자율성장 인공지능 기술

Self-Improving Artificial Intelligence Technology

송화전 (H.J. Song, songhj@etri.re.kr) 복합지능연구실 책임연구원
김현우 (H.W. Kim, kimhw@etri.re.kr) 복합지능연구실 책임연구원
정의석 (E. Chung, eschung@etri.re.kr) 복합지능연구실 책임연구원
오성찬 (S. Oh, sungchan.oh@etri.re.kr) 시각지능연구실 선임연구원
이전우 (J.W. Lee, ljwoo@etri.re.kr) 시각지능연구실 책임연구원
강동오 (D. Kang, dongoh@etri.re.kr) 시각지능연구실 책임연구원
정준영 (J.Y. Jung, jyjung21@etri.re.kr) 시각지능연구실 책임연구원
이윤근 (Y.K. Lee, yklee@etri.re.kr) 인공지능연구소 책임연구원/소장

ABSTRACT

Currently, a majority of artificial intelligence is used to secure big data; however, it is concentrated in a few of major companies. Therefore, automatic data augmentation and efficient learning algorithms for small-scale data will become key elements in future artificial intelligence competitiveness. In addition, it is necessary to develop a technique to learn meanings, correlations, and time-related associations of complex modal knowledge similar to that in humans and expand and transfer semantic prediction/knowledge inference about unknown data. To this end, a neural memory model, which imitates how knowledge in the human brain is processed, needs to be developed to enable knowledge expansion through modality cooperative learning. Moreover, declarative and procedural knowledge in the memory model must also be self-developed through human interaction. In this paper, we reviewed this essential methodology and briefly described achievements that have been made so far.

KEYWORDS 자율성장 AI, Strong AI, 절차적 지식, 기억모델, Human Interaction

Ⅰ. 서론

최근 인공지능 기술이 IT 분야는 물론 많은 산업분야에서 좋은 성과를 내면서 관심이 집중되고 있다. 특히, 딥러닝 기술이 영상인식, 음성인식 등의 분야에 적용되면서 기존 성능을 월등히 추월하고 있으며, 기존 기술 발전의 패러다임을 빅데이터 기반의 방법론으로 전이시키는 현상이 나타

* DOI: https://doi.org/10.22648/ETRI.2019.J.340405
* 본 연구는 한국전자통신연구원 연구운영지원사업의 일환으로 수행되었음[19ZS1100, 자율성장형 AI 핵심원천기술 연구].

본 저작물은 공공누리 제4유형
출처표시+상업적이용금지+변경금지 조건에 따라 이용할 수 있습니다.
©2019 한국전자통신연구원

나고 있다. 하지만 인공지능 기술에 대한 기대 못지않게 한계도 극명하게 드러나고 있는데, 첫째는 현재의 인공지능 기술이 수동 가공된 대용량 데이터에 기반한 기술이므로 데이터 확보에 많은 시간과 비용이 필요하다는 점이며, 둘째는 학습된 문제 이외에 다른 분야로 확대 적용하기 어렵다는 것이다.

이러한 문제점을 해결하기 위해 비지도학습형 인공지능 기술, 자율성장형 인공지능 기술, 상식 인공지능(Machine Common Sense) 등에 대한 연구가 시작되고 있으며, 본 논문에서도 이러한 방법론에 대하여 논하고자 한다. 또한, 현재의 인공지능 기술의 대표적 분야인 언어지능, 시각지능, 청각지능 등은 모두 인간의 지능 영역 중 한가지만을 다루고 있으며, 각각의 기술과 방법론이 별도로 발전하고 있는 상황이나 향후 인간과 유사한 인공지능 기술 개발을 위해서는 인간과 유사한 복합적 지능을 구현하는 기술이 필요하다.

인간의 지식은 크게 선언적 지식과 절차적 지식으로 나눌 수 있다. 선언적 지식은 '…을 안다(knowing that)'로 표현할 수 있는 지식이며, 주로 암기와 이해를 통해 학습한다. 절차적 지식은 '…을 할 수 있다', '…을 하는 방법을 알고 있다(knowing how)'로 표현되며, 주로 행함을 통해 습득한다.

선언적 지식을 모사한 대표적인 인공지능으로 IBM의 왓슨이나 국내 ETRI에서 개발한 엑소브레인 등을 예로 들 수 있다. 이 인공지능은 많은 지식을 암기하고 있다가 사용자가 묻는 질문(퀴즈, 의학지식, 법률지식 등)에 대한 답을 알려준다. 이러한 인공지능은 수많은 데이터(주로 문서)를 이용하여 지도학습의 과정을 통해 지식을 학습한다. 절차적 지식을 모사하는 인공지능은 주로 게임이나 로봇의 행동 등에 적용되며 강화학습에 의해 지식을 학습하는 방법을 주로 사용한다. 강화학습은 이미 '알파고'와 같은 고난이도의 게임에 적용되어 사람의 능력을 넘어서는 결과를 도출한 바 있으며, 최근에는 스타크래프트 등의 전략 시뮬레이션 게임에서도 성공적인 결과를 보이기도 했다. 하지만, 사람의 일상적 지능 활동인 대화, 상담 등을 인공지능이 수행하거나 특정 분야의 노하우를 인공지능이 학습하는 것에 대해서는 강화학습이 아직까지 성공적인 결과를 내놓지 못하고 있다. 즉, 인공지능이 상품에 대한 모든 정보를 가지고 있어도 전문적인 쇼핑호스트의 기능을 수행하기 어렵고, 법률지식을 충분히 학습해도 변호사의 역할을 수행하기 어려운 이유이다. 이와 같이 인간 전문가를 인공지능이 따라가기 어려운 이유는 강화학습에 필요한 보상모델(Reward model)이 게임과 같이 명확한 목표를 가진 태스크에 대해서는 잘 동작하지만 인간의 노하우와 같이 명확히 규격화하기 어려운 분야에는 적합하지 않기 때문이다.

본 논문은 이 문제를 해결할 수 있는 몇 가지 핵심기술에 대해 제안하며, 궁극적으로는 인공지능이 인간의 일상적인 절차적 지식을 스스로 학습할 수 있는 '교감형 자율성장 AI' 방법론을 연구하고자 한다. 이는 미지의 음성/텍스트/영상/생체 신호 등의 복합모달 입력에 대하여 이해, 추론하며, 인터랙션을 통해 경험지식을 성장시키는 메커니즘이라 할 수 있다.

Ⅱ장에서는 복합모달 기반 자율성장 인공지능에 대한 개략적인 형상과 기본이 되는 요소 기술에 대한 동향에 대하여 설명하며, Ⅲ장에서는 현재 개발 중인 복합모달 자율성장 인공지능 시스템에 대한 간략한 설명과 향후 개발 방향에 대해 설명한다. 마지막으로 Ⅳ장에서 결론을 맺는다.

Ⅱ. 복합모달 기반 자율성장 인공지능 개발을 위한 요소기술 동향

언어/음성/이미지/생체 신호 등의 복합모달 입력으로 구성된 지식을 처리하는 자율 성장 인공지능 개발은 현재 태동기에 있는 상태이며, 그림 1은 이에 대한 시스템 형상의 일례이다. 기본적으로 인간의 뇌에서 복합모달 지식을 처리하는 것처럼 단기기억과 장기기억 모듈을 구성하여 현재 입력에 대해 과거의 기억을 사용하여 사용자의 요구사항을 출력하게 된다. 또한 이에 대해 사용자의 피드백을 받아 지속적으로 지식을 성장시키는 구조를 가지고 있다.

하지만 그림 1과 같은 능력을 가진 인공지능 시스템을 구성하기에는 넘어야 할 장벽이 아직은 너무 많다. 이러한 어려움으로 인해 현재는 단일모달이고 각각의 모듈에 대해 선구적인 요소 기술들이 연구 개발되고 있는 상황이며, 본 장에서는 자율성장 인공지능 기술을 구성하는 가장 기본이 되는 기술들에 대해 기술한다.

1. 메모리 네트워크 기술

수많은 선언적·절차적 지식을 표현하기 다루기 위해서는 일반적인 심층 신경망으로는 그 한계가 명확하며, 이를 극복하기 위해 최근 폰 노이만 구조의 컴퓨팅 모델처럼 신경망도 논리 흐름 제어와 외부 메모리를 명시적으로 분리하고 처리함으로써 컨텍스트 정보를 효과적으로 사용할 수 있는 메모리 네트워크에 대한 연구가 진행되고 있다.

가. 신경 튜링 기계

신경 튜링 기계[1]는 앨런 튜링이 고안한 튜링 기계의 기본적인 구조와 신경망을 활용한 모델이다. 그림 2는 제어기와 메모리를 분리한 신경 튜링 기계의 구조를 보여준다. 신경망으로 구성된 제어기는 입력 신호를 받고, 읽고 쓸 메모리의 위치와 값을 결정하여 헤드를 통해 메모리에 접근한다. 또한 메모리에서 읽은 값과 입력 신호를 사용하여 출력 신호를 계산한다.

나. 종단 간 메모리 네트워크

종단 간 메모리 네트워크[2]에서 문장은 단어 단위로 BoW(Bag-of-Word) 기법을 적용하여 벡터로

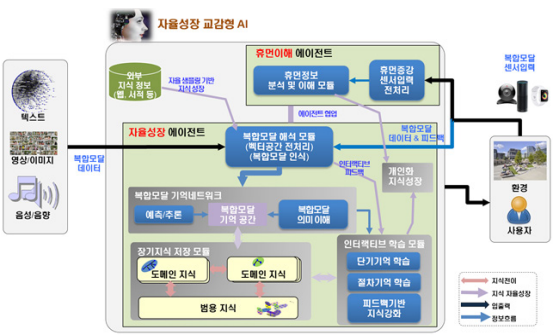
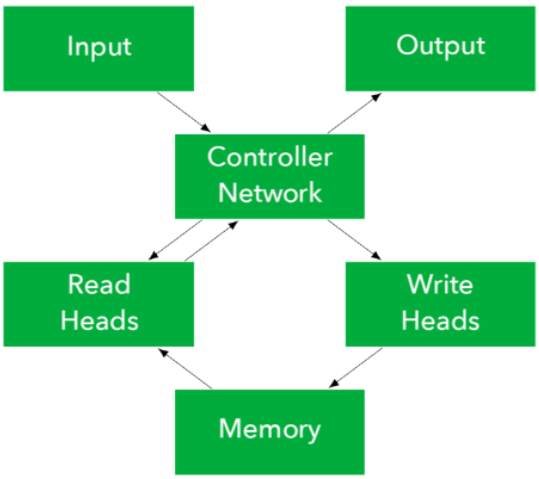


그림 1 복합모달 자율성장 교감형 AI 시스템 형상 예



출처 Kjerish [CC BY-SA 4.0 (https://creativecommons.org/licenses/by-sa/4.0)]

그림 2 신경 튜링 기계의 구조

변환되고, 이를 다시 선형 변환함으로써 임베딩된다. 입력 문장의 임베딩 벡터는 메모리에 저장되고, 질문의 임베딩 벡터는 집중 과정을 수행하여 메모리로부터 출력 신호를 계산한다. 이러한 과정을 반복 수행하여 답변을 예측한다.

2. 관계 추론 기술

관계 추론 기술은 사람의 추론 방식과 유사하게 객체 와 객체 간의 관계를 추론하는 기술을 말한다. 특정 개체의 속성을 추론하는 기술을 넘어 각 개체들 사이의 상대적 관계를 파악함으로써 문제 해결을 해야 한다. 구글 딥마인드에서는 이러한 관계형 추론을 할 수 있는 관계 네트워크[3]를 제안하였다. 합성곱 신경망(CNN: Convolutional Neural Network)과 LSTM 구조의 순환 신경망의 출력을 입력으로 받아 객체를 추출하고, 심층 신경망을 사용하여 객체 간의 관계를 추론한다.

3. 강화 학습 기술

강화 학습은 에이전트가 환경으로부터 현재의 상태를 인식하여, 선택 가능한 행동들 중 보상을 최대화하는 행동을 선택하는 방법을 학습하는 기술이다. 몬테 카를로 샘플링으로 인해 발생하는 높은 분산과 많은 학습 소요 시간을 해결하는 과정에서 A3C(Asynchronous Advantage Actor-Critic) 알고리즘[4]이 제안되었으며, 최근에는 거의 대부분의 강화학습 방식의 베이스라인으로 사용된다. 이러한 알고리즘은 정책망과 가치망 모두 신경망으로 구현하고, 상태 가치망을 평균으로 사용해서 현재의 행동이 평균적으로 얻을 수 있는 가치보다 얼마나 더 좋은가를 계산하며, 복수의 에이전트를 사용하여 훈련한다.

4. 벡터 표현 기반 텍스트 전처리 기술

가. 워드 임베딩 기술

워드 임베딩은 주어진 단어 목록의 각 단어에 벡터값을 할당하는 기술이다. 여기서 유사한 단어들은 주어진 벡터 공간상 근접한 거리에 위치하게 된다. 전형적인 접근방법으로 대용량 텍스트 데이터를 이용한 스킵그램(Skip-gram) 모델[5]이 있다. 하지만 스킵그램 모델은 일반적으로 해당 단어 목록에 포함되어 있지 않은 미등록어에 대한 처리를 따로 고려해야 하는 문제가 있다. 이를 해결하는 방법으로 주어진 단어를 구성하는 자질들(Sub-features)을 추출하여 해당 자질들에 대한 임베딩을 학습하는 서브워드 임베딩(Subword embeddings) 기술이 있으며, 이를 통해 미등록어 문제에 대응할 수 있다[6].

나. 문장 임베딩 기술

문장을 벡터 값으로 표현하는 접근 방법을 문장 임베딩이라 한다. 일반적으로 문장을 구성하는 단어들의 임베딩 벡터값의 평균으로 해당 문장 임베딩 값을 계산할 수 있다. 여기서 고려할 사항은 워드 임베딩의 경우 해당 단어의 주변 문맥 단어들 만을 고려하여 모델링한다는 점이다. 따라서, 문장 문맥을 고려하지 않는 문제가 있는데, 이런 이슈 자체를 고려한 접근 방법 중 하나로 Siamese CBOW 모델이 있다[7].

Siamese CBOW 모델은 문장 표현을 위한 단어 벡터 모델의 일종인 반면, 참고문헌 [8]은 매우 일반적인 문장 표현을 생성하기 위해 스킵서트(Skip-thought) 벡터를 제안했다. 이는 스킵그램 모델과 형상이 유사한데, 목표 단어와 문맥 단어의 모델을 목표 문장과 문맥 문장에 대한 모델로 치환했다. 세 개의 연속 문장으로 인코더, 디코더 모델을 사

용했는데, 인코더에 대한 하나의 목표 문장은 각각의 디코더를 통해 2개의 문맥 문장들을 생성하는 형상을 지닌다.

5. 시각속성 인식기술

영상으로부터 객체를 한정하는 특징인 시각속성(Visual Attributes)은 일반적으로 색상, 모양, 재료 및 부품과 같이 동일 종류의 객체 간 구분이 가능하고, 다른 종류의 객체 간에 공유도 가능한 의미론적 특성으로 정의된다. 시각속성은 이미지 검색, 제로샷(Zero-shot) 객체분류, 영상 설명, 사람 분석 및 식별 등 컴퓨터비전에서 광범위하게 연구되었다[11-14].

Nagarajan와 Grauman[15]은 속성과 객체가 결합된 클래스 구조로 인해 동일 속성의 서로 다른 객체들의 학습이 비효율적인 문제를 해결하기 위하여 속성을 명확히 분리하는 의미론적 임베딩을 학습하고, 속성과 객체가 결합되었을 때의 효과를 규칙화함으로써 미학습 속성과 객체의 조합을 인식하는 방법을 제안하였다.

Sarafianos 등[16]은 하나의 객체가 여러 속성을 갖는 다중 레이블(Multi-label) 특성, 클래스 불균형, 영상샘플 부족 등의 문제에 대응하기 위해 다중 스케일의 시각적 주의집중 마스크들을 생성하여, 클래스 및 인스탄스 레벨에서의 클래스 불균형을 반영하는 손실함수를 적용하고, 높은 예측 편차를 갖는 주의집중 마스크의 가중치를 차별적으로 높여 주는 시각속성 학습법을 시각 속성인식에 적용하였다.

Song 등[17]은 실세계 응용에 제로샷 객체분류 기술을 적용하기 위하여 사람이 정의한 속성(Human-defined Attributes)뿐만 아니라 미묘하게 구분 가능한 잔여속성(Residual Attribute)을 자동으로 구분하여 학습할 수 있고, 예측 신뢰도를 산출방안을 부가하여 신뢰도가 일정 수준 이하일 경우 예측을 포기하는 선택적 제로샷 분류기를 제안하였다.

시각속성 인식기술은 시장의 요구가 커서 상업화 가능성이 높고, 데이터 셋 확보가 비교적 용이한, 패션 관련 분야에서 관심이 높아지고 있다[11,14,18].

홍콩 중문대학에서는 속성예측(Attributes Prediction), 상점 내 상품 검색(In-shop Retrieval), 상점 밖 일반인 착용 의류와 동일한 상품 검색(Consumer-to-shop Retrieval), 패션 랜드마크 검출 등의 연구를 위하여 80만 장 이상의 의복 영상 및 태깅 데이터를 포함하는 대규모의 데이터 셋(DeepFashion)을 구축하여 벤치마크와 함께 공개하였다. DeepFashion 데이터 셋과 함께 발표된 FashionNet은 저수준 영상특징 추출을 위해 VGG16 심층신경망을 채용하고, 마지막 합성곱 레이어 다음 단에 패션랜드마크 및 랜드마크 가시성을 검출하는 포즈단, 랜드마크 주변의 특성을 추출하는 국지 외형단 및 패션의 전체적인 외형 특성을 추출하는 글로벌 외형단을 병렬로 구성하여 패션의 카테고리 분류, 속성분류를 처리한다[14].

Ak 등[18]은 AlexNet의 전 연결층을 제거하고 두 개의 합성곱 레이어 및 GAP(Global Average Pooling) 레이어를 추가하여 각각의 속성에 대한 속성 활성화 지도(AAMs: Attribute Activation Maps)를 생성함으로써 속성별 ROIs(Region of Interests)를 처리할 수 있도록 구성하여 영역별 속성이 다른 의류를 구별 및 검색할 수 있는 FashionSearchNet을 제안하였다.

6. 선언적 지식 자율 성장 기술

본 절에서는 기계가 사실, 명제, 뉴스 등의 선언

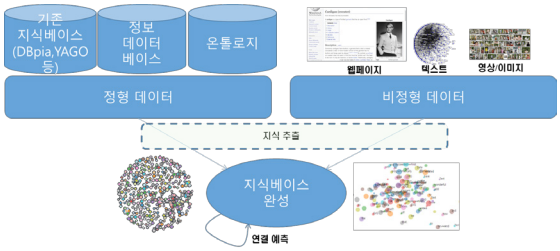


그림 3 일반적인 선언적 지식 자율 성장 구조

적 지식을 습득하여 스스로 지식을 성장시키는 기술에 대해 살펴본다. 선언적 지식의 자율 성장을 위해서는 정형/비정형 학습 데이터로부터 지식 정보를 추출하여 지식 표현으로 변환하는 기술과 추출된 지식 정보를 바탕으로 지식을 성장시켜 축적하는 기술이 필요하다[19]. 일반적으로 언어 기반 데이터로부터 자연어 처리/이해 기술을 통하여 지식 정보를 추출하고, 이를 지식 표현으로 변환하여 지식베이스의 지식을 성장시키는 연구가 많이 진행되어 왔다. 본 절에서는 이러한 기존의 언어 기반 선언적 지식 성장 기술을 소개하고 이를 복합모달 데이터에 적용하는 기술 연구 동향에 대해서 살펴본다(그림 3 참조).

가. 언어 기반 지식 추출 및 표현 기술

선언적 지식을 저장하여 데이터베이스와 같이 탐색과 추론을 할 수 있는 지식베이스는 표 1과 같은 것이 있다.

표 1의 지식베이스들은 일반적으로 위키피디아와 같이 여러 사람에 의하여 생성된 정형화된 지식이나 웹페이지로부터 크롤링한 비정형화된 데이터를 자연어 처리와 같은 전처리를 거친 후에 크라우드 소싱 기법으로 사람에 의하여 지식을 검증하고, 이를 그래프나 온톨로지 형식으로 저장한다.

비정형 언어 데이터로부터 객체 간의 관계를 자동 추출하여 이를 지식 표현으로 변환하는 기술로 TextRunner, WOE, Open IE, Reverb, OLLIE 등 원거리 감독(Distant Supervision) 모델 등과 자연어 처리나 이해 기술을 바탕으로 한 기술이 연구되었다. 근래에는 워드 임베딩, CNN, RNN 등의 딥러닝 기술을 활용하여 관계 추출의 성능을 높이고, 데이터 셋이 부족한 경우도 관계 추출이 가능하도록 하는 연구가 진행되고 있다[20,21]. 또, 기존 지식베이스의 검증된 정형화된 사실 지식을 활용하여 추출된 지식의 신뢰도를 검증하고 이를 지식화하는 Knowledge Vault 등과 같은 기술이 연구되었으며, 국내에서는 ETRI가 주관한 엑소브레인(Exo-Brain) 사업에서 유사한 관계 추출 및 온톨로지 확장 기술을 개발한 바 있다[22,23].

나. 지식베이스 완성 기술

정형/비정형 데이터로부터 지식을 추출하여 구성한 지식베이스는 완전하지 못하여 모든 객체가 정의된 모든 관계를 정의하지 못한다. 따라서 불완전한 관계를 보완하는 기술로 객체 간의 관계를 예

표 1 지식베이스 종류

속성	지식베이스 명칭
개방형	Yago, OpenIE, WIKIDATA, DBPedia, ConceptNet, OpenCYC 등
상용	MS Satori, Google Knowledge Graph, Facebook Entity Graph, Amazon Evi 등

표 2 연결 예측 기술 종류

접근 방식	알고리즘 이름
규칙 기반	FOIL, Path Ranking Algorithm 등
확률 기반	PSL, Multiple Relational Clustering 등
분해 방식	RESCAL 등
임베딩 방식	TransE, DistMult, ComplEx, HoIE, GAKE, ConvKB, RotatE 등

측하는 연결 예측(Link Prediction) 기술이 있으며, 표 2와 같은 기술들이 개발되었다[24]. 최근에는 객체 간의 관계를 임베딩된 공간상의 함수로 변환하여 딥러닝 학습을 통하여 관계를 예측하는 기술이 활발히 연구되고 있다[25,26].

다. 복합모달 지식베이스 성장 기술

딥러닝의 발달로 복합모달 데이터에 대한 처리 기술이 발달함에 따라 단일 출처의 언어 데이터뿐만 아니라 다양한 출처의 복수 데이터로부터 선언적 지식을 추출하여 성장시키는 기술이 연구되고 있다. 이미 추출되어 정형화된 지식을 가진 이질적인 복수의 지식베이스들로부터 지식을 융합하여 새로운 지식베이스를 생성하거나 확장하는 기술 연구가 진행되었다[27,28]. 또 웹 등에서 추출된 비정형화된 언어와 이미지 등이 융합된 지식베이스를 구축하고 이에 대해서 지식베이스를 완성하는 기술 연구도 진행되고 있다[29,30].

III. 복합모달 기반 자율성장 인공지능 기술 개발 현황

그림 1의 시스템을 한번에 개발하는 것은 달성하기 어려운 목표이므로 단계를 나누어 점진적으로 전체 시스템을 구성하는 방식이 효율적이다. 본 논문에서도 먼저 복합모달 입력에 대한 단기기억을 통한 시스템의 성장에 초점을 맞추었고, 이를 AI 패션 코디네이터 도메인으로 검증하고자 하였다. 실제 수집된 복합모달 및 인터랙션 데이터 일례를 그림 4에 나타내었다. 즉, 패션 코디네이션 전문가가 일반인에게 패션 코디를 제안하는 절차를 그대로 데이터로 나타내었으며, 각각의 의상에 대해서는 상세한 메타데이터 정보도 함께 구성하여, 선언적 지식으로 사용하도록 하였다. 그림 4의

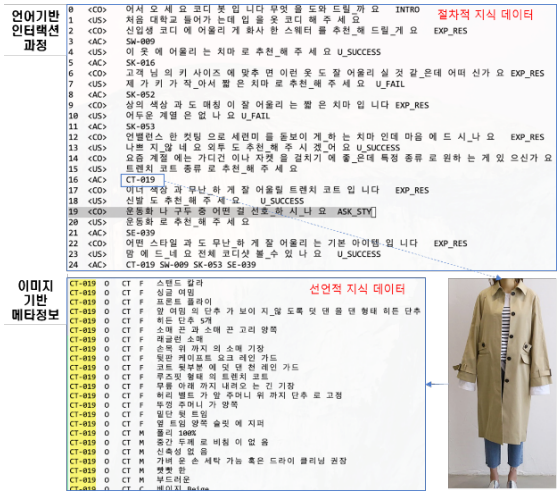


그림 4 복합모달 AI 패션 코디네이터 인터랙션 및 복합모달 데이터 일례

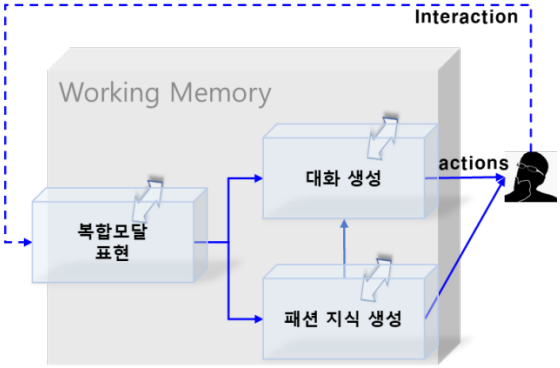


그림 5 복합모달 AI 패션 코디네이터 실증 시스템

복합모달 입력에 대해 임베딩 및 기억모델 등을 사용하여 추천 패션에 대해 지식을 생성하고 이를 바탕으로 대화가 생성되도록 하였다. 그림 5에서 이런 과정을 그대로 모방하도록 구성하였다. 단기 기억(Working Memory) 모듈은 모든 다른 모듈에서 접근이 가능하며 과거 이력을 포함하여 현재 요청한 정보를 추출하게 된다.

1. 패션 코디네이션 지식 생성 기술

패션 코디네이션 지식은 패션을 착용할 TPO

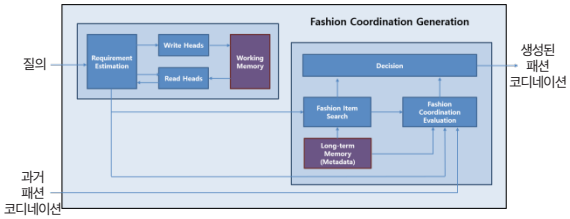


그림 6 패션 코디네이션 생성 구조

(Time, Place, Occasion)에 대한 사용자의 요구사항을 반영하여 패션 아이템의 조합을 생성하는 절차 지식을 말한다. 본 논문에서는 언어와 이미지를 통해 사용자의 요구사항을 파악하고, 사용자의 반응으로 강화 학습을 수행함으로써 패션 코디네이션 지식을 획득한다. 정확한 패션 코디네이션 지식을 생성하기 위해서는 사용자의 요구사항 컨텍스트 정보와 패션 이력을 충분히 활용할 필요가 있기 때문에 명시적인 메모리를 사용한다. 또한 패션 아이템들의 조합 과정에서는 패션 아이템 속성들 간의 관계형 추론이 필요하다.

그림 6은 패션 코디네이션 생성 구조를 보여준다. 우선 사용자의 요구사항 추정 과정에 메모리 네트워크를 사용한다. 현재 질의에 대해 작업 메모리에 접근하는 데 필요한 파라미터들을 생성하고, 그 파라미터를 사용하여 읽고 쓸 작업 메모리의 위치와 값을 생성한 후 읽기와 쓰기 작업을 수행하며, 읽어 온 작업 메모리 값으로 요구사항을 추정한다. 다음으로 사용자의 요구사항에 적합한 패션 아이템의 조합을 생성한다. 추정한 요구사항과 패션 아이템들의 메타데이터를 사용하여 요구사항에 맞는 패션 아이템의 확률을 산출하고, 새로 구성된 패션 코디네이션이 요구사항에 적절한지와 얼마나 어울리는지를 평가함으로써 최적의 패션 코디네이션을 생성한다. 이러한 패션 코디네이션 지식은 사용자 반응을 통한 강화학습으로 훈련함으로써 획득한다.

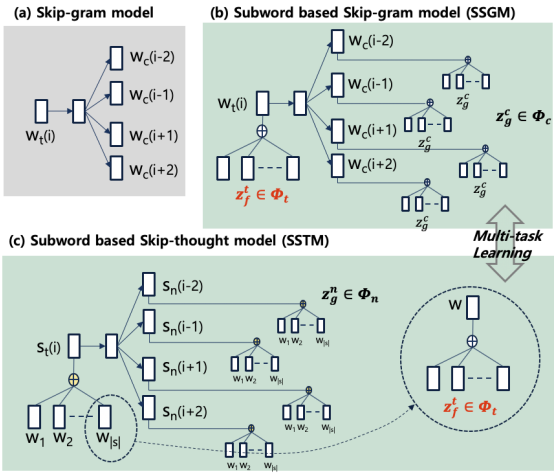


그림 7 서브워드 기반 스킵서트 모델

2. 서브워드 기반 스킵서트 모델

제안하는 워드 임베딩 기술의 특징은 문장 문맥 정보를 서브워드 벡터에 반영하는 접근 방법을 시도한다는 데 있다. 제안하는 모델의 기본 모델은 그림 7의 (a)에 기술되어 있다. 스킵그램 모델[5]은 목표 단어 w_t가 문맥 단어 w_c를 예측하는 확률 모델이다. 여기서, 목표 단어와 주변 문맥 단어는 각각 다른 임베딩 벡터로 표현된다.

서브워드 기반 스킵그램 모델(SSGM: Subword based Skip-Gram Model)은 그림 7의 (b)에 기술되어 있다. 스킵그램 모델과의 차이점은 각 단어의 임베딩 벡터가 하위 서브워드 벡터의 합으로 기술된다는 것이다. 여기서 Φ_t는 목표 단어의 서브워드 벡터 집합이고, Φ_c는 문맥 단어들의 서브워드 벡터 집합을 말한다. 그리고 서브워드 벡터 z^t는 단어 w_t의 구성 자질이면서 Φ_t에 속해 있고, 서브워드 벡터 z^c는 문맥 단어 w_c의 구성 자질이면서 Φ_c에 속해 있다. 모델 학습은 $p(w_c|w_t)$를 최대화하는 접근 방법을 취한다. 여기서 학습 결과는 서브워드 집합인 Φ_t와 Φ_c에 반영된다.

서브워드 기반 스킵서트 모델(SSTM: Subword

based Skip-Thought Model)은 그림 7의 (c)에 기술되어 있다. 기본 구조는 SSGM과 동일하다. 차이점은 목표 문장 s_t가 문맥 문장 s_c을 예측하고, 문장 임베딩 값이 문장 구성 단어의 하위 서브워드 자질 값에서 구해진다는 것이다. 본 논문에서는 SSGM과 SSTM을 동시에 학습하는 멀티 태스크 학습을 사용한다. 여기서 연결 지점은 SSGM과 SSTM 간에 공유되는 Φ_t이다. 최종 학습 결과 Φ_t를 단어 임베딩 작업에 사용할 수 있다.

3. 시각속성 인식 기술 개발 현황

가. 의복 시각속성 인식

본 논문에서는 시각속성 인식의 응용예로 의복에 대한 속성인식기를 개발 중이다. 구현 중인 시각속성인식기의 구성도는 그림 8과 같다.

나. 재귀적 구조를 사용한 의복 구조 추정

영상을 통해 관측되는 의복은 의복 자체의 형태뿐만 아니라 착용자의 자세도 여러 형태로 변화되어 관측된다. 본 논문에서는 다양한 의복 형태에도 불구하고 올바른 시각속성인식을 수행하기 위하여 의복의 구조적 특징을 추출 과정을 수행한다.

의복의 구조 추정은 그림 8과 같은 구조를 가지는 CNN(Convolutional Neural Network)을 이용하여 추정한다. 의복의 구조를 정의하는 특징은 Deep-Fashion 데이터 셋[14]에 정의된 특징점(landmark)과 의복영역으로 정하였다. 합성곱 레이어는 동일

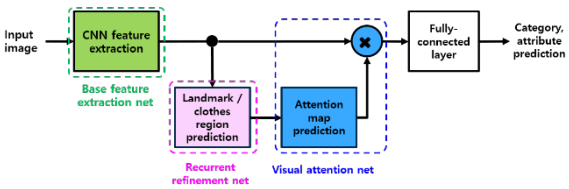


그림 8 시각속성 인식기의 구성도

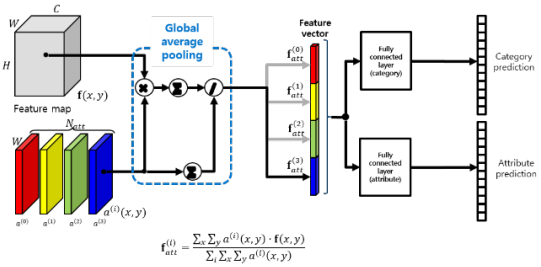


그림 9 시각주의 기반 속성인식기의 구성도

한 가중치를 공유하므로 재귀적인(recurrent) 특징을 가지며, 반복횟수에 따라 연산량과 추정정확도 사이의 절충이 가능하다.

다. 시각주의 기반 시각 속성 인식

본 논문에서 목표로 하는 시각속성 인식 기술은 객체의 전역적인 특징뿐 아니라 국부적인 속성에 대한 인식을 수행하여야 하기 때문에 특징 맵(Feature map)으로부터 각 국부영역의 특징을 시각주의(Visual attention)를 이용하여 분리하여 추출한다.

영상의 주의영역은 객체의 구조적인 특징과 화소별 특징에 결정되어야 하기 때문에 그림 9와 같이 특징 맵을 이용하여 추정된 복수의 시각주의 맵(Visual attention map)을 계산한다. 추정된 시각주의 맵을 가중치로 하여 특징맵에 대한 전역적 가중평균을 계산하여 1차원 특징벡터를 추출한다.

이상에서 서술한 방법으로 구한 특징벡터는 국부적인 특징 정보의 합으로 구성되어 있으므로 전역적인 특징을 필요로 하는 의복 분류와 국부적인 속성 인식에 동일하게 사용될 수 있다.

4. 인터랙션 기반 성장형 패션 코디네이션 시스템 구현

이상에서 설명한 기술들을 통합하여 AI 패션 코디네이터 프로토타입 시스템을 개발하였으며, 그

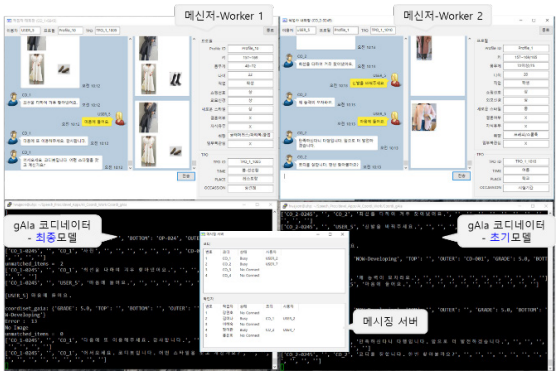


그림 10 인터랙션 기반 성장형 패션 코디네이션 시스템

림 10은 임의의 순간의 스틸 컷을 보여준다. 앞서 설명한 것과 같이 그림 4의 데이터를 그대로 모방하도록 그림 5와 같은 시스템을 학습시켰다. 프로토타입 시스템의 기본적인 동작은 카카오톡과 유사한 메시징 통신 형태이며, 사용자의 메신저에 그림 4에서 보여준 예제처럼 임의의 입력을 하여 전송하게 되면 터미널 형태인 AI 코디네이터가 이를 받아 사용자가 요구하는 적절한 코디셋과 적절한 답변을 출력하여 사용자 메신저로 다시 전송하게 된다. 그림 10의 아래 부분의 두 개의 터미널 중 오른쪽 AI 코디는 초기의 적은 양의 데이터로 훈련한 상태이며, 왼쪽 AI 코디는 인터랙션을 통해 보다 많은 경험지식 데이터로 훈련한 상태로 오른쪽 AI 코디보다 사용자가 요구하는 의상을 좀 더 적절하게 제시하여 준다. 비록 프로토타입 시스템이지만 현재 연구 개발하고자 하는 방향이 적절한지에 대한 검증을 수행한 것이며, 결과적으로 타당한 결론을 얻었다. 하지만 향후 지속적으로 연구 개발할 내용은 훨씬 더 복잡하고 난해하며 이에 대해서는 다음 절에서 간략하게 소개한다.

5. 향후 연구 개발 계획 현황

이상에서 설명한 AI 패션 코디네이터는 그림 1의 모든 기술을 포함하지는 않는다. 향후 장기 기억의 지식그래프를 자율적으로 성장시키는 기술과 복합모달에 대한 단기 및 장기 기억 간의 지식 상호 교환 및 저장 기술 및 사용자 인터랙션에서 암묵적인 리워드를 추정하는 역강화학습 기술도 필요하다. 그리고, 수동 가공된 데이터가 적은 환경에서 모델 훈련이 가능하도록 유저 시뮬레이터 개발이 요구되며, 적은 데이터로 훈련할 때 발생하는 오버피팅(overfitting) 문제해결도 중요한 연구개발 내용이다. 또한 과거 학습한 지식을 잃어버리지 않도록 지속적 학습(Continual learning) 기술이나 다중 도메인 지식을 다루는 멀티 에이전트 기술 개발도 필요하다. 뿐만 아니라 이러한 모든 기술들이 서로 시너지가 발생하도록 통합하는 것도 어려운 숙제이며, 장기간에 걸쳐 점진적으로 해결할 수 있도록 시스템을 구체적으로 설계, 수정 및 보완하는 것도 중요한 향후 계획이다.

Ⅳ. 결론

인간의 지식은 선언적 지식과 절차적 지식으로 이루어져 있으며, 또한 복잡한 복합모달 구조를 다룬다. 이러한 임의의 다양한 지식에 대해 오류없이 처리할 수 있는 인공지능 시스템을 개발하는 것은 현재 모든 인공지능 개발자의 궁극적인 목표이다. 본 논문에서는 이러한 시스템의 선도적인 기술 개발을 위해 대표적인 각 요소 기술에 대한 개발 동향을 살펴보았고, 또한 발생된 문제를 해결할 수 있는 몇 가지 핵심기술을 제안하였다. 이는 미지의 음성/텍스트/영상/생체 신호 등의 복합모달 입력에 대하여 이해, 추론하며 인터랙션을 통해 경험지식을 성장시키는 메커니즘으로, 궁극적으로는 인공지능이 인간의 일상적인 절차적 지식을 스스로 학습할 수 있는 '교감형 자율성장 AI' 방법론을

구축하는 원대한 목표이며, 본 논문에서는 아주 먼 여정을 위한 첫 걸음을 내딛고자 하였다.

약어 정리

A3C	Asynchronous Advantage Actor–Critic
BoW	Bag of Word
CNN	Convolutional Neural Networks
FOIL	First Order Inductive Learner
GAKE	Graph Aware Knowledge Embedding
GAP	Global Average Pooling
HolE	Holographic Embeddings
LSTM	Long Short–Term Memory
OLLIE	Open Language Learning for Information Extraction
PSL	Probabilistic Soft Logic
ROI	Region of Interest
SSGM	Subword based Skip–Gram Model
SSTM	Subword based Skip–Thought Model
TPO	Time, Place, Occasion
TransE	Translating Embeddings
WOE	Wikipedia–based Open Extractor

참고문헌

[1] A. Graves et al., "Neural turing machines," arXiv preprint: 1410.5401, 2014.

[2] S. Sukhbaatar et al., "End-to-end memory networks," in *Proc. NIPS*, 2015.

[3] A. Santoro et al., "A simple neural network module for relational reasoning," in *Proc. NIPS*, 2017.

[4] V. Mnih et al., "Asynchronous Methods for Deep Reinforcement Learning," in *Proc. ICML*, 2016.

[5] T. Mikolov et al., "Distributed representations of words and phrases and their compositionality," in *Proc. NIPS*, 2013.

[6] P. Bojanowski et al., "Enriching word vectors with subword information," in *Proc. TACL*, 2017.

[7] T. Kenter et al., "Siamese CBOW: optimizing word embeddings for sentence representations," in *Proc. ACL*, 2016.

[8] R. Kiros et al., "Skip-thought vectors," in *Proc. NIPS*, 2015.

[9] 정의석 외 3인, "한국어 음소열 기반 워드 임베딩 기술," 한국어정보학회 학술대회 논문집, 2017.

[10] S. Park et al., "Subword-level word vector representations for Korean," in *Proc. ACL*, 2018.

[11] R. S. Feris et al., "Visual Attributes," Springer International Publishing, 2017.

[12] A. Farhadi et al., "Describing Objects by their Attributes," in *Proc. IEEE Conf. Comput. Vision Pattern Recogn.*, 2009, pp. 1778-1785.

[13] O. Russakovsky and F. Li, "Attribute learning in large-scale datasets," in *Proc. Eur. Conf. Comput. Vision*, 2010, pp. 1-14.

[14] Z. Liu, et al., "DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations," in *Proc. IEEE Int. Conf. Comput. Vision Pattern Recogn.*, 2016, pp. 106-1104.

[15] T. Nagarajan and K. Grauman., "Attributes as Operators: Factorizing Unseen Attribute-Object Compositions," in *Proc. Eur. Conf. Comput. Vision*, 2018, pp. 169-185.

[16] N. Sarafianos et al., "Deep Imbalanced Attribute Classification using Visual Attention Aggregation," in *Proc. Eur. Conf. Comput. Vision*, 2018, pp. 708-725.

[17] J. Song et al., "Selective Zero-Shot Classification with Augmented Attributes," in *Proc. Eur. Conf. Comput. Vision*, 2018, pp. 474-490.

[18] K. E. Ak et al., "Learning Attribute Representations with Localization for Flexible Fashion Search," in *Proc. IEEE Conf. Comput. Vision Pattern Recogn.*, 2018, pp. 7708-7717.

[19] E. Gabrilovich and N. Usunier, "Constructing and Mining Web-scale Knowledge Graphs," in *Proc. SIGIR*, Tutorial Slide, 2016

[20] S. Kumar, "A survey of deep learning methods for relation extraction," arXiv:1705.03645, 2017.

[21] X. Han et al., "FewRel: A Large-Scale Supervised Few-Shot Relation Classification Dataset with State-of-the-Art Evaluation," in *Proc. EMNLP*, 2018, pp. 4803-4809.

[22] X. Dong et al., "Knowledge vault: A web-scale approach to probabilistic knowledge fusion," in *Proc. ACM SIGKDD Int. Conf. Knowledge Discovery Data Mining*, 2014, pp. 601-610.

[23] J. Kim and S.H. Myaeng. "Discovering relations to augment a web-scale knowledge base constructed from the web," in *Proc. Int. Conf. Web Intell., Mining Semantics*, 2016, pp. 16;1-12.

[24] D.Q. Nguyen, "An overview of embedding models of entities and relationships for knowledge base completion," arXiv:1703.08098, 2017.

[25] Q. Wang et al., "Knowledge graph embedding: A survey of approaches and applications," *IEEE Trans. Knowledge Data Eng.*, vol. 29, no. 12, 2017, pp. 2724-2743.

[26] 최현영 등, "지식 베이스 임베딩을 활용한 지식 완성 모델링 기법," 한국정보과학회 논문지, 제45권 제9호, 2018.9, pp. 895-903.

[27] H. Al-Mubaid and S. Bettayeb, "An Algorithm for Combining

Graphs Based on Shared Knowledge," in *Proc. ISCA BICoB*, 2012. pp. 137-142.

[28] Q. Xie et al., "An interpretable knowledge transfer model for knowledge base completion," arXiv preprint:1704.05908, 2017.

[29] G. Dihong and D. Z. Wang, "Extracting Visual Knowledge from the Web with Multimodal Learning," in *Proc. IJCAI*, 2017, pp. 1718-1724.

[30] P. Pouya, L. Chen, and S. Singh, "Embedding multimodal relational data for knowledge base completion," arXiv preprint:1809.01341, 2018.