

파라미터 튜닝을 통한 Relation Networks 성능개선

이현옥*, 임희석**

* 고려대학교 컴퓨터 정보 통신 대학원 빅데이터 융합학과

** 고려대학교 컴퓨터학과

e-mail : webpainter@korea.ac.kr, limhseok@korea.ac.kr

Improving the performance for Relation Networks using parameters tuning

Hyun-Ok Lee*, Heui-Seok Lim**

* Dept. of Big Data Convergence, Korea University Graduate School of Computer & Information Technology

**Dept. of Computer Science and Engineering, Korea University

요 약

인간의 추론 능력이란 문제에 주어진 조건을 보고 문제 해결에 필요한 것이 무엇인지를 논리적으로 생각해 보는 것으로 문제 상황 속에서 일정한 규칙이나 성질을 발견하고 이를 수학적 방법으로 법칙을 찾아내거나 해결하는 능력을 말한다. 이러한 인간인지 능력과 유사한 인공지능 시스템을 개발하는데 있어서 핵심적 도전은 비구조적 데이터(unstructured data)로부터 그 개체들(object)과 그들간의 관계(relation)에 대해 추론하는 능력을 부여하는 것이라고 할 수 있다. 지금까지 딥러닝(deep learning) 방법은 구조화 되지 않은 데이터로부터 문제를 해결하는 엄청난 진보를 가져왔지만, 명시적으로 개체간의 관계를 고려하지 않고 이를 수행해왔다.

최근 발표된 구조화되지 않은 데이터로부터 복잡한 관계 추론을 수행하는 심층신경망(deep neural networks)은 관계추론(relational reasoning)의 시도를 이해하는데 기대할 만한 접근법을 보여주고 있다. 그 첫 번째는 관계추론을 위한 간단한 신경망 모듈(A simple neural network module for relational reasoning)인 RN(Relation Networks)이고, 두 번째는 시각적 관찰을 기반으로 실제대상의 미래 상태를 예측하는 범용 목적의 VIN(Visual Interaction Networks)이다. 관계 추론을 수행하는 이들 심층신경망(deep neural networks)은 세상을 객체(objects)와 그들의 관계(their relations)라는 체계로 분해하고, 신경망(neural networks)이 피상적으로는 매우 달라 보이지만 근본적으로는 공통관계를 갖는 장면들에 대하여 객체와 관계라는 새로운 결합(combinations)을 일반화할 수 있는 강력한 추론 능력(powerful ability to reason)을 보유할 수 있다는 것을 보여주고 있다.

본 논문에서는 관계 추론을 수행하는 심층신경망(deep neural networks)중에서 Sort-of-CLEVR 데이터 셋(dataset)을 사용하여 RN(Relation Networks)의 성능을 재현 및 관찰해 보았으며, 더 나아가 파라미터(parameters) 튜닝을 통하여 RN(Relation Networks) 모델의 성능 개선방법을 제시하여 보았다.

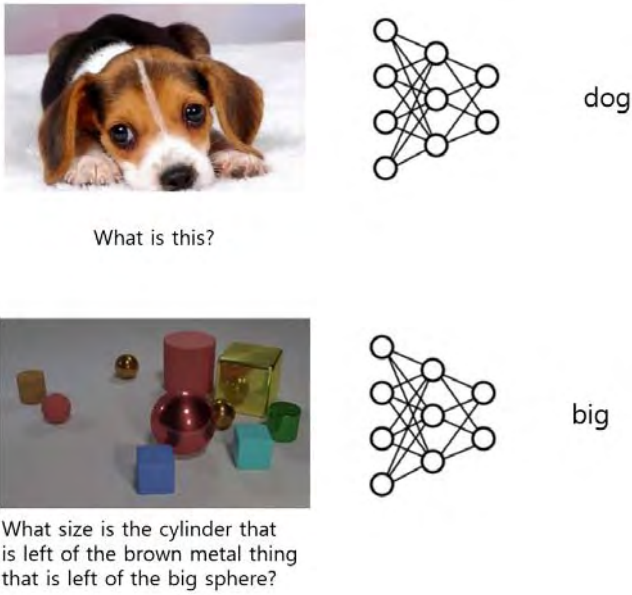
1. 서론

기존 시각기반의 질의응답 문제(Visual Question Answering)에 있어 심층 학습(deep learning)의 표준적 접근 방법은 질문(question), 이미지(image) 및 답변(answer)으로 구성된 대규모 데이터 셋(data set)을 수집하고 질문 및 이미지에서 답변으로 직접 매핑하는 단일 신경 네트워크를 교육하는 것이었다.

대부분의 질문이 아래의 첫 번째 사진의 경우처럼 익숙한 이미지 인식에 있다면 이러한 단일 신경망 네트

워크는 매우 효과적이다. 그러나 질문이 아래의 두 번째 사진과 같다면 결과는 이와 같은 질문에 대해서는 잘 작동하지 않는다는 것이다. 이러한 질문에 대한 해결책은 무엇일까?

해결책으로 제시할 수 있는것은 관계추론(relational reasoning)으로써 서로 다른 개체들 간의 관계를 이해하는 것을 학습하는 것이다. 관계추론(relational reasoning)은 지능(intelligence)의 기본적 특성으로 간주된다.



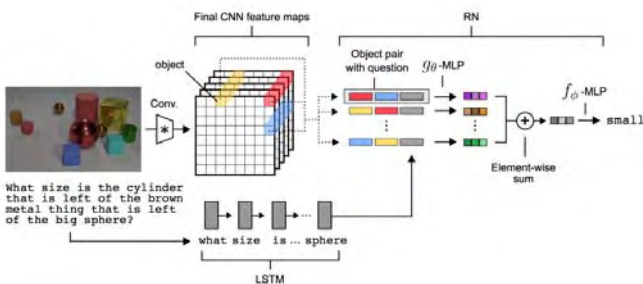
(그림 1) 기존 VQA vs 관계추론 VQA

이에 딥마인드(DeepMind)에서는 이러한 관계 추론 (relational reasoning)을 할 수 있는 RN(Relation Networks)을 개발하였다. RN (Relation Networks)은 간단하며, 다른 모델에 쉽게 붙일 수 있고, 유연한 관계추론에만 중점을 둘 수 있다. RN 을 CNN 과 LSTM 과 결합하여 시각기반의 질의응답 문제(Visual Question Answering)에 대해서 실험을 한 결과, 기존 모델은 76.6% 이하의 성능을, 사람은 92.6%, RN 을 적용한 모델은 95.5% 성능이 나왔다.

본 논문에서는 Sort-of-CLEVR 데이터 셋(dataset)을 이용, RN(Relation Networks)의 성능을 재현 및 관찰해 볼 것이며, 더 나아가 파라미터(parameters) 튜닝을 통하여 RN(Relation Networks) 모델의 성능 개선 방법을 제시해 보고자 한다.

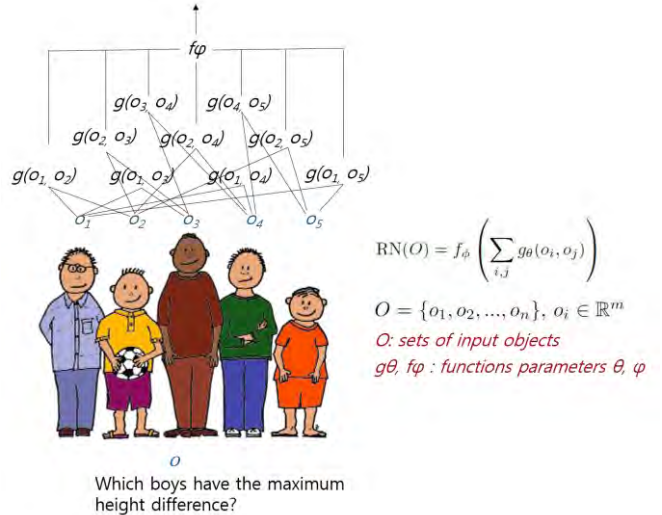
2. RN-augmented Visual QA architecture

RN(Relation Networks)은 CNN 과 LSTM 등의 모델에서의 출력을 입력 받아 객체를 추출하고 각 객체간의 관계를 추론한다. 아래 그림은 두 객체간의 관계를 학습할 수 있는 CNN(Convolutional Networks), LSTM(Long Short-Term Memory) 그리고 RN(Relation Networks)이 결합된 end to end neural network 모델의 아키텍처(architecture)이다.



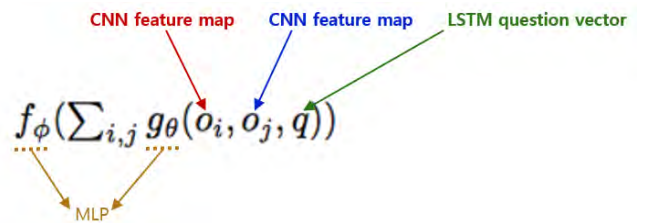
(그림 2) Visual QA architecture using RN

RN 의 기본수식은 다음과 같다. O 는 관계를 알고 싶어하는 object 를 의미한다. g_{θ} 는 한 쌍의 object 를 입력으로 받아서 관계를 도출 하는 함수로 한 쌍의 객체 사이의 관계를 계산하는 다층 퍼셉트론(MLP)이다. f_{ϕ} 는 모든 g_{θ} 의 합계를 취하여 모델의 최종 출력을 계산하는 다층 퍼셉트론(MLP)이다.



(그림 3) RN(Relation Networks) 수식

RN(Relation Networks)는 하나의 모듈로서 CNN (Convolutional Neural Networks)와 LSTM(Long Short-Term Memory)에 플러그인 될 수 있다. CNN 로 부터 나온 '객체(object)'와 LSTM 로 부터 나온 벡터(vector)를 사용하여 관계형 네트워크를 학습한다. 관계(relation)를 알고 싶은 각 객체(object) 쌍은 LSTM 의 질문 벡터(question vector)와 함께 신경망인 g_{θ} 의 입력으로 사용된다. g_{θ} 의 출력이 합쳐져 다른 신경망인 f_{ϕ} 의 입력으로 사용되고, f_{ϕ} 는 질문에 대한 답에 대해 최적화된다.



(그림 4) Relation Networks conditioned using LSTM

3. Sort-of-CLEVR 도메인 RN 모델 성능 재현 및 관찰

RN(Relation Networks)을 검증한 다양한 도메인 중 본 논문에서는 Sort-of-CLEVR 도메인을 사용하여 RN 모델의 성능을 재현 및 관찰을 해 볼 것이며, 더 나아가 파라미터(parameters) 튜닝을 통하여 RN 모델의 성능 개선방법을 제시해 보고자 한다. Sort-of-CLEVR 은 CLEVR 의 간소화된 버전으로 각각의 이미지에 10000 개의 이미지와 20 개의 질문(10 개의

관계형 질문-relational question 과 10 개의 비관계형의 질문- non-relational question) 으로 구성된 데이터 셋이다.

Sort-of-CLEVR 도메인을 사용한 RN 성능 재현 및 관찰 모델에서는 최적화 모델을 찾기 위하여 다음의 최적화(optimization) 방법을 사용하였다. 재현 모델에서는 배치 정규화(batch normalization)와 드롭아웃(dropout)을 사용하여 오버피팅(overfitting)을 억제하였고, 최적의 매개변수를 찾기 위해 Adam 을 사용, 하이퍼 파라미터(hyper parameter)의 편향을 보정 및 학습의 갱신강도를 적응적으로 조정하였다. RN 이 없는 CNN + MLP 을 baseline 으로 설정, RN 이 결합된 모델과 성능을 비교하여 보았다. 성능 비교에 사용된 질문은 관계형 (relational) 질문 / 비관계형 (non-relational) 질문이라는 두 가지 유형의 질문으로 나누었다. Sort-of-CLEVR 도메인을 사용한 RN 모델을 본 논문에서 성능 재현 및 관찰 결과는 다음과 같다.

<표 1> Sort-of-CLEVR 성능 재현 및 관찰 결과

RN 여부	질문유형	Relational Networks (20th epoch) -batch-size:64, learning rate:1e ⁻⁴	CNN + MLP (without RN, 50th epoch) -batch-size:64, learning rate:1e ⁻⁴
	Relational question	87~90%	71%
	Non-relational question	99%	63%

본 논문에서는 다양한 최적화 방법을 사용한 RN 모델의 성능 재현 및 관찰 외에 파라미터 튜닝을 통해 RN 모델의 성능 개선을 시도해 보았다. 튜닝에 사용한 파라미터로는 batch size 와 learning rate 를 정하였다.

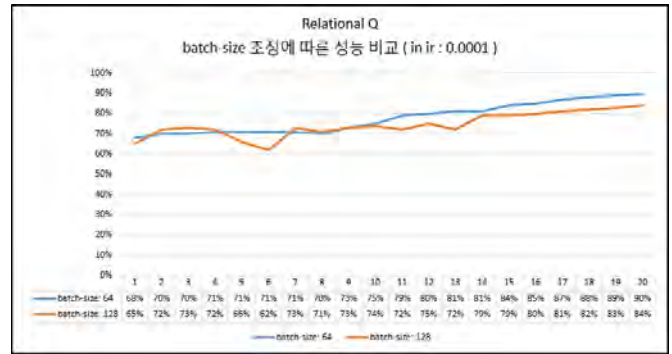
4. Batch- size 설정에 따른 모델 성능 비교

신경망은 입력을 한 번에 하나씩 처리하지 않는다. 처리량을 높이려면 데이터를 일괄 처리하는데, 이를 배치 사이즈(batch size)라고 한다. batch size 크기가 너무 작으면 일괄 처리 내에 많은 변화가 발생하고 training loss 곡선이 많이 바뀐다. batch 크기가 너무 크면 메모리가 부족해 지거나 training 이 너무 느리게 진행된다. 따라서 적절한 batch size 가 중요하다 할 수 있다.

적절한 batch size 를 찾기 위해 앞서 RN 성능 재현 및 관찰에서 구현한 Sort-of-CLEVR dataset 모델 (bs: 64) 대비 batch-size 를 두 배(bs: 128) 로 늘려 관계형 질문(relational question)과 비관계형 질문(non-relational question)에 대한 성능을 비교 분석해 보았다.

관계형 질문(relational question)의 경우 learning rate 1e⁻⁴ 기준 epoch 별 성능을 비교 해 본 결과 batch size 가 큰 경우가 작은 경우에 비해 불안정한 성능을 보였고, 20 으로 설정한 epoch 이 끝난 경우 batch size 를 64 로 설정한 모델은 90%의 성능을 보

였지만, 128 로 설정한 모델은 84%의 성능을 보여주었다.



(그림 5) batch size 조정에 따른 관계형 질문에 대한 RN (Relation Networks) 모델 성능 검증 결과

반면 비관계형 질문(non-relational question)에 대해서는 learning rate 1e⁻⁴ 기준 epoch 별 성능을 비교 해 본 결과 batch size 가 작게 설정된 모델이 batch size 를 두 배로 설정한 모델보다 epoch 4~14 까지의 구간에서 성능이 높게 나왔지만, 20 으로 설정한 epoch 이 끝난 경우 관찰해 보면 batch size 를 64 로 설정한 모델과 128 로 설정한 모델 모두 99%라는 높은 성능을 보여주었다.

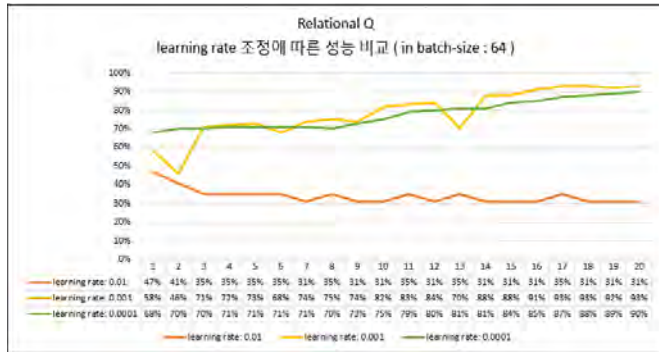
5. Learning rate 설정에 따른 모델 성능 비교

신경망 학습의 목적은 손실함수를 줄이는 것이다. 따라서 신경망 학습에서는 학습률(learning rate) 값이 중요하다. 이 값이 너무 작으면 학습 시간이 너무 길어져서 거의 갱신되지 않고 학습이 끝나고, 반대로 너무 크면 발산하여 학습이 제대로 이뤄지지 않는다. 따라서 적절한 학습률 조정이 중요하다. 적절한 학습률을 찾기 위해 앞서 RN 성능 재현 및 관찰에서 구현한 Sort-of-CLEVR dataset 모델(lr: 1e⁻⁴) 대비 learning rate 를 10 배(lr: 1e⁻³), 100(lr: 1e⁻²)배로 늘려 관계형 질문(relational question)과 비관계형 질문(non-relational question)에 대한 성능을 비교 분석해 보았다.

관계형 질문(relational question)의 경우 batch size 64 기준 epoch 별 성능을 비교 해 본 결과 learning rate 가 1e⁻² 인 경우 매우 낮은 성능을 보여주었고, learning rate 가 1e⁻³ 인 경우 초반에는 다소 낮았지만 epoch 13 인 경우를 제외하고는 learning rate 가 1e⁻⁴ 인 경우보다 성능이 다소 높게 나왔다. 20 으로 설정한 epoch 이 끝난 경우의 결과를 보면 learning rate 를 1e⁻² 로 설정한 모델은 31%의 성능을 보였고, learning rate 를 1e⁻³으로 설정한 모델은 93%의 성능을, learning rate 를 1e⁻⁴ 으로 설정한 모델은 두 번의 실험을 반복해 본 결과 87~90%의 성능을 보여주었다.

반면 비관계형 질문(non-relational question)의 경우 batch size 64 기준 epoch 별 성능을 비교 해 본 결과 learning rate 가 1e⁻² 인 경우 초반보다 epoch

진행 될수록 매우 낮은 성능을 보여주었고, learning rate 가 $1e^{-3}$ 인 경우 초반에는 다소 낮았지만 epoch 9 부터 learning rate 가 $1e^{-4}$ 인 경우보다 확연하게 성능이 높게 나왔다. Learning rate 가 $1e^{-4}$ 인 경우는 epoch 이 11 인 경우부터 높은 성능을 보여주었다. 20 으로 설정한 epoch 이 끝난 경우의 결과를 보면 learning rate 를 $1e^{-2}$ 으로 설정한 모델은 15%의 매우 낮은 성능을 보였고, learning rate 를 $1e^{-3}$ 으로 설정한 모델은 100%의 성능을, learning rate 를 $1e^{-4}$ 으로 설정한 모델은 두 번의 실험을 반복해 본 결과 두 번의 경우 모두 99%의 성능을 보여주었다.



(그림 6) learning rate 조정에 따른 관계형 질문에 대한 RN (Relation Networks) 모델 성능 검증 결과

6. 실험 결과 및 평가

본 논문에서는 Sort-of-CLEVR dataset 에 관한 최적화된 RN 모델을 찾기 위해 파라미터 튜닝을 통한 RN 모델의 성능 개선을 시도해 보았고 그 결과를 보면 다음과 같다. 표에서 보여주고 있듯이 질문의 유형에 상관없이 배치 사이즈(batch size)가 64 일 때와 학습률(learning rate)이 너무 낮지도 높지도 않은 $1e^{-3}$ 인 경우 높은 성능이 나왔음을 알 수 있다. 또한 batch size 보다 learning rate 조정에 의해 모델의 성능이 확연히 달라질 수 있음을 확인할 수 있었다.

<표 2> Sort-of-CLEVR 파라미터 튜닝 성능 개선 결과

batch-size	learning rate	Performance (%)	
		Relational question	Non-relational question
64	$1e^{-4}$	87~90%	99%
64	$1e^{-3}$	93%	100%
64	$1e^{-2}$	31%	15%
128	$1e^{-4}$	84%	99%
128	$1e^{-3}$	88%	100%

7. 결론 및 향후 연구

RN(Relational Networks)을 검증한 다양한 도메인 중 본 논문에서는 Sort-of-CLEVR 도메인을 사용한 RN 모델의 성능을 재현 및 관찰을 해 보았고, 재현 과정에서 모델의 최적화(optimization)를 위한 방법을 살펴 보았다.

또한 기존 모델의 파라미터(parameters) 튜닝- batch

size 와 learning rate-으로 RN(Relation Networks) 모델의 성능 개선방법을 제시해 보았다. 물론 전체 데이터 양과 epoch 의 횟수가 batch size 및 learning rate 에 영향을 줄 수 있기 때문에 해당 실험으로 파라미터의 최적 설정이라고 단정하기 어렵지만, 실험 단계에서 보았듯이 모델의 성능에 크게 영향을 미치는 요인이 learning rate 임을 알 수 있었다. 추후 RN(Relation Networks) 모델의 성능향상을 위하여 learning rate 의 더 섬세한 조정을 해 볼 것이며, 더 나아가 논문에서 구현한 Sort-of-CLEVR dataset 에 관한 RN 모델 성능실험에서 매개변수 갱신을 위해 사용한 Adam 기법 외에 Momentum, Nadam 과 같은 기법을 써보도록 하겠다. 또한 dropout 의 비율 조정을 통하여 모델의 오버피팅(overfitting)에 대한 보다 정교한 튜닝(fine tuning)을 해 볼 예정이다.

RN(Relation Networks)은 기존 신경망(neural network) 에 쉽게 추가할 수 있는 사용이 간단한 play and plug 모듈이다. RN 은 이미지(image)나 문장열(series of sentence) 과 같은, 구조화 되지 않은 입력을 받아들여 그 안에 포함된 객체간의 관계에 대해 암묵적으로 추론할 수 있다.

객체간의 암묵적 추론을 하는 RN(Relation Networks) 외에 관계추론(relational reasoning)의 또 다른 핵심적인 부분은 물리적인 장면에서의 미래를 예측하는 것이다. 이와 관련된 연구로서 딥마인드(DeepMind)에서는 이러한 능력을 모방한 Visual Interaction Network(VIN)을 개발하였다. VIN 은 시각적 모듈(visual module)과 물리적 추론 모듈(physical reasoning module)의 두 가지 매커니즘으로 구성되어 있다.

RN 과 VIN 은 관계추론 접근(relational reasoning approach)에 있어 밝은 전망을 보여주고 있다. 이러한 접근법은 확장이 가능하고, 수준 높은 추론모델을 구축하는 등과 같은 수 많은 일에 응용될 수 있으며 인간의 강력하고 유연한 지능에 대한 핵심적 요소를 이해하는데 많은 도움을 줄 수 있다.

이에 추후 연구에서는 VIN (Visual Interaction Networks)에 대해서 알아보고, 더 나아가 본 과제에서 다루었던 RN(Relation Networks)과 비교해 보는 시간을 갖도록 하겠다.

참고문헌

- [1] Adam Santoro, David Raposo, David G.T. Barrett, Mateusz Malinowski, Razvan Pascanu, Peter Battaglia, Timothy Lillicrap. A simple neural network module for relational reasoning, arXiv: 1706.01427v1
- [2] Nicholas Watters, Andrea Tacchetti, Theophane Weber, Razvan Pascanu, Peter Battaglia, Daniel Zoran. Visual Interaction Networks, arXiv:1706.01433v1
- [3] Peter Battaglia, Razvan Pascanu, Matthew Lai, Danilo Jimenez Rezende, et al. Interaction networks for learning about objects, relations and physics. In NIPS, 2016.