

Evaluations

Ki Hyun Kim

nlp.with.deep.learning@gmail.com

Experimental Setup

- Preprocessing
 - en: MosesTokenizer + BPE
 - Vocabulary Size: 20,525
 - ko: Mecab + BPE
 - Vocabulary Size: 29,411

- Dataset

set	lang	# lines	# tokens	# characters
train	en	1,200,000	43,700,390	367,477,362
	ko	1,200,000	39,066,127	344,881,403
validation	en	200,000	7,286,230	61,262,147
	ko	200,000	6,516,442	57,518,240
validation-1000	en	1,000	36,307	305,369
	ko	1,000	32,282	285,911
test-1000	en	1,000	35,686	298,993
	ko	1,000	31,720	280,126

- Hyper-parameters

parameter	Seq2seq
batch_size	64*4
word_vec_size	512
hidden_size	768
n_layers	4
n_epochs	40
dsl_n_warmup_epochs	30
dsl_lambda	1e-2
max_grad_norm	1e+8 → 1.0
optimizer	Adam (1e-3)

Equation Comparison

- Paper

$$G_f = \nabla_{\theta_{xy}} (1/m) \sum_{j=1}^m [\ell_1(f(x_j; \theta_{xy}), y_j) + \lambda_{xy} \ell_{\text{duality}}(x_j, y_j; \theta_{xy}, \theta_{yx})];$$

$$G_g = \nabla_{\theta_{yx}} (1/m) \sum_{j=1}^m [\ell_2(g(y_j; \theta_{yx}), x_j) + \lambda_{yx} \ell_{\text{duality}}(x_j, y_j; \theta_{xy}, \theta_{yx})];$$

- Ours

$$\theta_{x \rightarrow y} \leftarrow \theta_{x \rightarrow y} - \eta \nabla_{\theta_{x \rightarrow y}} \mathcal{L}(\theta_{x \rightarrow y})$$

$$\mathcal{L}(\theta_{x \rightarrow y}) = \sum_{i=1}^N \left(-\log P(y_i | x_i; \theta_{x \rightarrow y}) + \lambda \times \left((\log P(x_i; \phi) + \log P(y_i | x_i; \theta_{x \rightarrow y})) - (\log P(y_i; \psi) + \log P(x_i | y_i; \theta_{y \rightarrow x})) \right)^2 \right)$$

$$\theta_{y \rightarrow x} \leftarrow \theta_{y \rightarrow x} - \eta \nabla_{\theta_{y \rightarrow x}} \mathcal{L}(\theta_{y \rightarrow x})$$

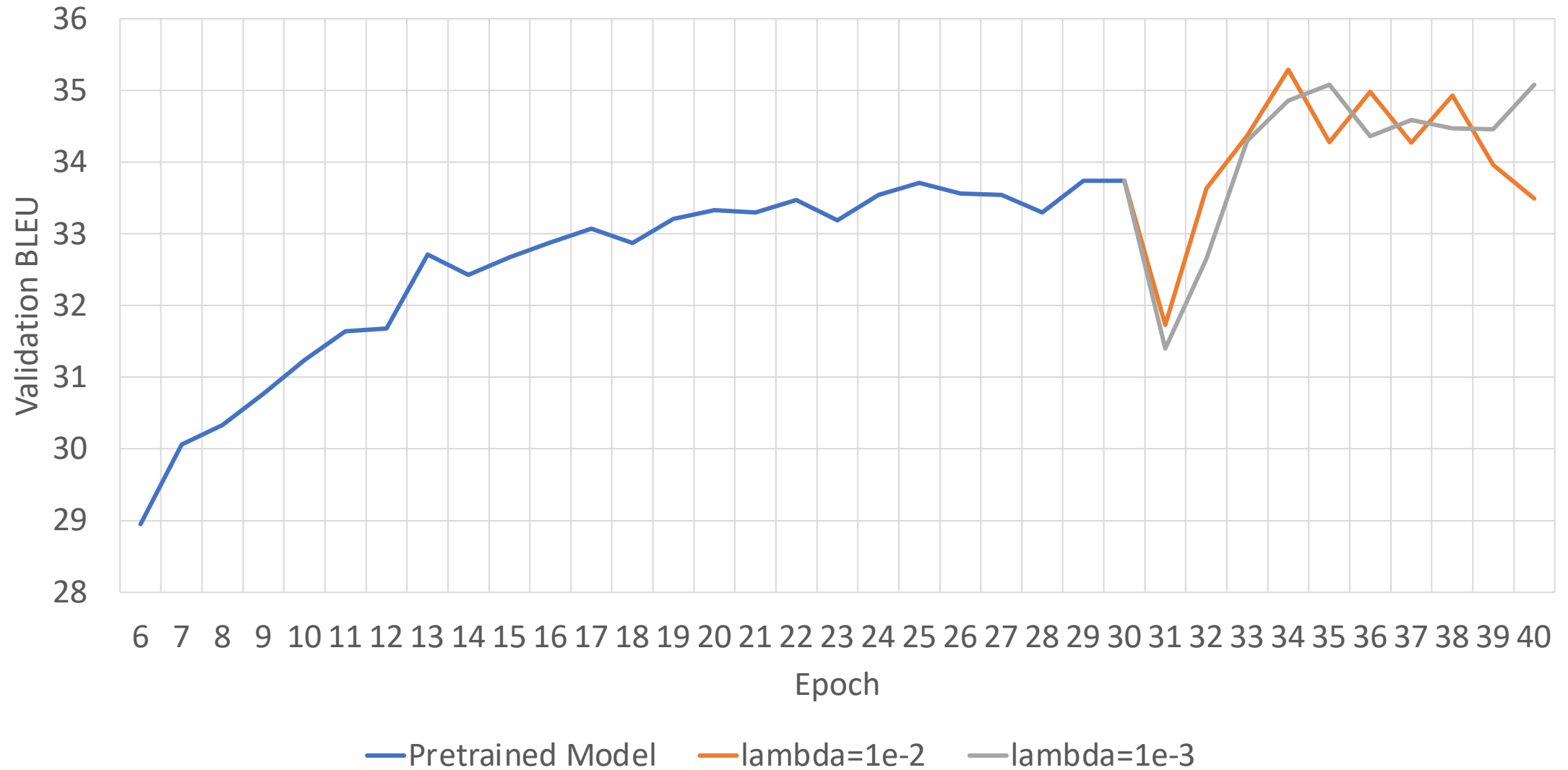
$$\mathcal{L}(\theta_{y \rightarrow x}) = \sum_{i=1}^N \left(-\log P(x_i | y_i; \theta_{y \rightarrow x}) + \lambda \times \left((\log P(x_i; \phi) + \log P(y_i | x_i; \theta_{x \rightarrow y})) - (\log P(y_i; \psi) + \log P(x_i | y_i; \theta_{y \rightarrow x})) \right)^2 \right)$$

Test BLEU

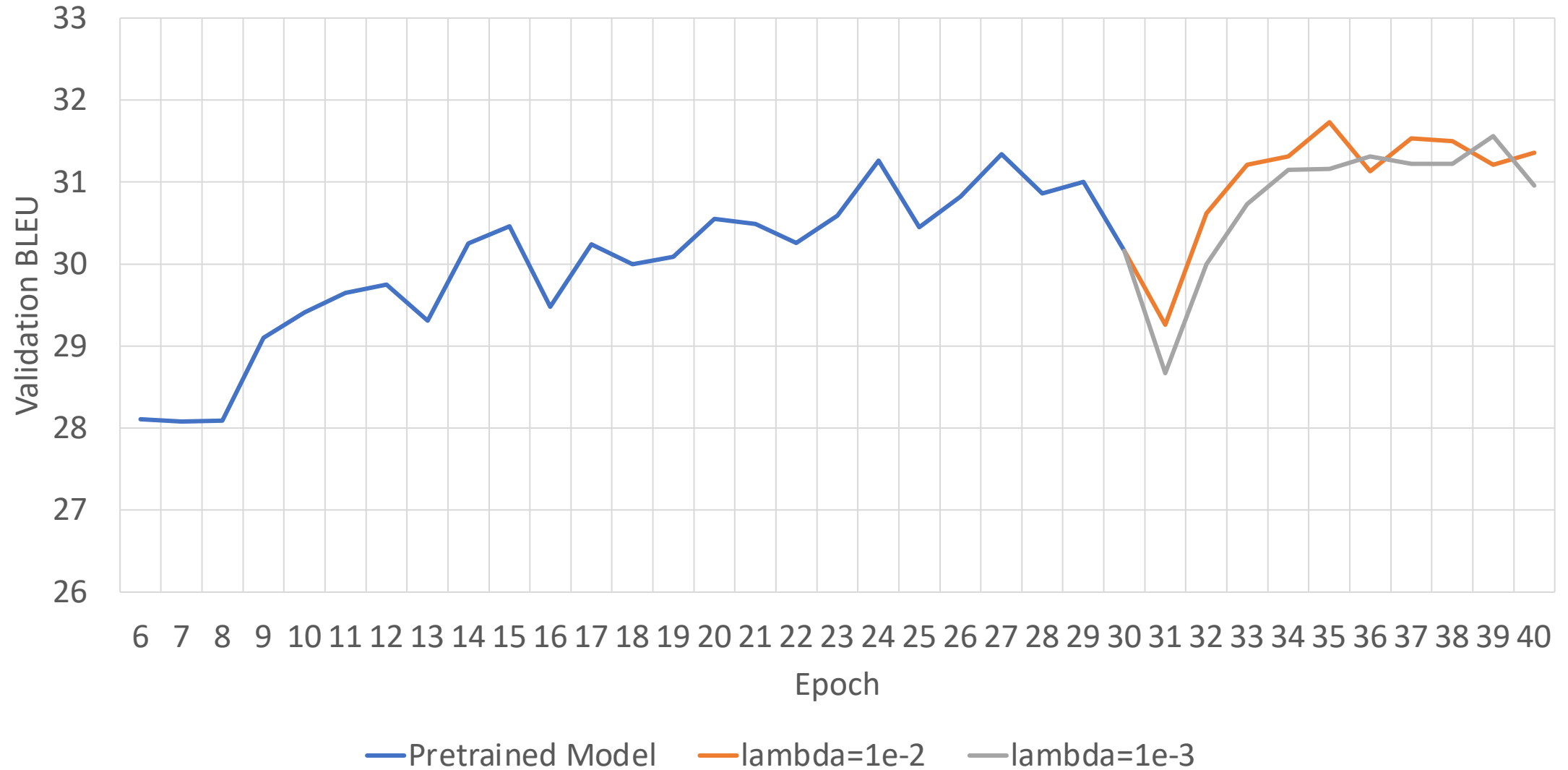
	MLE	MRT	DSL
ENKO	32.53	34.04	33.47
KOEN	29.67	31.24	31.00

REF	Sequence-to-Sequence	MRT	DSL
또한, 이견 미세 혈관을 손상해 당신이 빨간 코를 가지게 할 수 있습니다. 예비신혼부부도 1년 이내 혼인 사실이 증명 가능하면 청약할 수 있다. 택배 박스의 빈 곳에 한국 주소도 같이 붙여주시면 감사하겠습니다. 수백만 가입자가 일상 환경에서 초고속 데이터 전송과 음성통화를 이상 없이 이용하도록 망 안정화와 업그레йд, 커버리지 확대를 지속하는 일은 최우선 과제다. 단기적으로 시장의 변동성은 지속될 것으로 전망되지만 역사적 관점에서 볼 때 최근 시장의 변동성이 특별히 극심한 것은 아니며 장기적인 시장 펀더멘털 역시 여전히 견고해 보인다. 기존 플라이닝 프로브는 제품의 가격이 너무 고가인 데다 ICT와 같은 기존 시험 장비에 비해 시험 시간이 길어서 양산 라인에 적용하기 어렵다는 단점이 있다. 장우혁의 이번 공연이 성황리에 마무리된 것은 놀랄 만한 일이다. 문희상 의원은 조문 후 기자들과 만나 “나라가 소용돌이 한복판에 놓였을 때 국가의 큰 어른인 가르침이 그 어느 때보다 아쉬운데 돌아가시게 돼 진심으로 애도를 표한다”고 말했다.	또한 미세한 코를 가질 수 있는 마이너스를 손상시킬 수 있습니다. 결혼을 계획하는 부부는 결혼식이 1년 이내인 경우에도 가입할 수 있다. 택배 박스의 빈 공간에 한국 주소를 첨부하여 주시면 감사하겠습니다. 가입자 수백만 명이 일상 환경에 문제없이 초고속 데이터 전송과 음성 호출을 이용하는 것은 네트워크 안정화, 커버리지 고도화, 커버리지 확대 등이 최우선 과제다. 단기적으로 시장 변동성은 지속될 것으로 전망되지만 역사적 관점에서 보면 최근 시장의 변동성이 특별히 극단적이지 않고 장기간 시장 펀더멘털도 남아 있는 것으로 보인다. 기존 비행조사는 제품 가격이 너무 높고 시험시간도 ICT 등 기존 테스트 장비보다 길기 때문에 양산 라인을 적용하기가 어렵다. 장우혁의 활약이 성황리에 마무리되었다는 점은 놀라운 일이다. 문재인 의원은 조의 끝에 기자들과 만나 “국가가 소용돌이에 빠져 있을 때 큰 고령화를 줄 수 있는 사람이 분실이라는 점에 진심으로 애도를 표한다”고 말했다.	또한 미세혈관을 손상할 수 있어서 빨간색 코를 가지게 할 수 있습니다. 결혼을 계획하는 부부도 1년 이내 결혼식이 발생할 수 있으면 가입할 수 있다. 택배함의 빈 공간에 한국 주소를 첨부해주시면 감사하겠습니다. 수백만 명의 가입자가 일상 환경에 문제 없이 초고속 데이터 전송과 음성통화를 이용하는 것이 네트워크 안정화, 업그레йд, 확대 안정성이 최우선이다. 단기적으로 시장 변동성이 이어질 것으로 전망되지만 역사적 관점에서 최근 시장의 변동성은 특별히 극적으로 극단적이지 않고 장기적인 시장 펀더멘털도 남아 있는 것으로 보인다. 기존의 비행 탐사선은 제품 가격이 너무 높고, 시험시간이 ICT 등 기존 테스트 장비보다 길기 때문에 양산 라인에 적용하기 어렵다. 장우혁의 공연이 성공적으로 마무리되었다는 점은 놀라운 일이다. 문재혁 의원은 애도한 뒤 기자들과 만나 “소용돌이 한복판이 있을 때 큰 가르침을 줄 수 있는 사람의 손실이라는 것에 진심으로 애도를 표한다”고 말했다.	또한, 당신을 빨간 코로 이끌 수 있는 마이크로셀들을 손상시킬 수 있습니다. 결혼을 계획하는 부부는 1년 이내 결혼식이 성사될 경우 가입할 수도 있다. 택배 상자의 빈 공간에 한국 주소를 첨부해 주시면 감사하겠습니다. 가입자 수백만명이 일상 환경에서 문제없이 고속 데이터 전송과 음성통화를 이용할 수 있도록 네트워크를 안정화하고, 커버리지를 확대·확대하는 것이 최우선 과제다. 단기적으로 시장 변동성이 지속될 것으로 전망되지만 역사적 측면에서 보면 최근 시장의 변동성은 특별히 극단적이지 않고 장기적인 시장 근본성도 남아 있는 것으로 보인다. 기존 플라이닝 프로브는 제품 가격이 너무 높고, 시험시간이 ICT 등 기존 시험 장비보다 길어 양산 라인에 적용하기 어렵다. 장우혁의 활약이 성공적으로 마무리되어 놀라움을 금할 수 없다. 문희상 의원은 조문 끝에 기자들과 만나 “국가가 소용돌이에 한창일 때 훌륭한 가르침을 줄 수 있는 사람의 상실이라는 애도를 표한다”고 말했다.
한 은행권 관계자는 “기존에는 자체 DSR 기준에 걸려도 심사역이 승인하면 대출이 가능했는데, 관리지표로 본격 도입되면 사실상 대출이 어려워질 것”이라고 전했다.	은행권 관계자는 “예전에는 자체 DSR 기준에 해당돼도 심사팀이 대출을 승인하면 대출을 할 수 있었는데, 경영지표로 도입되면 실질적으로 대출을 받기 어려울 것”이라고 말했다.	은행권 관계자는 “기존에는 자체 DSR 기준에 해당하더라도 심사단이 대출을 승인하면 대출할 수 있었지만 관리지표로 도입하면 사실상 대출을 받기 어려울 것”이라고 말했다.	한 은행 관계자는 “과거에는 자체 DSR 기준에 해당하더라도 심사팀이 대출을 승인하면 대출을 할 수 있었지만 경영지표로 도입되면 실제로는 대출을 받기 어려워진다”고 말했다.
사회적 감사는 업무가 법규나 규정에는 부합하더라도 공정성이나 형평 등 사회적 규범에 맞지 않는다면 이를 개선하도록 권고해 조직에 변화를 유도하는 것이다. 마찰이 있더라도 국익 차원에서 인터넷 역차별 해소 정책을 강력하게 추진해야 한다는 목소리가 힘을 얻기 시작했다.	사회적 감사는 법이나 규정을 준수하더라도 정의나 형평성 등 사회적 규범에 부합하지 않으면 업무 윤리 개선을 시키도록 유도하는 등 조직의 변화를 유도하는 것이다.	사회적 감사는 법이나 규정을 준수하더라도 정의나 형평성 등 사회적 규범을 충족하지 않으면 업무 윤리 개선을 유도해 조직의 변화를 격려하고 있다.	사회적 감사는 법이나 규정을 준수하더라도 정의나 형평성 등 사회적 규범에 부합하지 않을 경우 업무 윤리 개선을 유도해 조직의 변화를 유도하도록 유도한다.
마찰이 있더라도 국익 차원에서 인터넷 역차별 해소 정책을 강력하게 추진해야 한다는 목소리가 힘을 얻기 시작했다.	마찰이 일어나더라도 인터넷 역차별 정책의 해법을 구현하는 목소리가 국익을 향한 힘을 얻기 시작했다.	마찰이 발생하더라도 인터넷 역차별 정책의 해법을 구현하는 목소리가 국익을 위해 힘을 얻기 시작했다.	마찰이 발생하더라도 인터넷 역차별 정책의 해법을 구현해야 한다는 목소리가 국민적 관심에 힘을 받기 시작했다.
오비맥주가 주요 맥주 제품의 가격을 인상키로 하면서 하이트진로와 롯데주류 등도 조만간 맥주 가격을 올릴 가능성이 높아졌다.	오비맥주가 각종 맥주 상품 가격을 올리기로 결정한 만큼 하이트진로와 롯데주류도 가까운 시일 내에 맥주 가격을 인상할 가능성이 높다.	오비맥주가 다양한 맥주 상품의 가격을 올리기로 결정하면서 하이트진로와 롯데주류도 가까운 미래에 맥주 가격을 인상할 가능성이 높다.	오비맥주가 다양한 맥주 상품의 가격 인상을 결정한 만큼 하이트진로와 롯데주류도 가까운 시일 내에 맥주 가격을 인상할 가능성이 높다.
이 사고로 통학차량에 타고 있던 학생 7명(남3, 여4) 중 고교생 3명과 여중생 1명이 다쳐 인근 병원으로 옮겨져 치료를 받고 있다.	이 사고로 통근차량에 타고 있던 학생 7명 가운데 고등학생 3명과 여중생 1명이 부상을 입고 인근 병원으로 옮겨져 치료를 받고 있다.	이 사고로 통학차량에 타고 있던 학생 7명 중교생 3명과 여중생 1명이 다쳐 인근 병원으로 옮겨져 치료를 받고 있다.	통학차량에 타고 있던 학생 7명 가운데 고등학생 3명과 여중생 1명이 다쳐 인근 병원으로 옮겨져 치료를 받고 있다.
반경 1km 내에 용현초, 신광초, 신흥여중 등 다수의 교육시설이 위치해 있고 홈플러스와 현대유비스 병원 등이 도보 5분 거리 내에 위치해있다.	1km 이내에는 용원초등학교, 신광초등학교, 신흥여중 등 수많은 교육시설이 있으며, 걸어서 5분 거리에 홈플러스, 현대위비병원 등이 있다.	1km 안에는 용현초등학교, 신광초등학교, 신흥여학교 등 수많은 교육시설이 있으며, 5분 도보 거리에는 홈플러스와 현대유비병원이 있다.	1km 이내에는 용현초등학교, 신왕초, 신흥여중 등 수많은 교육시설이 있으며 도보 5분 거리에 홈플러스와 현대유비병원이 있다.

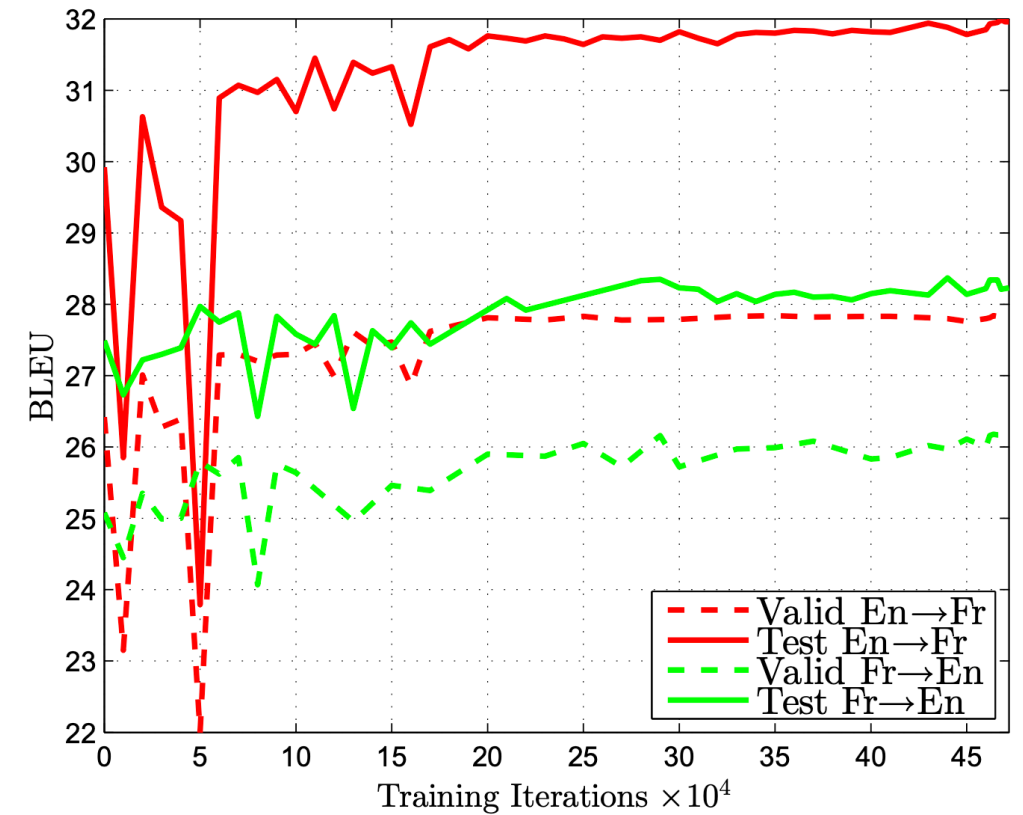
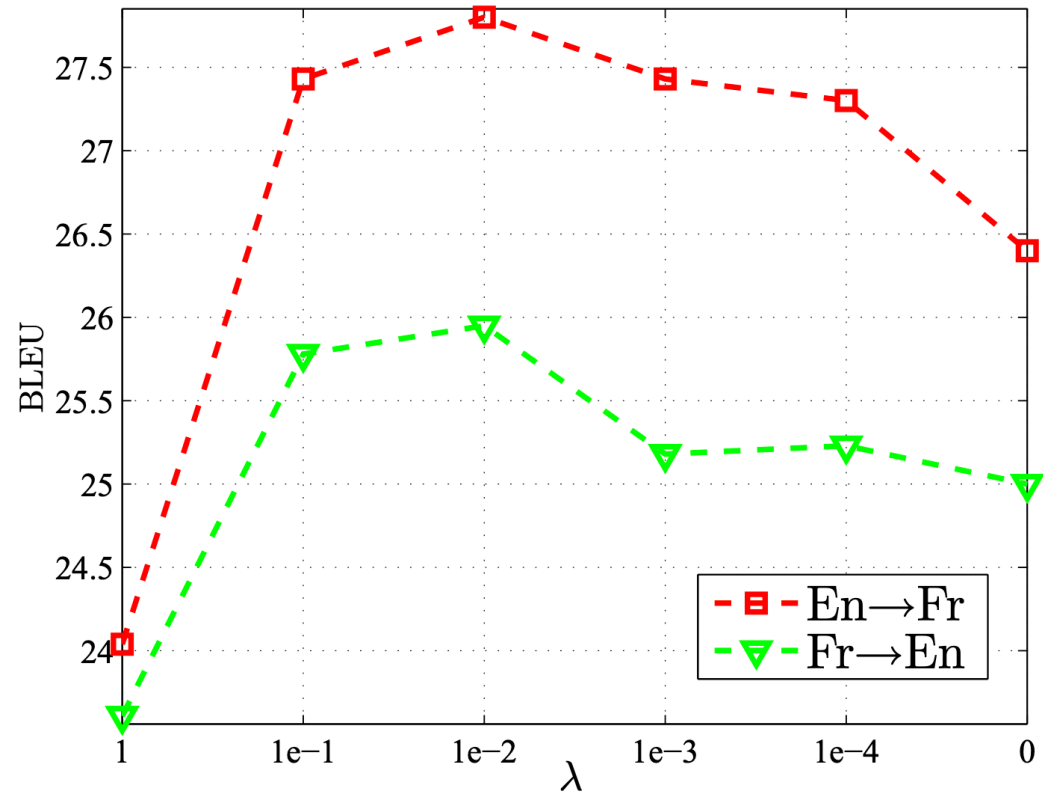
lambda에 따른 성능 비교 (ENKO)



lambda에 따른 성능 비교 (KOEN)



Evaluation Result on the Paper



Summary

- 한 epoch 당 소요 시간(3시간)이 더 걸리지만 두 모델이 동시에 학습됨
 - 튜닝 하는데 많은 시간이 소요됨
- λ 이외에도 다양한 hyper-parameter의 영향을 받았음
 - max_grad_norm
 - SGD 대신 Adam을 통해 학습
 - Valid BLEU에 따른 LR decay는 생략
- Paper만큼의 성능 향상은 확보하지 못했지만, MLE에 비해 성능 향상됨