당 이미지에 대한 메타 정보가 미리 구축되어 있어야 했다. 본 연구는 대화셋 구축을 위하여 채팅 도구 및 의상셋 검색 도구를 개발하여 적용하였다. 자세한 사항은 생략한다. 최종적으로 평균 10턴 분량의 7,200셋

수준의 대화셋을 구축하였다.

발화	발화 유형
INTRO	대화 도입부
EXP_RES_*	추천의상 설명
USER_SUCCESS	추천의상 성공
USER_SUCCESS_PART	일부 추천의상 성공
USER_FAIL	추천의상 실패
FAIL	의상 추천 실패
ASK_*	사용자가 원하는 의상유형이나 스타일, 색상등에 대한 질문
CONFIRM_*	확인 질문
SUCCESS	의상 추천 성공
CLOSING	대화 종료
WAIT	대기 요청
SUGGEST_*	제안 발화
NONE	의상 없음
HELP	사용자 지원

표 2. 발화 유형

2. 의상 이미지와 메타 정보


의상ID	유형	메타 정보
BL-001 	형태	단추 여밈의 전체 오픈형 / 스탠드 칼라와 브이넥 네크라인의 결합스타일 / 손목까지 내려오는 일자형 소매 / 여유로운 핏 / 어깨에서 허리까지 세로 절개에 풍성한 러플 장식 / 와이드 커푸스
	재질	면 100% / 구김이 가기 쉬운 / 드라이클리닝 권장
	색상	시원해 보이는 소라색(SKY BLUE) / 단색의 깔끔한 느낌
	감성	여성스러운 / 페미닌한 / 세련된 / 사랑스러운 / 깔끔한 / 오피스룩 / 로맨틱한 데이트룩 / 포멀한 이미지 / 단정한 오피스걸 특이 이미지

표 3. 의상 ID와 메타 정보

표 3은 의상 ID와 해당 의상의 메타 정보를 기술한다. 메타 정보는 형태, 재질, 색상, 감성 표현으로 기술되어 있다. 기술된 언어들 가능한 의상 전문 용어는 배제하려 노력했다. 의상 ID에 따른 의상 이미지는 하나의 의상 ID에 하나의 이미지로 유지하였다. 이미지는 전면 이미지 만을 대상으로 하였다.

표 1의 “<AC>”는 최대 4개의 의상으로 구성되어 있다. 해당 의상은 겹옷(outer), 상의(top), 하의(bottom), 신발(shoes)로 구성된다. 각각의 의상 종류는 표 4에 기술되어 있다. 표 3의 BL-001은 상의(top)에 속한 블라우스를 의미한다는 것을 알 수 있다. 신발을 포함한 의상들은 총 2,600개를 구축하였고, 각각 이미지 정보를 포함하고 있다. 의상은 데이터 수집 및 작업자 선정 등의 이유로 인해 여성복에 한정하였다.

대화셋과 의상셋 구축은 지속적으로 진행되고 있으며, 향후 관련된 추가 정보를 공개할 예정이다. 또한, 의상 전문가를 통한 의상 코디셋을 독립적으로 구축하였고, 해당 코디셋에 대한 사용자 평가점수를 작성하여 대화셋 구축에 활용하였다.

의상 분류	의상 ID 유형
겉옷(outer)	'CT': '코트', 'CD': '가디건', 'VT': '조끼', 'JK': '자켓', 'JP': '후드'
상의(top),	'KN': '저지', 'SW': '스웨터', 'SH': '셔츠', 'BL': '블라우스'
하의(bottom),	'SK': '스키트', 'PT': '팬츠', 'OP': '원피스'
신발(shoes)	'SE': '신발'

표 4. 의상 분류에 따른 ID 종류

3. 의상 코디 평가셋 및 평가 방식

화자	발화
<CO>	안녕하세요. 코디붐입니다. 무엇을 도와드릴까요?
<US>	락 페스티벌에 가는데 독특하고 특이한 룩으로 추천해주세요.
<CO>	치마, 바지, 원피스 중 어떤 옷이 포함된 코디를 추천해드릴까요?
<US>	치마로 추천해주세요.
<CO>	짧은 기장의 치마를 추천해드릴까요?
<US>	네.
RANK1	
RANK2	
RANK3	

표 5. 평가셋1 예제

의상 코디셋 추천 시스템 개발을 위해 시스템의 코디셋 추천 능력을 평가할 수 있는 절차의 필요성이 제기되었다. 이를 위해 FASCODE는 두가지 평가셋을 구상하였다. 첫번째 평가셋은 3-4턴 분량의 대화셋을 질의로 하고, 3가지 코디셋 후보를 rank1, rank2, rank3 순서로 구축하였다. 평가 방식은 코디셋 후보의 rank를 정렬하는 접근 방법으로 평가 기준을 세웠다. 표 5는 평가셋1의 예제이다. 해당 예제의 rank 구축 방법은 다음과 같다. 1) 대화셋 질의로부터 컨텍스트 충돌요소를 파악: “평범

한 옷”, “긴 치마(계절 추측)”, “코디와 어울리지 않은 의상”. 2) rank1 코디를 대화셋을 통해 선정. 3) 1개의 충돌 요소로 rank2 선정: “코디와 어울리지 않는 상의”, 4) 2개 이상의 충돌 요소를 통해 rank3 선정: “코디와 어울리지 않는 상의”, “계절에 어울리지 않는 상의”.

평가셋 1의 평가 척도는 Weighted Kendall's Tau를 사용한다.³ 켄달타우(Kendall's tau)는 순위 상관 계수(rank correlation coefficient)의 한 종류이며 두 변수들간의 순위를 비교하여 연관성을 계산한다. 가중치 켄달 타우(Weighted Kendall's Tau, WKT)는 여기서 순위에 따라 가중치가 고려된 상관 계수를 측정한다. 평가셋 1은 총 900세트로 구성되어 있다.

평가셋2는 대화컨텍스트와 의상코디셋간의 관계성을 절대 평가 형식으로 0-10스케일의 점수를 다수의 평가자가 참여하여 구축하였다. 단어간의 유사도를 측정하는 평가셋인 WordSim353에 영향을 받았다.⁴ 총 450쌍으로 평가셋 2는 구성되어 있다. 해당 평가셋을 위해서는 대화컨텍스트와 코디셋간의 유사도 점수를 시스템이 계산해 낼 수 있어야 한다. 평가셋1, 2의 개발 의도는 해당 점수를 예측할 수 있는 시스템을 개발하기 위함이다. 평가셋2의 검증은 향후 진행 예정이다. 표 6은 평가점수 가이드라인에 대하여 기술하고 있다.

점수	기준
컨텍스트와 코디셋의 충돌요소 없음 (8-10)	
10	코디셋에 주관적/객관적 관점에서 동의
9	코디셋에 주관적에는 동의할 수 없으나 객관적 관점에서는 동의
8	코디셋에 동의할 수 없음
컨텍스트와 코디셋의 충돌요소 1개 (5-7)	
7	코디셋에 주관적/객관적 관점에서 동의
6	코디셋에 주관적에는 동의할 수 없으나 객관적 관점에서는 동의
5	코디셋에 동의할 수 없음
컨텍스트와 코디셋의 충돌요소 2개 (0-4)	
4	코디셋에 주관적/객관적 관점에서 동의
3	코디셋에 주관적에는 동의할 수 없으나 객관적 관점에서는 동의
2	코디셋에 동의할 수 없음
1	컨텍스트와 코디셋의 충돌요소가 3개 이상이나 코디셋은 동의
0	컨텍스트와 코디셋의 충돌요소가 3개 이상이고 코디셋은 동의 못함

표 6. 평가셋2 점수 가이드라인

4. 실험

실험에 사용한 베이스라인의 구성도는 그림 1과 같다. 평가셋1은 표 5처럼 대화와 후보 의상 코디들로 구성되는데, 베이스라인의 목적은 대화를 토대로 가장 적절한

³ scipy.stats.weightedtau

⁴ <http://alfonseca.org/eng/research/wordsim353.html>

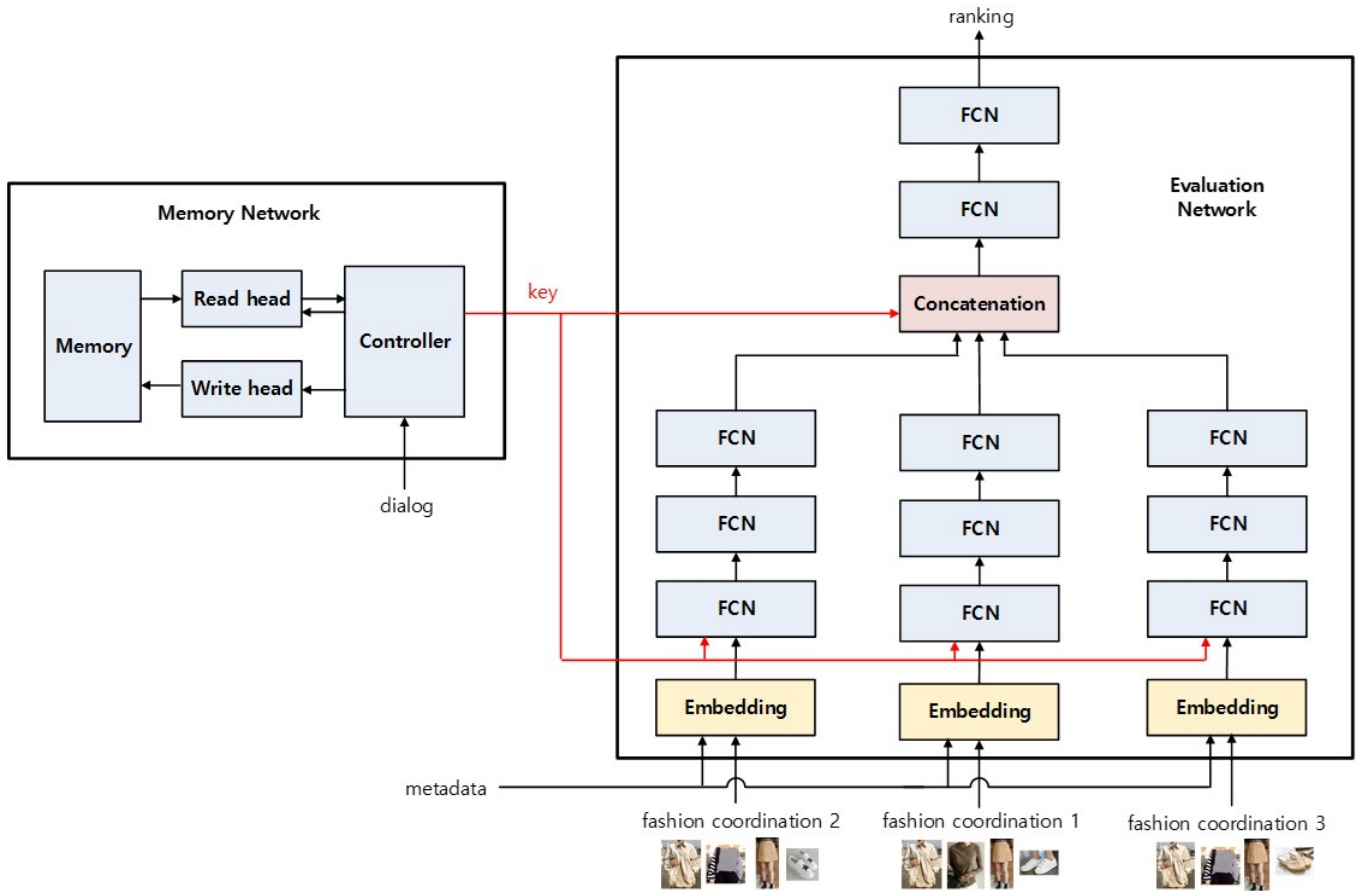


그림 1. 베이스라인 구성도

의상 코드의 순위를 매기는 것이다.

베이스라인은 우선 기억망을 사용하여 대화로부터 핵심 요구사항을 추정한다. 여기서 기억망으로는 종단간 기억망 [5]을 사용한다. 대화의 문장은 ETRI가 자체 개발한 서브워드 임베딩 벡터를 사용하여 BoW(Bag-of-Word) 기법으로 문장 임베딩 벡터로 변환된다. 입력 문장의 임베딩 벡터는 메모리에 저장되고, 질문의 임베딩 벡터는 주의집중 과정을 반복적으로 수행하여 메모리로부터 대화의 핵심 요구사항을 추정한다.

다음으로 다수 계층의 FCN(Fully-Connected Network)로 구성된 평가망을 사용하여, 후보 의상 코드가 대화에 얼마나 적합한지를 기준으로 순위를 산정한다. 후보 의상 코드는 의상 아이템의 메타 정보를 사용하여 임베딩한다. 의상 코드는 4개의 의상 아이템으로 구성되어 있고, 의상 아이템의 메타 정보는 형태/재질/색상/감성 등 4가지 특징으로 구성되어 있다. 특징 별로 ETRI가 자체 개발한 서브워드 임베딩 벡터를 사용하여 BoW 기법으로 특징 임베딩을 수행하고, 4가지 특징 임베딩을 연결하여 의상 아이템 임베딩을 수행하며, 4개의 의상 아이템을 연결하여 의상 코드 임베딩을 수행한다. 임베딩한 후보 의상 코드와 대화의 핵심 요구사항을 연결한 것을 입력으로 3개 계층의 FCN을 사용하여 1차 평가를 출력한다. 3개 후보 의상 코드에 대한 1차 평가 결과를 토대로 2개 계층의 FCN으로 최종 순위를 산정한다. 최종 순위의 구현은 3개 후보 코드의 조합 개수인 6개를 분류하는 형

태로 한다. 따라서 학습은 6개 분류에 대한 크로스 엔트로피를 기준으로 확률적 경사 하강법으로 수행된다.

학습을 수행하기 위해서는 대화셋에서 평가셋1처럼 후보의상 코드를 선정할 필요가 있다. 여기서는 대화에서 최종 제안한 의상 코드로부터 대화의 역방향으로 중복 없이 3개의 의상 코드를 후보 의상 코드로 결정한다. 또한 성능을 높이기 위해 데이터를 증강한다. 대화의 최종 제안한 의상 코드에서 임의로 의상 아이템을 선택하고 그 아이템과 코사인 유사도가 적은 아이템으로 교체함으로써 새로운 데이터를 생성한다. 후보 의상 코드의 순서는 임의로 배정한다.

900셋의 평가셋1중 200셋은 개발셋으로 하여 하이퍼패러미터 튜닝을 진행하였고, 700셋으로 평가를 진행하였다. 표 7은 이에 대한 결과를 보여 준다. 비교는 3개의 순서를 랜덤하게 셔플한 결과와의 WKT값을 Dev와 Test 셋 각각 10회 구하고 이를 평균하였다. 그림 1의 베이스라인 접근 방법은 Dev에서 0.5296 그리고 Test에서 0.5249로 Random 결과에 비해 월등하게 좋은 결과를 보였다.

평가셋1	Random WKT	Baseline WKT
Dev set	0.0065	0.5296
Test set	0.0076	0.5249

표 7. 실험 결과

5. 결론

본 논문은 인터랙션 기반 의상 추천 시스템 개발을 위한 의상 코디 추천 대화셋, 의상 메타 정보 그리고 의상 코디 평가셋에 대한 구성과 구축 결과에 대하여 기술하였고, 평가셋 일부에 대한 검증에 대하여 기술하였다. 해당 검증을 통해 평가셋 자체의 타당성을 확인할 수 있었다. 향후, 대용량 사전학습 모델 연구, 연속학습, 퓨샷(few-shot)학습 등을 접목하여 구축 데이터를 다양한 AI영역에 활용할 예정이다.

감사의 글

본 연구는 한국전자통신연구원 연구운영지원사업의 일환으로 수행되었음. [20ZS1100, 자율성장형 복합인공지능 원천기술 연구]

참고문헌

- [1] L. Zhou, H. Palangi, L. Zhang, H. Hu, J. J Corso, and J. Gao, "Unified Vision-Language Pre-Training for Image Captioning and VQA." AAAI. 2020.
- [2] 송화진, 김현우, 정의석, 오성찬, 이진우, 강동오, 정준영, 이윤근, "자율성장 인공지능 기술", 전자통신동향분석, 제34권, 제4호, pp43-54, 2019.
- [3] X. Guo, H. Wu, Y. Gao, S. Rennie, and R. Feris, "Fashion IQ: A New Dataset towards Retrieving Images by Natural Language Feedback." arXiv preprint arXiv:1905.12794. 2019.
- [4] Guo, X., Wu, H., Cheng, Y., Rennie, S., Tesauro, G., & Feris, R. (2018). Dialog-based interactive image retrieval. In Advances in neural information processing systems (pp. 678-688).
- [5] S. Sukhbaatar et al., "End-to-end memory networks," NIPS2015.