

# 20210624 Samsung Card NLP

Spoken Language Dataset

Knowledge Distillation

# 구어체 데이터셋

**AI Hub** <https://aihub.or.kr/aihub-data/natural-language/about>

- 감성 대화 말뭉치
- 한국어 대화
- 한국어 대화 요약
- 한국어 음성 (음성 파일 + 전사 텍스트 파일)

**국립국어원 모두의 말뭉치** <https://corpus.korean.go.kr/>

- 일상 대화 말뭉치 2020
- 일상 대화 음성 말뭉치 2020 (음성 파일 + 전사 텍스트 파일)
- 구어 말뭉치

# 구어체 언어분석 API (ETRI)

형태소분석 기술, 개체명인식 기술

축약이나 기존 맞춤법을 벗어나는 표현이 많은 구어체 특성상 새로운 형태소분석이 필요

형태소분석은 MeCab보다 우수함

데이터 부족 해결: 전이학습(transfer learning), 데이터 증강(data augmentation)

# 구어체 언어분석 API (ETRI)

형태소/개체명

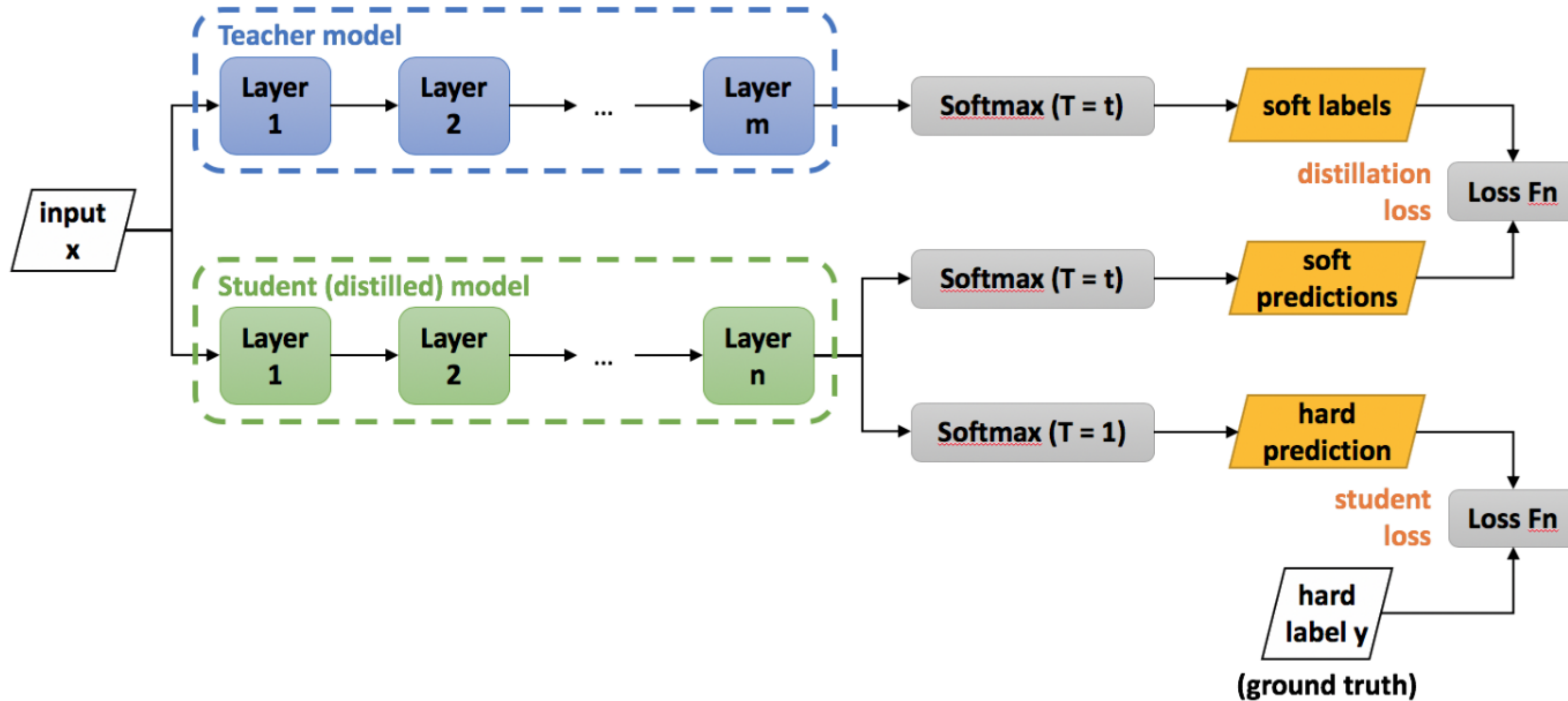
No	Word	형태소			개체명		
		단어	태그	설명	단어	태그	설명
0	톰	톰	NNP	고유명사	톰 행크스	PS_NAME	사람 이름
1	행크스면	행크스 이 면	NNG VCP EC	일반명사 긍정지정사 연결어미			
2	믿고	믿 고	W EC	동사 연결어미			
3	봍니당	봍 봍니당	W EF	동사 종결어미			

# Knowledge Distillation (지식 증류)

NLP에서의 distillation:

- transfer learning의 일종인 knowledge distillation 기법
- model compression, cross-lingual training, domain adaptation 등에 활용
- 큰 모델(teacher model)이 충분히 학습한 지식을 작은 모델(student model)이 전달받아서 잘 기능할 수 있도록 하는 것
- 작은 모델: 규모가 작거나 (더 적은 계산량) 자료가 부족한 데이터/task에 적용하면 좋음!

# Knowledge Distillation (지식 종류)



# Knowledge Distillation (지식 증류)

- **distillation loss**: teacher model과 student model 사이의 (지식의) 거리를 좁히는 역할
  - 예: 문장 임베딩 사이의 Mean Squared Errors (Reimers and Gurevych., 2020<sup>1</sup>)  
임베딩 사이의 Cross Entropy loss (Hinton et al., 2015<sup>2</sup>)
- **student loss**: student model이 적용하고자 하는 task에서의 loss

1 Reimers, N., & Gurevych, I. (2020). *Making monolingual sentence embeddings multilingual using knowledge distillation*. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. pp. 4512-4525.

2 Hinton, G., Vinyals, O., & Dean, J. (2015). *Distilling the knowledge in a neural network*. NIPS Deep Learning and Representation Learning Workshop.

# Knowledge Distillation (지식 종류)

- soft labels/predictions <--> hard labels/predictions
  - 정답으로 예측되지 않은 다른 label에 대한 정보도 남겨두어 학습에 도움이 되도록

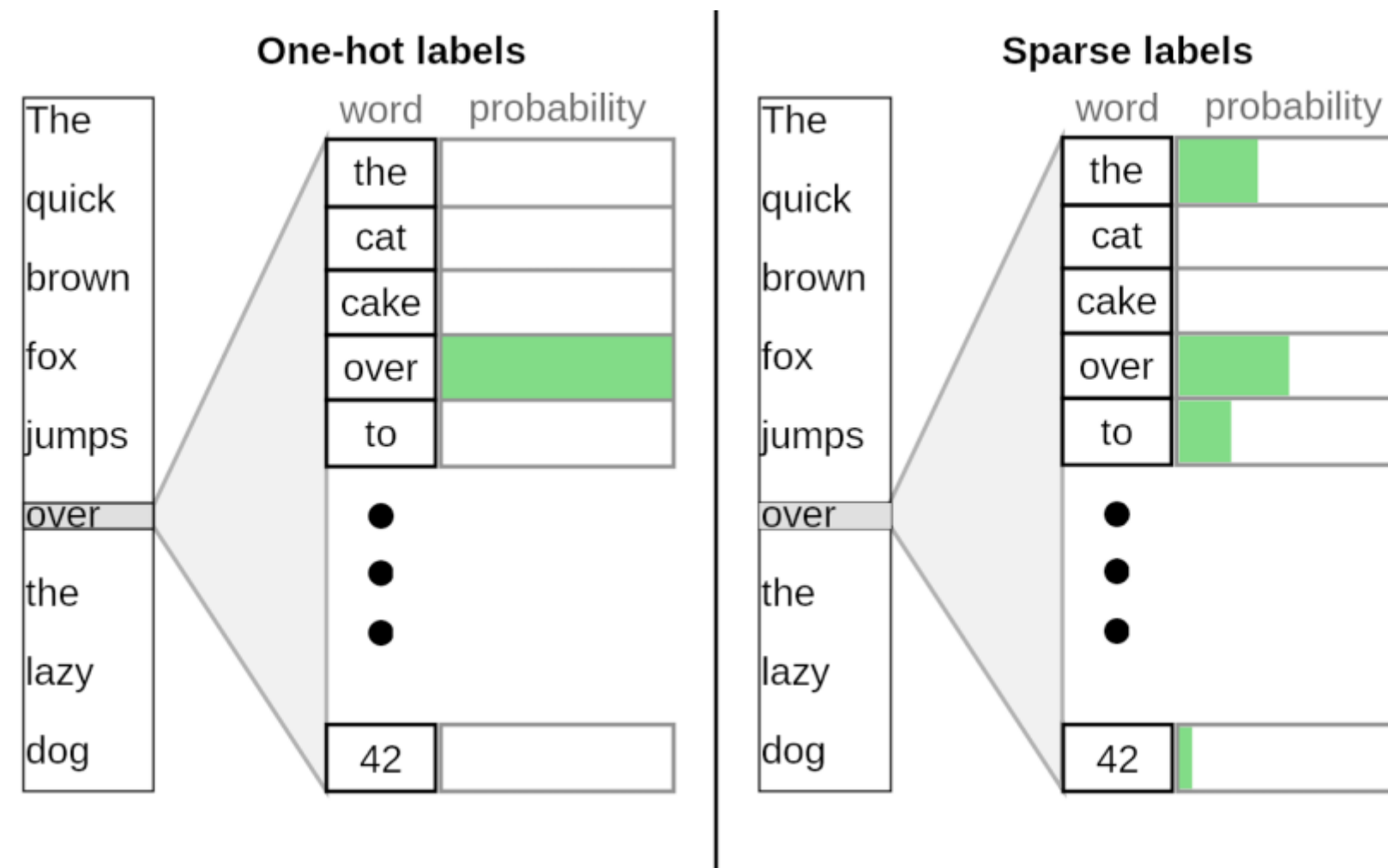


image source: <https://blog.floydhub.com/knowledge-distillation/>



# Knowledge Distillation (지식 증류)

- 모델 조합 예시:

**teacher model** - original BERT, **student model** - small-scaled task-specific model

- Tan et al. (2019)<sup>3</sup>: individual monolingual *translation* models -> multilingual model(s)
- Chen et al. (2020)<sup>4</sup>: finetuned BERT -> conventional seq2seq model (text generation)
- Ryu and Lee (2020)<sup>5</sup>: source domain (finetuned) BERT -> target domain BERT
- Reimers and Gurevych (2020): English S-BERT -> multilingual RoBERTa (translation)

3 Tan, X., Ren, Y., He, D., Qin, T., Zhao, Z., & Liu, T. Y. (2019). *Multilingual neural machine translation with knowledge distillation*. Proceedings of the 7th International Conference on Learning Representations.

4 Chen, Y. C., Gan, Z., Cheng, Y., Liu, J., & Liu, J. (2019). *Distilling Knowledge Learned in BERT for Text Generation*, Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 7893-7905.

5 Ryu, M., & Lee, K. (2020). *Knowledge Distillation for BERT Unsupervised Domain Adaptation*. arXiv preprint arXiv:2010.11478.