세계 한국어한마당 2021 World Convention of the Korean Language

함의관계 주석 자원을 활용한 한국어 인공지능 평가

Assessing the Korean AIs with the Commitment Data

송상헌 고려대학교 부교수

Sanghoun Song *Linguistics, Korea University*

문법

- 문법(文法): Grammar
- N-gram
- 어떤 문법이 좋은 문법인가?
  - 특정 언어의 현상을 잘 설명할 수 있는 체계
- 잘 설명한다는 것은 무엇일까?
  - 정문과 비문을 구분할 수 있고,
  - 비문의 경우 어떠한 제약을 위반하였기 때문인지를 설명할 수 있어야 한다.
- 어떤 언어모형이 좋은 언어모형인가?
  - 문법도 평가의 대상이 된다.
  - 언어모형을 일종의 문법으로 간주한다.
- 대표적인 문법 체계 (문장 단위 혹은 그 이상)
  - 통사론
  - 의미론
  - 화용론

인공지능을 위한 종합시험

- 대학수학능력시험 大學修學能力試驗
- Generalized Language Understanding Evaluation

Rank	Name	Model	URL	Score	CoLA	SST-2	MRPC	STS-B	QQP
1	T5 Team - Google	T5		89.7	70.8	97.1	91.9/89.2	92.5/92.1	74.6/90.4
2	ALBERT-Team Google Language	ALBERT (Ensemble)		89.4	69.1	97.1	93.4/91.2	92.5/92.0	74.2/90.5
3	王玮	ALICE v2 large ensemble (Alibaba DAMO NLP)		89.0	69.2	97.1	93.6/91.5	92.7/92.3	74.4/90.7
4	Microsoft D365 AI & UMD	FreeLB-RoBERTa (ensemble)		88.8	68.0	96.8	93.1/90.8	92.4/92.2	74.8/90.3
5	Facebook AI	RoBERTa		88.5	67.8	96.7	92.3/89.8	92.2/91.9	74.3/90.2
6	XLNet Team	XLNet-Large (ensemble)		88.4	67.8	96.8	93.0/90.7	91.6/91.1	74.2/90.3
7	Microsoft D365 AI & MSR AI	MT-DNN-ensemble		87.6	68.4	96.5	92.7/90.3	91.1/90.7	73.7/89.9
8	GLUE Human Baselines	GLUE Human Baselines		87.1	66.4	97.8	86.3/80.8	92.7/92.6	59.5/80.4
9	Stanford Hazy Research	Snorkel MeTaL		83.2	63.8	96.2	91.5/88.5	90.1/89.7	73.1/89.9
10	XLM Systems	XLM (English only)		83.1	62.9	95.6	90.7/87.1	88.8/88.2	73.2/89.8

통사론

- Corpus of Linguistic Acceptability (수용성)

Label	Sentence	Source
*	The more books I ask to whom he will give, the more he reads.	Culicover and Jackendoff (1999)
✓	I said that my father, he was tight as a hoot-owl.	Ross (1967)
✓	The jeweller inscribed the ring with the name.	Levin (1993)
*	many evidence was provided.	Kim and Sells (2008)
✓	They can sing.	Kim and Sells (2008)
✓	The men would have been all working.	Baltin (1982)
*	Who do you think that will question Seamus first?	Carnie (2013)
*	Usually, any lion is majestic.	Dayal (1998)
✓	The gardener planted roses in the garden.	Miller (2002)
✓	I wrote Blair a letter, but I tore it up before I sent it.	Rappaport Hovav and Levin (2008)

Table 3: CoLA random sample, drawn from the in-domain training set (✓= acceptable, *=unacceptable).

의미론

- Semantic Textual Similarity (진리조건)

2	*The two sentences are not equivalent, but share some details.*
	They flew out of the nest in groups. They flew into the nest together.
1	*The two sentences are not equivalent, but are on the same topic.*
	The woman is playing the violin. The young lady enjoys listening to the guitar.
0	*The two sentences are completely dissimilar.*
	The black dog is running through the snow. A race car driver is driving his car through the mud.

5	*The two sentences are completely equivalent, as they mean the same thing.*
	The bird is bathing in the sink. Birdie is washing itself in the water basin.
4	*The two sentences are mostly equivalent, but some unimportant details differ.*
	Two boys on a couch are playing video games. Two boys are playing a video game.
3	*The two sentences are roughly equivalent, but some important information differs/missing.*
	John said he is considered a witness but not a suspect. “He is not a suspect anymore.” John said.

화용론

- Natural Language Inference (행간의 의미)

Met my first girlfriend that way.	FACE-TO-FACE **contradiction** C C N C	I didn't meet my first girlfriend until later.
8 million in relief in the form of emergency housing.	GOVERNMENT **neutral** N N N N	The 8 million dollars for emergency housing was still not enough to solve the problem.
Now, as children tend their gardens, they have a new appreciation of their relationship to the land, their cultural heritage, and their community.	LETTERS **neutral** N N N N	All of the children love working in their gardens.
At 8:34, the Boston Center controller received a third transmission from American 11	9/11 **entailment** E E E E	The Boston Center controller got a third transmission from American 11.
I am a lacto-vegetarian.	SLATE **neutral** N N E N	I enjoy eating cheese too much to abstain from dairy.
someone else noticed it and i said well i guess that's true and it was somewhat melodious in other words it wasn't just you know it was really funny	TELEPHONE **contradiction** C C C C	No one noticed and it wasn't funny at all.

Table 1: Randomly chosen examples from the development set of our new corpus, shown with their genre labels, their selected gold labels, and the validation labels (abbreviated E, N, C) assigned by individual annotators.

함의 분석

- 함의 含意
  - 머금다, 품다
  - 들어있음: 包含 含蓄 含量 含有
- 내포 명제의 진리치에 기여(commitment)하는 화자/필자의 확신성의 정도를 파악하는 연구
  - 대상 담화와 그 대상 담화의 내포절이 주어지면 해당 내포절의 진리치에 기여하는 화자의 확신성을 함의(entailment), 중립(neutral), 그리고 모순(contradiction)으로 분류하여 리커트 척도를 통해 분석
  - 구축한 함의 분석 말뭉치를 활용하여 KorBERT, KR-BERT 등 최신 한국어 인공지능 모델의 자연어이해 수준을 평가하고 결과를 제시

함의 분석

한국어는 보문소와 술어의 결합으로 내포 명제의 (비)사실성이 결정

선행문장	영희는 어제 눈이 온 것을 보았다. 영희는 영수에게 질문을 한다.
대상문장	“철수는 어제 눈이 온 것을 아니?”
후행문장	영수는 모르겠다고 대답했다.

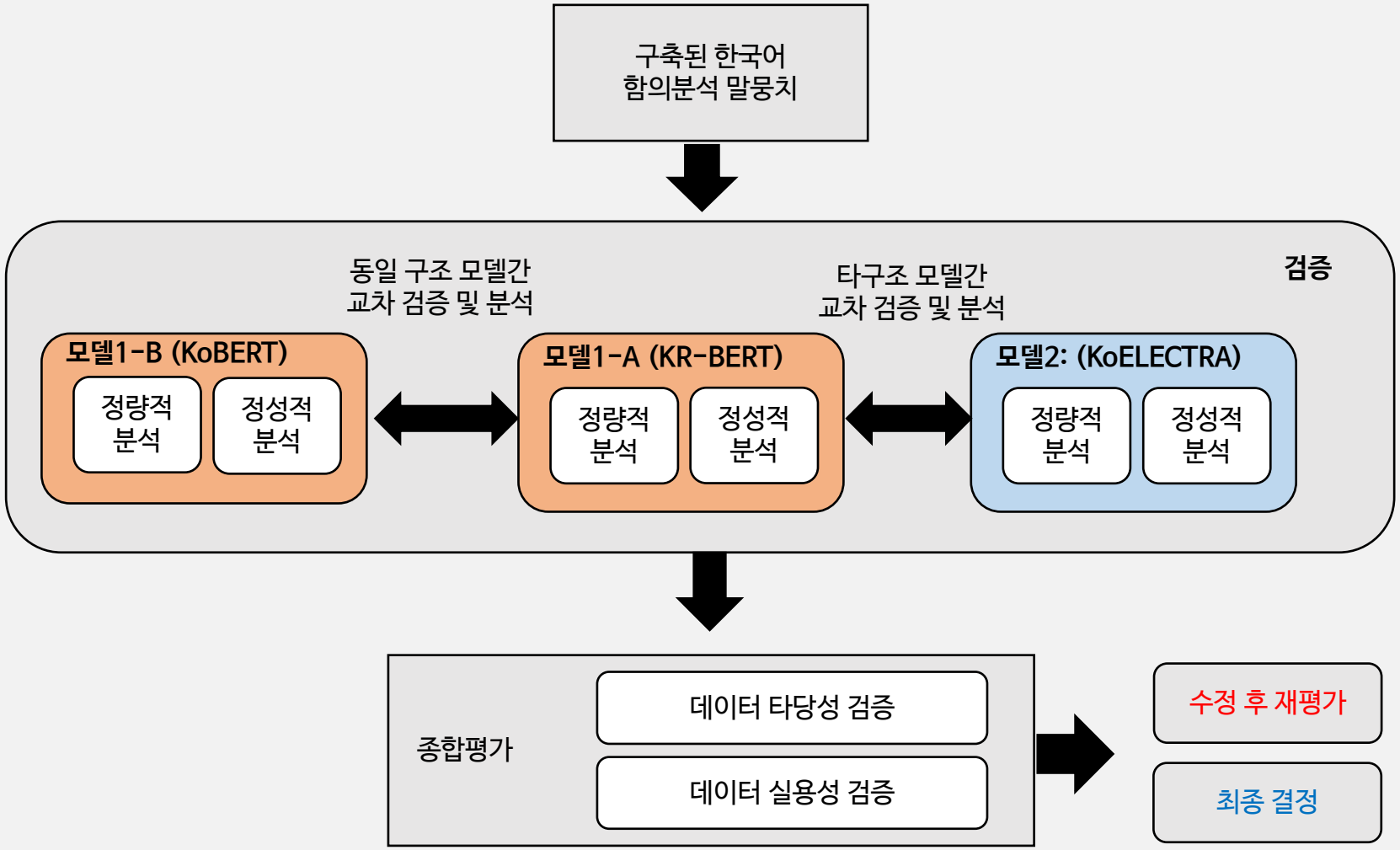
↓

내포명제	어제 눈이 왔다.

내포 명제의 진리치 대한 화자(영희)의 확신성 정도는?

함의 분석

시스템 구현을 통한 한국어 함의 분석 말뭉치 검증의 흐름도



인공지능 평가는 쉬운 과제일까?

세계 한국어한마당 World Convention of the Korean Language

- 응용언어학 (Applied Linguistics)
- 난이도 조절
  - 운전면허시험 / 대학수학능력시험 / 고시
  - 물수능 vs. 불수능
- Artifact: 표본 제작 과정에 투입된 여러 물질이나 기법 때문에 생성된 의도치 않은 부정적 부산물
  - 찍기 신공
  - 출제자의 의도
- 시험문제와 실제 실력
  - 토익만점자의 영어실력