

2019년 5월 30일 목요일 뉴랄브리핑

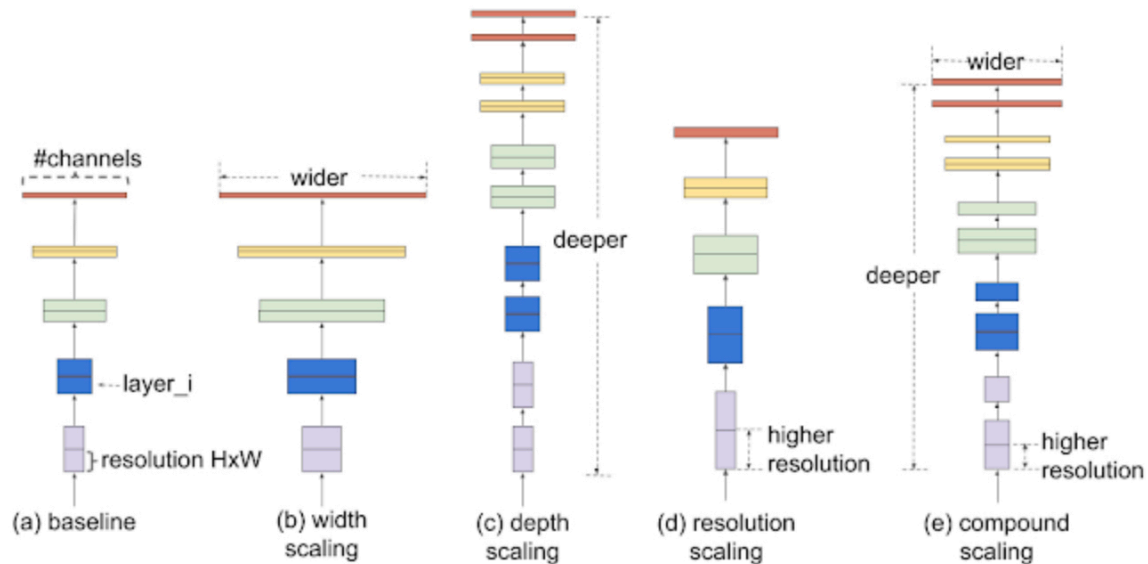
제 0001호 뉴랄브리핑(2019년 5월 30일)

구글블로그

- [EfficientNet: Improving Accuracy and Efficiency through AutoML and Model Scaling](#)
 - The conventional practice for model scaling is 1) to arbitrarily increase the CNN depth or width, or 2) to use larger input image resolution for training and evaluation. While these methods do improve accuracy, they usually require tedious manual tuning, and still often yield suboptimal performance. What if, instead, we could find a more **principled method to scale up a CNN** to obtain better accuracy and efficiency?
 - [본문 중]
 - 논문 이라 하는 것의 인트로는 항상 이렇게 삼단계 스텝으로 말해줍니다.
 - [요즘에는 무얼 하는데]=>[그게 뭐가 문제인데]=>[그래서 좋은걸 보여드립니다]
 - The conventional practice for model scaling is (=요즘의 방법은)
 - 1) to arbitrarily increase the CNN depth or width, or
 - 2) to use larger input image resolution for training and evaluation.
 - While these methods do improve accuracy, they usually require (=안좋은 점이 이러이러한 것인데)
 - tedious manual tuning, and still often yield suboptimal performance.
 - What if, instead, we could find a more principled method to scale up a CNN to obtain better accuracy and efficiency? (=그래서 우리가 제안을 하였다)
 - 그래서 제안하는 바는 EfficientNet인데 이것은 일종의 앙상블 (=a family

of models)이다.

- **[토막상식]** 뉴럴넷 스케일링 (scaling)이란 아래 그림처럼, 너비를 주욱 늘이거나, 깊이를 주욱 늘이거나 해상도를 주욱 올리거나, 셋 다 합쳐서 (compound) 하거나 함.



Comparison of different scaling methods. Unlike conventional scaling methods (b)-(d) that arbitrarily scale a single dimension of the network, our compound scaling method uniformly scales up all dimensions in a principled way.

AWS Machine Learning Blog

- **Powering a search engine with Amazon SageMaker**
 - 애네 블로그는 매번 자기네 제품 (SageMaker) 자랑만 하는듯. 써봤더니 잘났더라 하는 글
 - This is the case of Ibotta, one of a online shopping mall(?) exploiting Amazon SageMaker.
 - In this post, I discuss the architecture of Ibotta's search engine and how we use Amazon SageMaker with other AWS services to integrate real-time ML into the search experience of our mobile application.
 - Foundational search infrastructure => ML-enhanced search infrastructure => Building to scale => Conclusion
 - Conclusion: The AWS services that we rely on enable us to do rapid iteration and testing that would otherwise be out of reach or impossibly slow to implement.

Facebook Research Blog

- [Facebook launches testing and verification research awards at ICSE 2019](#)
 - 소프트웨어 엔지니어링 (친구 <https://www.facebook.com/bntejn>, 번역글: [인공신경망 학습 레시피](#)). [토막상식] software Testing and Verification (TAV) 이라는 것인데, 소프트웨어를 만들어놓고 잘 도나 안 도나 죽나 안죽나 왜 죽나 얼마나 solid하게 개발을 해놓은 것인가 테스트 해보는 분야임.

Machine teaching with Dr. Patrice Simard

- Machine teaching with Dr. Patrice Simard
 - This is a recent post (episode 78) from the series of interviews of ML researchers in Microsoft. MS 리서치에서 발간하는 팟캐스트임. 이 분야의 연구자를 데려다놓고 인터뷰함. 매일 올라오는것은 아닌듯.

[arXiv](#)

- [On the Expressive Power of Deep Polynomial Neural Networks](#)
 - [요약] Polynomial activation에 대하여 논해봄. polynomial neural network라는 것은 weight를 다항식(polynomial)로 매핑하는 것임. (원래는 weight를 다항식이 아니라 NxM matrix로 매핑하는게 일반적임). Dimension of this variety라는 것을 제안하는데, poly-뉴럴넷의 설명력(expressive power)를 측정하는데 사용하는 개념임.
- [Where is the Information in a Deep Neural Network?](#)
 - [요약] 뉴럴넷을 훈련시키면 정보가 weight에 저장되는 것인데, 훈련(training)시킬때 들어있던 정보가 테스트(test)시에는 어디로 어떻게 영향을 미치는지를 아무도 잘 모름. 애초에 네트워크 안에서 정보(information)이라는거를 어떻게 정의하고 또 측정할 것인지도 합의된 바가 없음. 그래서 weight안에서의 정보량이라는 것을 우리가 나름대로 정의하고 또 측정해보는 시도를 하였음. weight안의 information이라는 것을 정확도와 네트워크 복잡도간의 최적의 트레이드오프라고 정의해보면 어떨까 함("the optimal trade-off between accuracy of the network and complexity of the weights"). 그러면 Information에 대한 Shannon의 정의와도 잘 부합하고, 또 네트워크라는 상황에 대해 flexible하게 적용이 됨. 이것이 네트워크의 구조뿐만 아니라 '어떻게 트레이닝되는가'에도 매우 유관하다는 것이

놀라울 따름임. 그래서 덜 복잡한 모델일수록 일반화가 잘 되는 현상도 설명이 되고, 테스트 데이터가 가지는 특징을 트레이닝시에도 변함없이 배운다는 현상도 설명이 됨.

- [Pre-training Graph Neural Networks](#)

- **[토막상식]** Pre-training 모델은 어떤 task가 구체적으로 주어지기 전에 일반적인 개념상식?을 탑재하도록 훈련시켜놓은 모델임. 이것을 갖다가 딱 사용할 시점에 구체적인 task에 맞춰서 레이어 한두개만 추가하는 식으로 요기 붙였다 저기 붙였다 하면서 사용하게 됨. Language Model에서 대표적인 pre-training 모델로는 [BERT](#)가 있음.

- [Solving NP-Hard Problems on Graphs by Reinforcement Learning without Domain Knowledge](#)

- **[토막상식]** NP-Hard 문제란: at least as difficult as NP-complete. 최소한 NPC문제만큼 어려운 문제의 집단임. Traveling Salesman Problem이 대표적임. 바둑도 마찬가지로 NP-hard 문제임. 일반적인 컴퓨터로 합리적인 시간 안에 풀이를 찾기 어려움. 일반적인 컴퓨터란 가능한 모든 솔루션 세트를 전부 일일이 체크하고서야 결론을 내놓기 때문임. 본 논문에서는 그것을 도메인 지식 없이 강화학습으로 풀어보겠다 하는 것.

[레딧](#)

- 오늘 나온 위의 [EfficientNet](#) 에 대한 품평글. 물건사고 댓글놀이 하는것이라 보시면 됨. **[요약]** Brief summary: scaling depth, width, or resolution in a net independently tends not to improve results beyond a certain point. They instead make $\text{depth} = \alpha_\phi$, $\text{width} = \beta_\phi$, and $\text{resolution} = \gamma_\phi$. They then constrain $\alpha \cdot \beta^2 \cdot \gamma^2 \approx c$, and for this paper, $c = 2$. Grid search on a small net to find the values for α, β, γ , then increase ϕ to fit system constraints. **[의의]** 그러니까 뉴럴넷을 잡아늘여서 스케일링(depth, width, resolution)해갖고 성능이 높아지는데는 어떤 상한선 (upper bound)가 있다는 것임. 그래갖고 α (=depth), β (=width), γ (=resolution) 를 따로 보지 말고 식 하나로 조립해갖고 변수 하나 (ϕ)에만 의존하게 만든 다음에 서 작은 넷(small net)에서 α, β, γ 를 grid search로 찾아갖고 고정한 다음에 ϕ 만을 증가시켜서 최적점을 찾는다는 말씀임. 이것의 의의는 무엇이나 하면, 어떤 문제를 푸는 CNN의 하이퍼 파라미터 크기를 사람이 매번 찾는게 아니고 자동으로 짚싸게 찾는다는 의의가 있는 것임

- IEEE가 화웨이 소속 리뷰어들을 모조리 내쫓았다 함여 [\[Link\]](#). 갑질을 하는 중인듯
- N개짜리 데이터셋을 갖다가 NxN개짜리 데이터셋으로 확대시술하는 방법을 개발 해봄 [\[Link\]](#)

트위터 ([Yann LeCun](#))

- [MNIST reborn](#) [\[의의\]](#) 오래도록 사랑받았던(?) MNIST 데이터셋 (60,000 트레이닝 +10,000 테스트)을 복구(restore)및 확장(expand)했음. 원래 있던 것에서 50,000개의 training samples가 추가됨. 그래서 총 12만개짜리 숫자데이터가 됨. 10,000개짜리 테스트셋을 여러번 사용하느라 테스트셋에 overfit되는 경우가 있기도 했는데, 이번 기회에 새로운 데이터셋으로 테스트를 해보길. 관련해서 레딧 댓글[\[Link\]](#)
-

[파비](#)

- [작정하고 쓰는 머신러닝 강의 비판 \(2\)](#)
 - “좀 위험한 발언일 수도 있지만, 지금의 과장된 붐에 혜택을 입은 Python은 길어봐야 3-4년 내에 Julia로 대체될 것이라고 생각한다. “. [\[의의\]](#) 파이썬을 사용하는게 붐인데, 기술의 자세한 내막은 모르고 그냥 덮어놓고 작동한다 우와 만 외치는 것에 대한 경종을 울리는 것임. 미래에 Julia가 쓰일 거라는 보장을 하는건 아니고, 근데 뭔지도 모르고 신나서 쓰는 것에는 거품이 끼기 마련이므로, Julia처럼 사용자가 기반지식이 어느정도 있어야만 사용할 수 있는 언어가 깊이가 있으므로 각광을 받을 것이라는 말임.
- [무슨 데이터가 빅데이터인지 모르는 데이터 전문가들](#)
 - “축척”한 데이터를 쓰는게 아니라, 유저들이 어떤 행동을 해서 어떤 결정을 내리는지를 볼 수 있도록 “추적”한 데이터를 이용해야 한다. “축척”이 아니라 “추적”.
 - 기업들에서 자기네가 인공지능 상품을 만들고 싶은데 어떻게 해야고 물어볼 때 필자가 제일 먼저 묻는 질문은 “초단위로 개인별 행동 데이터 있으십니까?”이다.

기타소식

- Reddit의 [Slack](#) 초대방: 현재(20190530)기준 2575명이 한 방에 모여서 머신러닝 대화를 하는 중. 각 분야별로 따로 채널이 있음. 주워들어 배울 것이 매우 많을듯

키워드

- [AutoML](#)

Useful links

- [Microsoft Machine Teaching Group](#)
- [Microsoft Research Potcast](#)