



상점 데이터 분석을 통한 업종 예측 모델 개발

박찬현¹, 박성현¹, 어승준¹, 장석원¹,
전희국², 임동혁¹

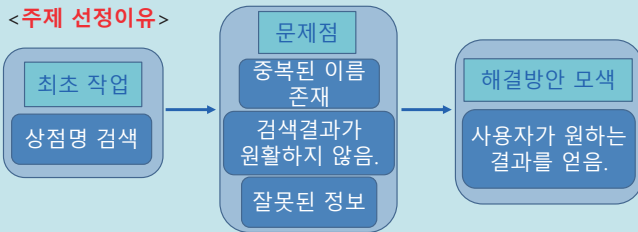
¹ 광운대학교 정보융합학부, ² ㈜ 오픈업

서론(주제소개 & 선정이유)

<배경 및 필요성>

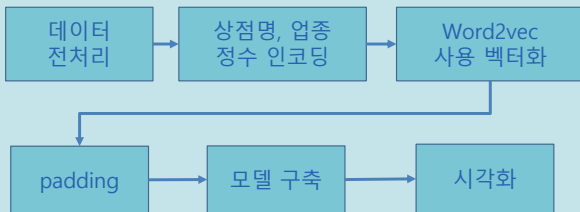
→ 스마트폰의 활용도가 높아짐에 따라 사람들이 **필요한 정보**를 얻으려 할 때 얻는 **불편함**을 제거하기 위해 사용자를 위한 더욱 **편리하고 정확한 모델**을 만들 **필요성**을 느낌.

<주제 선정이유>

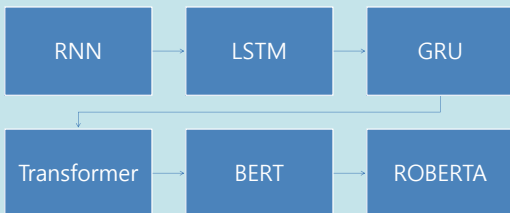


• 상점 이용을 위한 스마트폰의 활용도가 높아지면서 상점명을 "000"이라고 검색하면 검색한 상점이 어떤 업종인지 사용자에게 정보를 제공함으로써 사용자의 궁금증을 사전 해결해주며 유용하게 활용될 수 있음 → 업종별 **예측 모델 개발**

알고리즘



사용한 딥러닝 모델



- **RNN** - 입력과 출력을 시퀀스 단위로 처리하는 시퀀스 모델
- **LSTM** - RNN의 일종, 시계열 데이터, 자연어 처리에 사용
- **GRU** - LSTM의 구조를 간단하게 개선한 모델
- **Transformer** - RNN을 사용하지 않고 인코더-디코더 구조
- **BERT** - 사전 훈련 Embedding을 통해 특정 과제의 성능을 향상시킬 수 있는 언어 모델
- **ROBERTA** - BERT 모델의 하이퍼 파라미터 및 학습 데이터 사이즈 등을 조절하며 기존 BERT 모델보다 성능이 뛰어난

데이터 및 모델 선정 과정

Raw Data

shop_name	shop_type	
0	신목부대리가 오프라인점	고기 뷔페
1	도원동점	고기 뷔페
2	하우촌	고기 뷔페
3	고고상고기	고기 뷔페
4	현대식당	고기 뷔페
...
131647	도마기 (남도왕산)	회/수산물 일반
131648	명태어장(서울대점)	회/수산물 일반
131649	미스터요정어 선릉점	회/수산물 일반
131650	인기영선유도	회/수산물 일반
131651	상도왕브리 서울구왕점	회/수산물 일반

131,652 x 2(행/열)

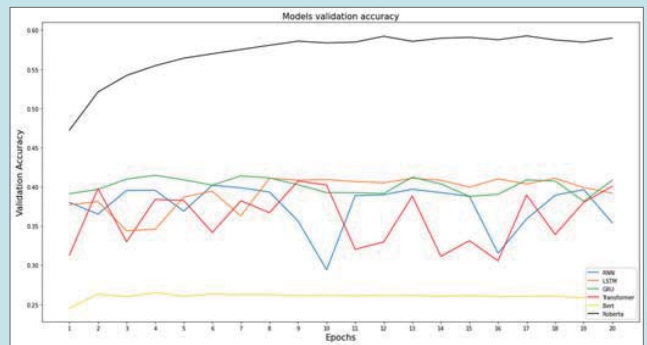
전처리

작은 세탁소 업종 제거
업종명 정수 인코딩

최종 사용할 Data

shop_name	shop_type	
0	신목부대리가 오프라인점	0
1	도원동점	0
2	하우촌	0
3	고고상고기	0
4	현대식당	0
...
131647	도마기 (남도왕산)	59
131648	명태어장(서울대점)	59
131649	미스터요정어 선릉점	59
131650	인기영선유도	59
131651	상도왕브리 서울구왕점	59

131,650 x 2 (행/열)



- **Train : Validation : Test** 데이터 셋의 비율을 **6:2:2**의 비율로 분할 후 테스트 진행
- Validation Data에 대한 **예측 정확도**를 위의 그래프를 통해 비교해본 결과 **ROBERTA** 모델의 **Validation Accuracy**가 가장 뛰어났음을 확인 → 최종 모델로 **ROBERTA** 모델 선택

결론

- WORD2VEC, GLOVE와 같은 **워드임베딩 방식**을 다방면으로 활용할 수 있었으며, **딥러닝**을 이용한 데이터 분석을 통해 Validation loss와 accuracy와 같은 **유의미한 결과**를 도출함.
- 이러한 과정을 통해 결과적으로 가장 높은 성능을 보이는 예측 모델은 **ROBERTA 모델**이라는 것을 확인할 수 있었으며, 이러한 모델들을 통해 다양한 사용자들이 **상점 정보**를 쉽게 얻을 수 있도록 하는 최종 목표를 달성함.

참고문헌

- [1] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.
- [2] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.