

14기
텍스트 세미나

ToBig's 14기 민거홍

Practical Tips for Final Projects

Contents

Unit 01 | Starting Research

Unit 02 | Review of gated neural sequence models

Unit 03 | A couple of MT topics

Unit 04 | Research Evaluation, etc.

14기
텍스트 세미나

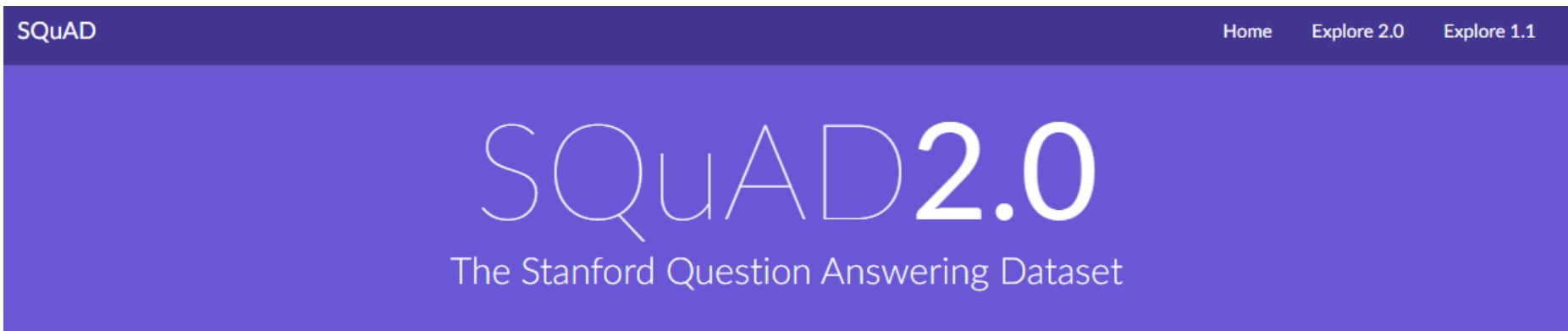
ToBig's 14기 민거홍

01. Starting Research

Unit 01 | Starting Research

CS 224n Final Project

- **CS224n 파이널 프로젝트**는 CS224n 수업의 큰 비중을 차지하며 (전체 성적의 43%), 총 4부분으로 나누어져 있습니다.
- 나뉜 부분: 프로젝트 프로포절 (5%), 프로젝트 마일스톤 리포트 (5%), 포스터 (3%), 리포트 (30%)
- 1~3인이 한 조를 이뤄서 같이 프로젝트를 진행하고, 언어는 PyTorch를 권장한다고 합니다!
- 프로젝트 주제는 1) SQuAD Q&A 프로젝트나 2) 자유 주제 중 하나를 골라 진행할 수 있습니다.



Unit 01 | Starting Research

SQuAD란?

- 스탠포드 대학의 NLP 그룹에서 크라우드 소싱을 통해 만든 위키피디아 아티글에 대한 107,785개의 질문-대답 데이터셋입니다!
- 현재 2.0버전까지 있고, 2.0버전에는 50,000개의 답변을 할 수 없는 질문들까지 추가합니다.
- 한국에는 KorQuAD가 있습니다! (자세한 내용은 10주차 강의자님이 설명하실 것 같습니다 ㅎㅎ)

T: [Bill] Aken, adopted by Mexican movie actress Lupe Mayorga, grew up in the neighboring town of Madera and his song chronicled the hardships faced by the migrant farm workers he saw as a child.

Q: In what town did Bill Aiken grow up?

A: Madera [But Google's BERT says <No Answer>!]



Unit 01 | Starting Research

누가 주어진 파이널 프로젝트를 해야할까요?

- 아직까지 많은 연구 경험을 갖고 있지 않은 사람
- 어떤 것을 해야 하는지 모르는 사람
- 진행 방향에 대한 도움을 받고 싶은 사람 & 리더보드가 있으면 좋겠다는 사람!

Leaderboard

SQuAD2.0 tests the ability of a system to not only answer reading comprehension questions, but also abstain when presented with a question that cannot be answered based on the provided paragraph.

Rank	Model	EM	F1
	Human Performance Stanford University (Rajpurkar & Jia et al. '18)	86.831	89.452
1 Apr 06, 2020	SA-Net on Albert (ensemble) QIANXIN	90.724	93.011
2 May 05, 2020	SA-Net-V2 (ensemble) QIANXIN	90.679	92.948

- EM Score: Exact Match; Reference 정답과 같은 답을 한 비율
- F1 Score [1]:

F1 Score : 정밀도와 재현율의 조화 평균

$$F_1 = 2 \cdot \frac{1}{\frac{1}{\text{recall}} + \frac{1}{\text{precision}}} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

한마디로 표현해보면

정확도 : 예측이 정답과 얼마나 정확한가?

정밀도 : 예측한 것중에 정답의 비율은?

재현율 : 찾아야 할 것중에 실제로 찾은 비율은?

F1 Score : 정밀도와 재현율의 평균

Unit 01 | Starting Research

이 수업은 모든 학생들한테 중요합니다!

- 전반적으로 연구를 어떻게 하는 지에 대해 아는 건 중요합니다.
- 또한, 이번 수업에서는 연구에서 중요시하는 다양한 요소들에 대해 알아볼 예정입니다
 - Baseline
 - Benchmark
 - Evaluation (결과에서 유의미한 결론 유추하기)
 - Error Analysis (오류 분석)
 - Paper Writing (논문 작성법)

Unit 01 | Starting Research

연구는 왜 중요할까요???



이상



현실

- 새로운 지식을 만듭니다!
- (인류의) 삶의 질을 높입니다!!
- 전반적으로 사회를 발전시킵니다!
- 중요한 문제에 대한 관심을 끌어 모읍니다.

Unit 01 | Starting Research

좋은 연구란?

- 연구는 새로운 지식을 창출하고 그 지식을 효과적으로 전달하는 것을 뜻합니다.
- 연구는
 - Novel (새로워야하고)
 - Generalizable (일반화되어야하고)
 - Valuable (새로운, non-trivial한 insight를 제공해야 하고)
 - Valid (검증될 수 있어야 합니다)

Unit 01 | Starting Research

연구 주제를 찾는 법...

두가지 기본적인 시작점

- **[Nails]** 문제를 생각해보고, 기존의 연구들이 이 문제에 대해 제시한 해법들보다 더 좋은 해법을 찾는 것
- **[Hammers]** 기존의 technical approach에 대해 생각해보고, 이 approach를 개선할 수 있는 방법이나 추가할 수 있는 내용들을 생각하는 것
- 많은 사람들이 연구하고 있는 주제를 벗어나라 (Herb Simon, Turing award winner, Nobel prize winner)

Unit 01 | Starting Research

프로젝트 종류

대체적으로 5가지의 카테고리 안에 들어간다:

- 모델의 application을 찾아보고 어떻게 효율적으로 적용할 지 찾거나
- complex neural architecture를 구현해보고 특정 데이터에 대한 performance를 측정해보거나
- new, variant NN model을 구현해서 실험적인 데이터로 향상을 보여주거나
- 모델의 동작법을 분석하거나
- rare theoretical project, 그냥 개쩌는 걸 가져오거나..?

Unit 01 | Starting Research

모델의 application을 찾아보고 어떻게 효율적으로 적용할 지 찾는 연구

**Deep Poetry: Word-Level and Character-Level
Language Models for Shakespearean Sonnet
Generation**

Stanley Xie, Ruchir Rastogi and Max Chang

Gated LSTM

Thy youth 's time and face his form shall cover?
Now all fresh beauty, my love there
Will ever Time to greet, forget each, like ever decease,
But in a best at worship his glory die.

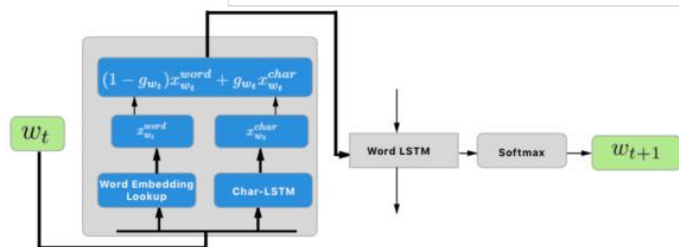


Figure 1: Architecture of the Gated LSTM

Unit 01 | Starting Research

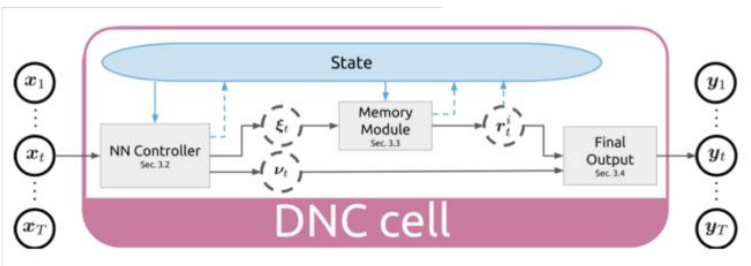
complex neural architecture를 구현해보고 특정 데이터에 대한 performance를 측정하는 연구

Implementation and Optimization of Differentiable Neural Computers

Carol Hsin

Graduate Student in Computational & Mathematical Engineering

We implemented and optimized Differentiable Neural Computers (DNCs) as described in the Oct. 2016 DNC paper [1] on the bAbI dataset [25] and on copy tasks that were described in the Neural Turing Machine paper [12]. This paper will give the reader a better understanding of this new and promising architecture through the documentation of the approach in our DNC implementation and our experience of the challenges of optimizing DNCs.



Unit 01 | Starting Research

new, variant NN model을 구현해서 실험적인 데이터로 향상을 보여주는 연구

Improved Learning through Augmenting the Loss

Hakan Inan

inanh@stanford.edu

Khashayar Khosravi

khosravi@stanford.edu

We present two improvements to the well-known Recurrent Neural Network Language Models(RNNLM). First, we use the word embedding matrix to project the RNN output onto the output space and already achieve a large reduction in the number of free parameters while still improving performance. Second, instead of merely minimizing the standard cross entropy loss between the prediction distribution and the "one-hot" target distribution, we minimize an additional loss term which takes into account the inherent metric similarity between the target word and other words. We show with experiments on the Penn Treebank Dataset that our proposed model (1) achieves significantly lower average word perplexity than previous models with the same network size and (2) achieves the new state of the art by using much fewer parameters than used in the previous best work.

Unit 01 | Starting Research

새로운 연구!

Word2Bits - Quantized Word Vectors

Maximilian Lam

maxlam@stanford.edu

Abstract

Word vectors require significant amounts of memory and storage, posing issues to resource limited devices like mobile phones and GPUs. We show that high quality quantized word vectors using 1-2 bits per parameter can be learned by introducing a quantization function into Word2Vec. We furthermore show that training with the quantization function acts as a regularizer. We train word vectors on English Wikipedia (2017) and evaluate them on standard word similarity and analogy tasks and on question answering (SQuAD). Our quantized word vectors not only take 8-16x less space than full precision (32 bit) word vectors but also outperform them on word similarity tasks and question answering.

Unit 01 | Starting Research

연구는 어디서부터 시작해야 할까요?

열심히 논문/레퍼런스들을 읽는다...

- NLP논문에 대한 ACL Anthology 참조 (<https://aclanthology.info>)
- 주요 ML 컨퍼런스들의 논문들을 참조한다 (NeurIPS, ICML, ICLR, ...)
- 기존의 프로젝트들을 참조한다
- 출판 전 논문들을 본다 (<https://arxiv.org>)

더 좋은 방법은 우리 세상 속에 존재하는 재미있는 문제들을 찾는 것!

Unit 01 | Starting Research


Arxiv Sanity Preserver (스탠포드 사람이 만든거)

Top papers mentioned on Twitter over last day:

Shaping the Narrative Arc: An Information-Theoretic Approach to Collaborative Dialogue
Kory W. Mathewson, Pablo Samuel Castro, Colin Cherry, George Foster, Marc G. Bellemare
1/31/2019 cs.HC | cs.AI | cs.CL | cs.LG
20 pages, 9 figures

1901.11528v1 pdf
[show similar](#) | [discuss](#)

We consider the problem of designing an artificial agent capable of interacting with humans in collaborative dialogue to produce creative, engaging narratives. In this task, the goal is to establish universe details, and to collaborate on an interesting story in that universe, through a series of natural dialogue exchanges. Our model can augment any probabilistic conversational agent by allowing it to reason about universe information established and what potential next utterances might reveal. Ideally, with each utterance, agents would reveal just enough information to add specificity and reduce ambiguity without limiting the conversation. We empirically show that our model allows control over the rate at which the agent reveals information and that doing so significantly improves accuracy in predicting the next line of dialogues from movies. We close with a case-study with four professional theatre performers, who preferred interactions with our model-augmented agent over an unaugmented agent.

17 tweets: 

Learning and Evaluating General Linguistic Intelligence
Dani Yogatama, Cyprien de Masson d'Autume, Jerome Connor, Tomas Kocisky, Mike Chrzanowski, Lingpeng Kong, Angeliki Lazaridou, Wang Ling, Lei Yu, Chris Dyer, Phil Blunsom
1/31/2019 cs.LG | cs.CL | stat.ML

1901.11373v1 pdf
[show similar](#) | [discuss](#)

Unit 01 | Starting Research

State of the Art (최첨단) 기술 도장깨기??

Machine Translation

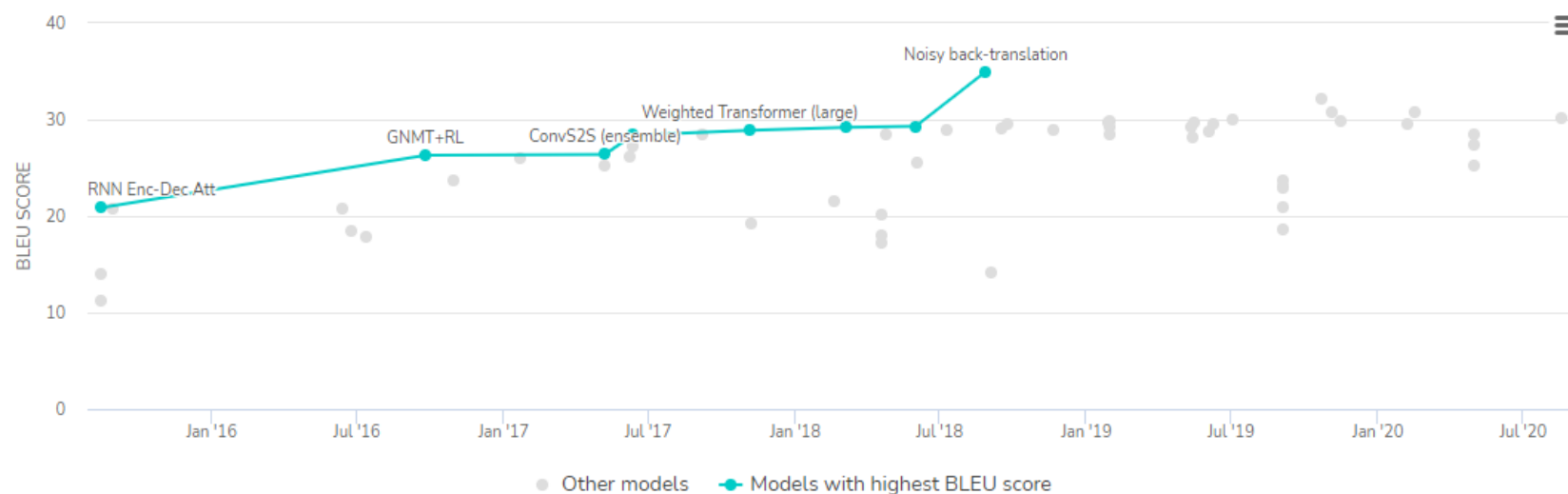
223 papers with code · Natural Language Processing

Machine translation is the task of translating a sentence in a source language to a different language.

State-of-the-art leaderboards

Dataset	Best Method	Paper title	Paper	Code
WMT2014 English-French	Transformer Big + BT	Understanding Back-Translation at Scale	Paper	Code
WMT2014 English-German	Transformer Big + BT	Understanding Back-Translation at Scale	Paper	Code
IWSLT2015 German-English	Transformer	Attention Is All You Need	Paper	Code
WMT2016 English-Romanian	ConvS2S BPE40k	Convolutional Sequence to Sequence Learning	Paper	Code

Machine Translation on WMT2014 English-German



Unit 01 | Starting Research

NLP연구에서 매우매우 중요한 것: 데이터

적절한 데이터를 찾는 건 매우 중요하다 (최소 10,000개의 레이블 된 데이터)

- 1) 직접 데이터를 구하기
- 2) 기존의 프로젝트에서 구한 데이터 / 기업에서 갖고 있는 데이터
- 3) 하지만, 대다수의 사람들은 공개되고, 잘 관리된 dataset을 활용할 것이다.

실행가능한, 적절한 Task를 찾는 것도 중요하다.

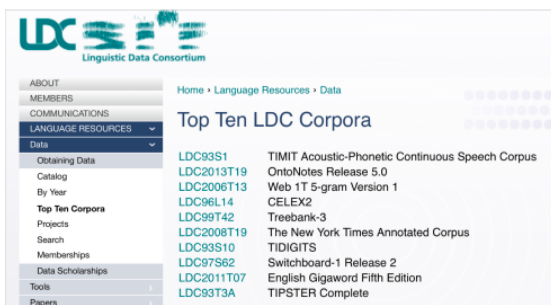
Task를 찾으면 이를 평가할 수 있는 자동화된 metric도 있어야 한다!

Unit 01 | Starting Research

데이터는 어디서?

Linguistic Data Consortium

- <https://catalog.ldc.upenn.edu/>
- Stanford licenses data; you can get access by signing up at: <https://linguistics.stanford.edu/resources/resources-corpora>
- Treebanks, named entities, coreference data, lots of newswire, lots of speech with transcription, parallel MT data
 - Look at their catalog
 - Don't use for non-Stanford purposes!



Machine translation

- <http://statmt.org>
- Look in particular at the various WMT shared tasks

Sitemap

- SMT Book
- Research Survey Wiki
- Moses MT System
- Europarl Corpus
- News Commentary Corpus
- Online Evaluation
- Online Moses Demo
- Translation Tool
- WMT Workshop 2014
- WMT Workshop 2013
- WMT Workshop 2012
- WMT Workshop 2011
- WMT Workshop 2010
- WMT Workshop 2009
- WMT Workshop 2008
- WMT Workshop 2007
- WMT Workshop 2006

Statistical Machine Translation

This website is dedicated to research in statistical machine translation, i.e. the translation of text from one human language to another by a computer that learned how to translate from vast amounts of translated text.

Introduction to Statistical MT Research

- [The Mathematics of Statistical Machine Translation](#) by Brown, Della Petra, Della Pietra, and Mercer
- [Statistical MT Handbook](#) by Kevin Knight
- [SMT Tutorial \(2003\)](#) by Kevin Knight and Philipp Koehn
- ESSLI Summer Course on SMT (2005), [day1](#), [2](#), [3](#), [4](#), [5](#) by Chris Callison-Burch and Philipp Koehn.
- [MT Archive](#) by John Hutchins, electronic repository and bibliography of articles, books and papers on topics in machine translation and computer-based translation tools

Unit 01 | Starting Research

데이터는 어디서?

온라인에는 많은 데이터셋들이 존재한다.

- 캐글!
- 기존의 연구 논문
- NLP 데이터셋를 모아둔 곳 (<https://machinelearningmastery.com/datasets-natural-languageprocessing/>, <https://github.com/niderhoff/nlp-datasets>)
- 한국은



14기
텍스트 세미나

ToBig's 14기 민거홍

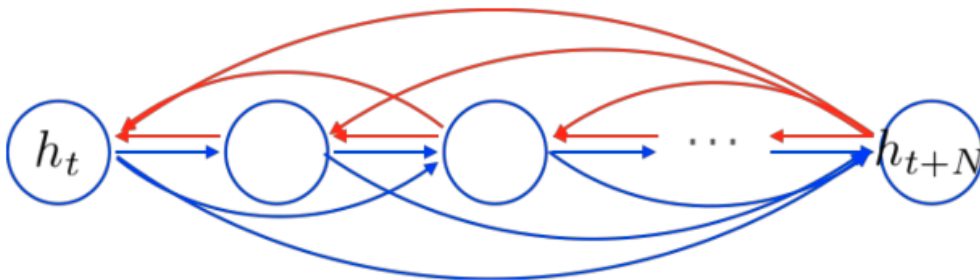
02. Review of gated neural sequence models

Unit 02 | Review of gated neural sequence models

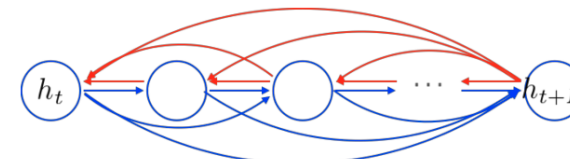
Gated Recurrent Units

간단히 정리하면

- 직관적으로 RNN을 이해하는 게 중요하다. (RNN measure the influence of past & future)
- Vanishing gradient 문제는 아주 문제있다. Gradient가 0이 되면 뭐가 문제인지 모른다. (t, t+n dependency 문제인지, 아니면 파라미터 configuration 문제인지)
- Backpropagation을 위해서 shortcut connection을 만들거나 adaptive하게 만들 수 있다.



Perhaps we can create *adaptive* shortcut connections.



$$f(h_{t-1}, x_t) = u_t \odot \tilde{h}_t + (1 - u_t) \odot h_{t-1}$$

Candidate Update $\tilde{h}_t = \tanh(W[x_t] + U h_{t-1} + b)$

Update gate $u_t = \sigma(W_u[x_t] + U_u h_{t-1} + b_u)$

14기
텍스트 세미나

ToBig's 14기 민거홍

03. A couple of MT Topics

Unit 03 | A couple of MT Topics

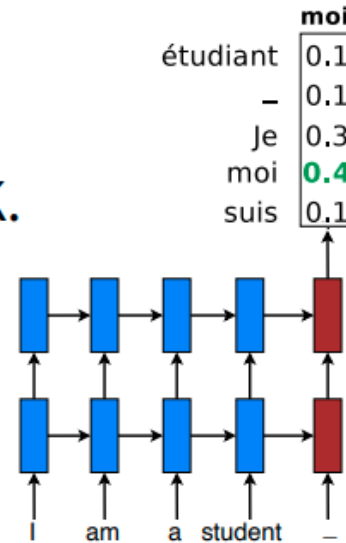
Word generation problem (Softmax가 비싸서 일어난다..)

- Word generation problem
 - Vocab used are usually modest: 50K.

The ecotax portico in Pont-de-Buis
Le portique écotaxe de Pont-de-Buis



The <unk> portico in <unk>
Le <unk> <unk> de <unk>



해결 방법

- Hierarchical Softmax
- Large Vocabulary Set을 몇 개의 모델들로 나눠서 train한 후에 알맞은 번역 거르기
- Attention! 그냥 사전을 검색해봐도 된다.
- 추후에 배울 Word pieces, char models 같은 것들

Unit 03 | A couple of MT Topics

MT Evaluation

수동?!

- Adequacy and Fluency (5점 / 7점 척도)
- 오류 분석
- 번역 순위 매기기

자동!

- BLEU (Bilingual Evaluation Understudy)

Unit 03 | A couple of MT Topics

BLEU

BLEU

BLEU(Bilingual Evaluation Understudy) score란 성과지표로 데이터의 X가 순서정보를 가진 단어 들(문장)로 이루어져 있고, y 또한 단어들의 시리즈(문장)로 이루어진 경우에 사용되며, 번역을 하는 모델에 주로 사용된다. 3가지 요소를 살펴보자.

- n-gram을 통한 순서쌍들이 얼마나 겹치는지 측정(precision)
- 문장길이에 대한 과적합 보정 (Brevity Penalty)
- 같은 단어가 연속적으로 나올때 과적합 되는 것을 보정(Clipping)

$$BLEU = \min(1, \frac{\text{output length}(\text{예측 문장})}{\text{reference length}(\text{실제 문장})}) (\prod_{i=1}^4 \text{precision}_i)^{\frac{1}{4}}$$

- 보통 BLEU4를 활용하는데 n-gram 길이가 4까지 계산한다.
- 4단어 미만 문장은 Bleu 0 점!
- 레퍼런스 번역 대비 번역의 정확성을 비교하는 지표

Few human translations will attain a score of 1, since this would indicate that the candidate is identical to one of the reference translations. For this reason, it is not necessary to attain a score of 1. Because there are more opportunities to match, adding additional reference translations will increase the BLEU score.[\[4\]](#)

Unit 03 | A couple of MT Topics

예제

Translation task를 생각해보자. 영어 sentence 을 모델에 넣으면, 그 모델은 한글 sentence 을 출력한다. 그럼 그 한글 sentence 이 얼마나 잘 예측된 것인지 알려면 사람이 정확히 번역한 문장 (true sentence)과 비교해 보면 된다.

1. n-gram(1~4)을 통한 순서쌍들이 얼마나 겹치는지 측정(precision)

- 예측된 sentence : 빛이 썬는 노인은 완벽한 어두운곳에서 잠든 사람과 비교할 때 강박증이 심해질 기회가 훨씬 높았다
- true sentence : 빛이 썬는 사람은 완벽한 어둠에서 잠든 사람과 비교할 때 우울증이 심해질 가능성이 훨씬 높았다

$$\bullet \text{ 1-gram precision: } \frac{\text{일치하는 1-gram의 수 (예측된 sentence중에서)}}{\text{모든 1-gram 쌍 (예측된 sentence중에서)}} = \frac{10}{14}$$

$$\bullet \text{ 2-gram precision: } \frac{\text{일치하는 2-gram의 수 (예측된 sentence중에서)}}{\text{모든 2-gram 쌍 (예측된 sentence중에서)}} = \frac{5}{13}$$

$$\bullet \text{ 3-gram precision: } \frac{\text{일치하는 3-gram의 수 (예측된 sentence중에서)}}{\text{모든 3-gram 쌍 (예측된 sentence중에서)}} = \frac{2}{12}$$

$$\bullet \text{ 4-gram precision: } \frac{\text{일치하는 4-gram의 수 (예측된 sentence중에서)}}{\text{모든 4-gram 쌍 (예측된 sentence중에서)}} = \frac{1}{11}$$

$$\left(\prod_{i=1}^4 precision_i \right)^{\frac{1}{4}} = \left(\frac{10}{14} \times \frac{5}{13} \times \frac{2}{12} \times \frac{1}{11} \right)^{\frac{1}{4}}$$

Unit 03 | A couple of MT Topics

2. 같은 단어가 연속적으로 나올때 과적합 되는 것을 보정(Clipping)

위 예제에서 단어 단위로 n-gram을 할 경우 보정할 것이 없지만, 영어의 한 예제에서 1-gram precision를 구하면, 예측된 문장에 중복된 단어들(**the** :3, **more** :2)이 있다. 이를 보정하기 위해 **true sentence**에 있는 중복되는 단어의 max count(**the** :2, **more** :1)를 고려하게 된다. (Clipping). 다른 n-gram도 같은 방식으로 처리하면 된다.

- **예측된 sentence** : **The more** decomposition **the more** flavor **the** food has
- **true sentence** : **The more the** merrier I always say

- 1-gram precision: $\frac{\text{일치하는 1-gram의 수 (예측된 sentence중에서)}}{\text{모든 1-gram쌍 (예측된 sentence중에서)}} = \frac{5}{9}$
- (보정 후) 1-gram precision: $\frac{\text{일치하는 1-gram의 수 (예측된 sentence중에서)}}{\text{모든 1-gram쌍 (예측된 sentence중에서)}} = \frac{3}{9}$

Unit 03 | A couple of MT Topics

- Brevity Penalty
 - Can't just type out single word "the" (precision 1.0!)

3. 문장길이에 대한 과적합 보정 (Brevity Penalty)

같은 예제에 문장길이에 대한 보정계수를 구하면 다음과 같다.

- **예측된 sentence**: 빛이 썩는 노인은 완벽한 어두운곳에서 잠들
- **true sentence**: 빛이 썩는 사람은 완벽한 어둠에서 잠든 사람과 비교할 때 우울증이 심해질 가능성이 훨씬 높았다

$$\min\left(1, \frac{\text{예측된 sentence의 길이(단어의 갯수)}}{\text{true sentence의 길이(단어의 갯수)}}\right) = \min\left(1, \frac{6}{14}\right) = \frac{3}{7}$$

Machine translation:

The American [?] international airport and its the office all receives one calls self the sand Arab rich business [?] and so on electronic mail, which sends out ; The threat will be able after public place and so on the airport to start the biochemistry attack , [?] highly alerts after the maintenance.

Unit 03 | A couple of MT Topics

BLEU score

- 아래 예시의 BLEU score를 계산하면 다음과 같다.
- 예측된 sentence**: 빛이 썩는 노인은 완벽한 어두운곳에서 잠든 사람과 비교할 때 강박증이 심해질 기회가 훨씬 높았다
- true sentence**: 빛이 썩는 사람은 완벽한 어둠에서 잠든 사람과 비교할 때 우울증이 심해질 가능성이 훨씬 높았다

$$\begin{aligned} BLEU &= \min(1, \frac{\text{output length}(\text{예측 문장})}{\text{reference length}(\text{실제 문장})}) (\prod_{i=1}^4 \text{precision}_i)^{\frac{1}{4}} \\ &= \min(1, \frac{14}{14}) \times (\frac{10}{14} \times \frac{5}{13} \times \frac{2}{12} \times \frac{1}{11})^{\frac{1}{4}} \end{aligned}$$

14기
텍스트 세미나

ToBig's 14기 민거홍

04. Research Evaluation, etc.

Unit 04 | Research Evaluation, etc.

연구는 어떻게?

1. Summarization이라는 task를 정하자

2. Dataset을 정하자

1. search for academic datasets
 1. newsroom summarization dataset을 사용한다!
2. define your own data (harder one)
 1. 트윗, 블로그 포스트, 뉴스들도 데이터셋이 될 수 있다.

3. Dataset hygiene (preprocessing)

1. 시작할 때 바로 테스트셋같은 것은 분리해두자

4. 평가 Metric도 정하자!

1. 온라인에서 활용 가능한 메트릭을 찾아보자
2. summarization은 Rouge*같은 것도 쓸 수 있다.

* (Recall-Oriented Understudy for Gisting Evaluation)

5. Baseline을 정하자

1. 너무 잘나오면 문제가 너무 쉬웠던 거다. 다시하자

6. Always be close to your data!

1. visualize the dataset
2. collect summary dataset
3. look at errors
4. analyze how different hyperparameters affect performance

7. 다른 시도도 많이 해보자

Unit 04 | Research Evaluation, etc.

보통 데이터셋들...

보통 데이터셋들은 train/dev/test 데이터셋이 있다

- 연구자 윤리를 위해서 개발이 끝난 후에만 test set을 사용하자
- 이렇게 분리하면 데이터셋이 꽤 크다는 뜻이다
- 만약 데이터 셋이 없으면 알아서 분리한다!
- 이렇게 데이터 셋이 분리하면 모든 사람들이 같은 데이터를 가지고 평가를 진행해서 좋지만, **만약 하나의 데이터 셋에만 이상 값들이 많이 존재하면 문제가 생길 수 있다.**

Unit 04 | Research Evaluation, etc.

보통 데이터셋들...

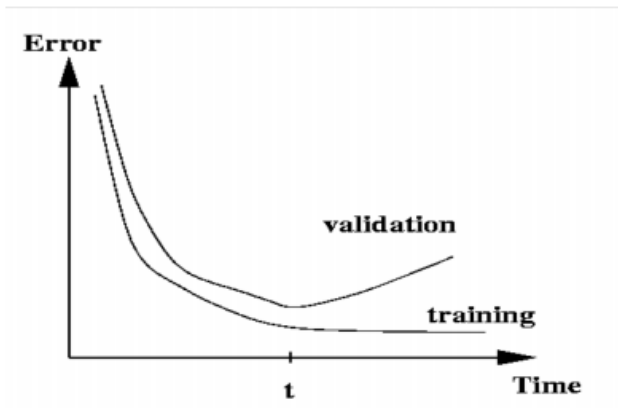
- **Training Set**에서 트레이닝을 진행한다.
- **Tuning Set**에서 하이퍼파라미터 튜닝을 한다. (Tuning Set은 하이퍼파라미터의 Training Set)
- **Dev Set**에서 트레이닝이 잘 되었는지 확인한다. 단, Dev Set을 많이 쓰면 오버피팅이 있을 수 있기 때문에 **dev2 set**도 두면 좋다.
- **마지막에만 test set을 쓴다! 한두번만 쓰는 게 좋다고 합니다...**
- 모든 데이터셋들은 독립적이어야 합니다!

Unit 04 | Research Evaluation, etc.

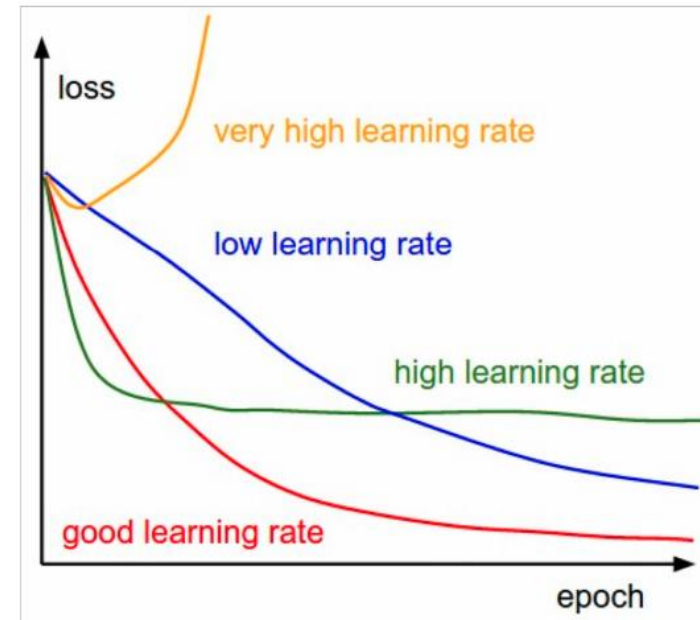
오버피팅 주의! Learning Rate 주의!

트레이닝을 많이 하면 오버피팅이 일어날 수 있다!

The way to monitor and avoid problematic overfitting is using **independent** validation and test sets ...



적절한 Learning Rate을 설정해주어야 한다.



Unit 04 | Research Evaluation, etc.

연구 방법론 (1)

Work Incrementally! 한 단계씩 연구를 진행하는 게 좋다.

- 처음에는 아주 간단한 모델에서부터 시작한다.
- 만약 잘 작동하면 간단한 모델에 많은 것들을 추가한다.

데이터셋도 마찬가지다

- 처음에는 아주 작은 데이터셋에서부터 시작하는 게 좋다. (수업에서는 8개 데이터가 좋다고 한다)
- 인위로 제작한 데이터도 좋다. 오류나 버그를 보기가 쉬울 것이다
- 작은 데이터로 시작하면 100%를 달성할 수 있도록 한다.

Unit 04 | Research Evaluation, etc.

연구 방법론 (2)

데이터셋의 크기를 점점 더 키우자.

- 데이터셋을 키워도 100%에 가까운 정확도가 나오는 게 좋다. 안 나오면 모델을 바꿀 시간이다.
- Training Set 오버피팅은 DL에서는 괜찮다.

These models are usually good at generalizing because of the way distributed representations share statistical strength regardless of overfitting to training data

오류 분석을 진행하자

- Summary statistics, 모델의 출력값, 오류 분석을 진행하자
- 하이퍼파라미터 튜닝이 성공적인 Nnets 모델에 제일 중요한 요소 중 하나다!

Unit 04 | Research Evaluation, etc.

기타 (RNN 트레이닝은?)

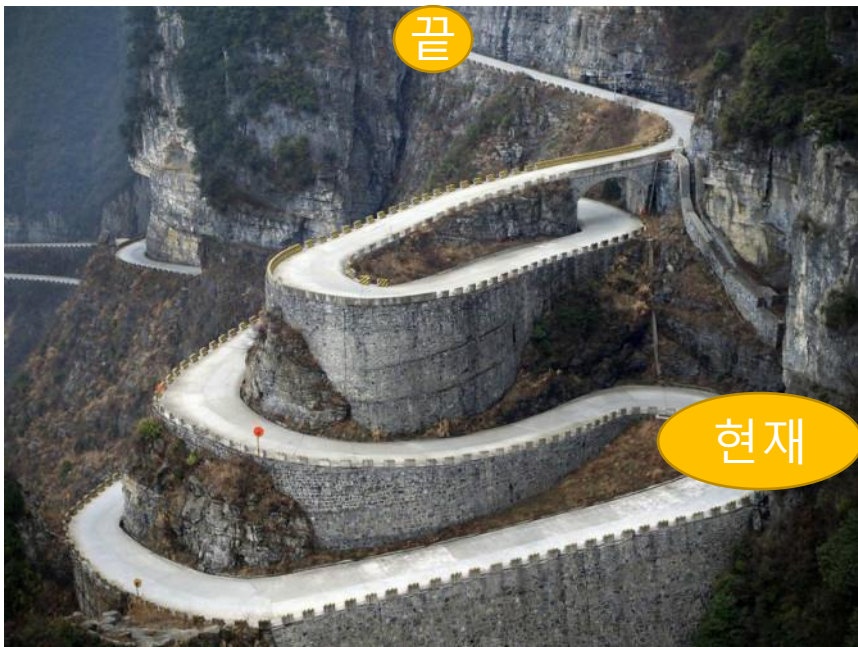
RNN을 학습할 땐 아래와 같은 내용도 살펴보자

1. LSTM이나 GRU를 써보자
2. orthogonal하게 recurrent matrices를 초기화하자
3. 다른 matrices들을 sensible scale로 만들자
4. forget gate bias를 1로 두자 (default to remembering이다)
5. adaptive learning rate를 사용하자
6. clip the norm of the gradient. (1 ~ 5가 적당한 threshold이다)
7. dropout을 vertically하게 적용시키거나 Bayesian Dropout을 사용하자
8. 학습은 좀 기다리자

Unit 04 | Research Evaluation, etc.

끝맺음..

NLP 컨퍼런스도 하는데 할 게 많아보이네요ㅠ ㅎㅎ;;



Reference ~

Reference

- [1] <http://blog.naver.com/PostView.nhn?blogId=sogangori&logNo=220986343741>
- BLEU 관련 <https://donghwa-kim.github.io/BLEU.html>
- 전반적으로 <https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1194/slides/cs224n-2019-lecture09-final-projects.pdf>
- 전반적으로 <https://jeongukjae.github.io/posts/cs224n-lecture-9-practical-tips-for-final-projects/>

Q & A

들어주셔서 감사합니다.