

Predicting Daily Emotional Distribution and Mental Health Index via LLM and Further Transfer Learning

2024 Data Mining Final Project

인공지능학과 석사과정
120240328정가연

Table of Contents

1. Introduction

2. Proposed Method

2.1 Construct Sentiment Classification Model

2.2 Inference Youtube Data

2.3 Prediction Sentiment Distribution

& Mental Health Index

3. Experiment

4. Conclusion

Inference

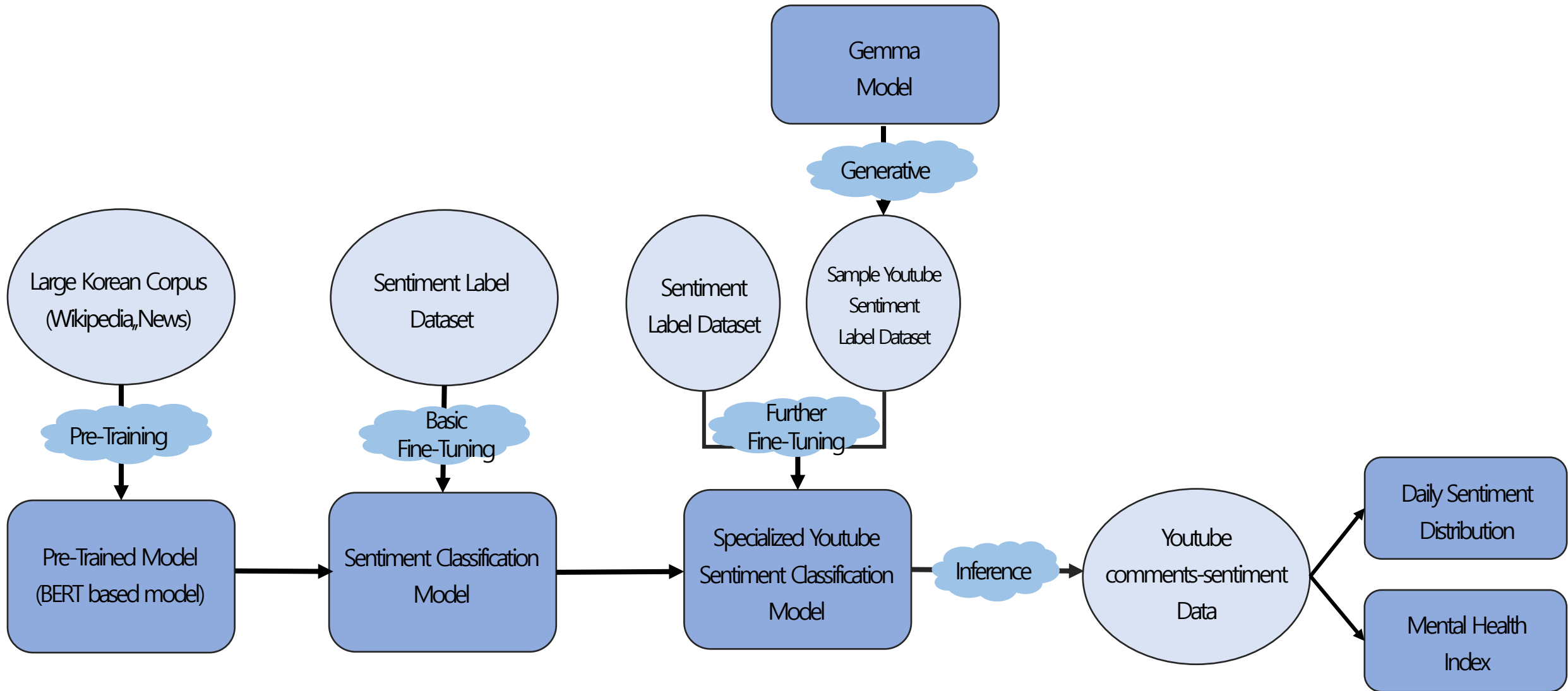
Motivate

- EIU의 정신건강 및 통합 보고서에 의하면 한국의 정신보건 통합지수 75.9점 -> OECD 국가 평균보다 낮은 현황
- 우울증 유병률 세계 1위, 자살률 OECD 평균치 상회
- 한국의 정신건강 지수 측면에서 개선 필요
- AI를 활용하여 정신건강 지수를 예측함으로써 효율적인 대응책 마련 가능

Theme

- 국민들의 정신건강 상태를 예측하기 위해서 LLM과 SNS 데이터를 사용하여 감정 분포와 정신건강 지수를 예측
- 감정 레이블 데이터를 기반으로 4가지 감정을 예측하는 다중 분류 모델 구축
- Basic Transfer Learning, Hyper parameter Optimization, Generating Youtube Sentiment Label, Further Transfer Learning
- 감정 별 분포 및 정신건강 지수 도출

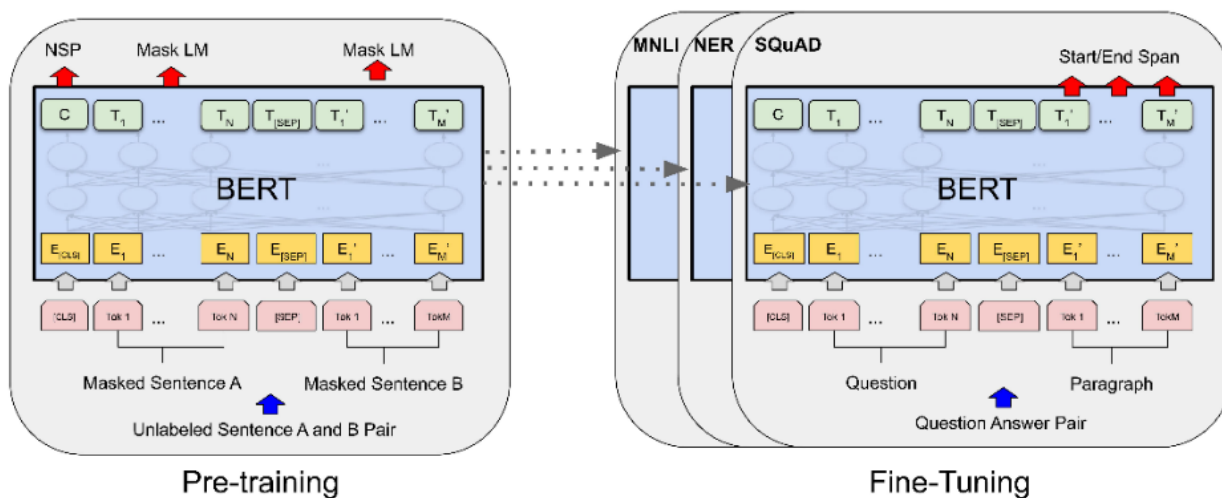
Proposed Method



Construct Sentiment Classification Model

1. Basic Transfer Learning

- Pre-Training된 언어 모델을 불러와 학습한 뒤 목표 테스트에 맞게 Fine-Tuning을 수행
- Pre-training 모델로 KLUE-BERT, KLUE-RoBERTa-base, KLUE-RoBERTa-large, KoELECTRA 모델 사용



- KLUE-BERT : 한국어 자연어 처리 작업에 최적화된 BERT 기반 모델
- KoELECTRA : 한국어 기반의 ELECTRA 모델로 텍스트 생성하는 Generator & 생성된 텍스트가 원래 텍스트인지 아닌지 판별하는 Discriminator로 구성
- KLUE-RoBERTa : BERT보다 더 큰 배치 사이즈와 더 긴 텍스트 시퀀스 사용하여 성능 극대화한 모델

Construct Sentiment Classification Model

2. Hyper Parameter Tuning

- 감정 분류 모델의 성능 고도화를 위해 하이퍼 파라미터 최적화 수행
- 여러가지 조합을 시도해 보며 모델의 최적 조합 탐색
- Optuna의 TPES 최적화 방법 사용 : 확률적으로 좋은 후보 값을 선택하기 위해 이전에 시도한 후보 값들의 분포를 고려
- -> 매 시도마다 새로운 후보 값들의 분포 추정, 더 좋은 후보 값 탐색

3. Generate Sample Youtube Sentiment Label Dataset

- Generative AI 모델인 Gemma를 로드
- 유튜브 댓글이 들어오면 4가지 감정(행복, 분노, 슬픔, 중립) 중 하나의 감정을 예측하도록 Prompt Tuning 수행
- 9,000개의 유튜브 댓글에 대한 감정 매핑된 데이터셋 구축

Construct Sentiment Classification Model

4. Further Transfer Learning

- 감정 분류 테스트에 Fine-Tuning된 기존 분류 모델에 (유튜브 댓글 라벨링 데이터셋 + 기존 감정 라벨링 데이터셋)을 사용해서 한번 더 Fine-Tuning 진행
- => 유튜브 댓글의 특징을 반영한 고도화된 감정 분류 모델 구축
- 기존 감정 분류 모델보다 성능 개선됨

Inference Youtube Data

- Further Transfer Learning을 통해 구축된 **유튜브 댓글 특화 감정 분류 모델**을 활용해서 전체 유튜브 댓글 감정 추론
- 원본 유튜브 데이터 전처리 -> batch size 로 분할 -> 각 batch에 대해 추론 수행
- 각 batch에 대한 처리 과정 : 텍스트 토큰화 -> 모델 평가 모드로 전환 -> 입력 데이터 기반 추론 수행
- -> 출력 logits에 softmax 함수 적용하여 각 함수에 대한 확률 계산 -> 각 텍스트에 대해 가장 높은 확률을 가진 감정 예측

Text	Top_Sentiment	Prob_Happy	Prob_Angry	Prob_Sad	Prob_Neutral	Mental_index_score
한동훈 너무 뻔뻔스럽다, 대한 민국 법이 너무 웃켜, 법인이 큰소리 치고있고만.	happy	0.623844	0.011767	0.356197	0.008191	0.255880
애들아고맙다""물났리나서'내 가살았다"코로나가나를인기 올려주네'그래'그래잘한다'애 들...	happy	0.999581	0.000219	0.000033	0.000167	0.999328
핵발전의 축복! 물부족 국가에 서 물폭탄 국가로, 중국에 핵발 전소 올 연말이면 49기...	happy	0.999403	0.000472	0.000109	0.000016	0.998823

Predict Sentiment Distribution & Mental Health Index

- 일자별 평균 감정 분포 산식

$$\text{daily sentiment ratio}_d = \frac{\sum_{i=1}^{\text{Number of Comments}_d} \text{Probability of Sentiment}_i}{\text{Number of Comments}_d}$$

- 월별 평균 감정 분포 산식

$$\text{monthly sentiment ratio}_m = \frac{\sum_{i=1}^{\text{Number of Comments}_m} \text{Probability of Sentiment}_i}{\text{Number of Comments}_m}$$

- 정신건강 지수 산식

$$\begin{aligned} \text{Mental Health Index} = & (\text{Probability of happy} * +1) + (\text{Probability of angry} * -1) \\ & + (\text{Probability of sad} * -1) + (\text{Probability of neutral} * 0) \end{aligned}$$

Experiment

Dataset

데이터셋	감정 레이블	문장 수
감성대화말뭉치	분노, 슬픔, 행복	27,884
감정 분류를 위한 대화 음성 데이터셋	중립	7,421

분노 문장 수	슬픔 문장 수	중립 문장 수	행복 문장 수	합계
10,417	10,128	7,421	7,339	35,305

-감성대화말뭉치, 감정 분류를 위한 대화 음성 데이터셋-

videoDate	videoTitle	text
2020-01-06	'백인들의 잔치' 골든글로브 뒤흔었다...또 새 역사 (2020.01.06뉴스데스크MBC)	사실 기생충이란 영화는 그 많은 요소들에 감독이 내포하는 의미가 대단해서 영화를 정...
2020-01-06	'백인들의 잔치' 골든글로브 뒤흔었다...또 새 역사 (2020.01.06뉴스데스크MBC)	골든글로브건 오스카건 여전히 인종에 대한 편견의 벽을 깨야할 부분이 남아있다고 생각...
2020-01-06	'백인들의 잔치' 골든글로브 뒤흔었다...또 새 역사 (2020.01.06뉴스데스크MBC)	각본상이 once upon a time in hollywood 한테 간건 진짜 예바...

-유튜브 댓글 데이터-

	publishedAt	videoTitle	clean_text	clean_sentiment
0	2020-08-01T07:32:13Z	2명 사망·이재민 150명..충청-전북에 오늘 또 물폭탄 - [LIVE]MBC 뉴스투...	기상청 날씨하나 맞추는게 어렵니	슬픔
1	2020-07-31T10:41:35Z	2명 사망·이재민 150명..충청-전북에 오늘 또 물폭탄 - [LIVE]MBC 뉴스투...	원 뉴스에 광고가 이리도 많나요 징허네 진짜	분노

-유튜브 댓글-감정 매핑 샘플 데이터셋-

Evaluation Metric

-Accuracy : 전체 예측 중에서 실제 레이블과 일치한 예측의 비율을 의미하며, 모델이 얼마나 자주 올바르게 예측하는지 나타내는 척도로 사용

$$Accuracy = \frac{\text{Number of Correct Prediction}}{\text{Total Number of Predictions}}$$

-F1 Score : 모델이 True라고 예측한 것 중에서 실제 True인 비율을 나타내는 Precision과 실제 True 중에서 모델이 True라고 정확하게 예측한 비율을 나타내는 Recall의 조화 평균

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

-Per-Class Accuracy : Confusion Matrix를 통해 각 클래스의 정확도를 계산한다 예측 결과와 실제 레이블 간의 매핑을 행렬 형태로 나타냄으로써, 각 클래스의 대각선 값을 추출하며 이는 해당 클래스의 정확도 나타냄

$$Accuracy_k = \frac{CM_{k,k}}{\sum_j CM_{k,j}}$$

Sentiment Classification

[Basic Transfer Learning Performance]

Accuracy	F1 Score	Per-class Accuracy			
		Happy	Angry	Sad	Neutral
90.29%	91.22%	95.15%	84.25%	86.04%	99.61%

BERT-base

Accuracy	F1 Score	Per-class Accuracy			
		Happy	Angry	Sad	Neutral
90.26%	91.23%	95.15%	83.97%	86.34%	99.48%

KoELECTRA-base

Accuracy	F1 Score	Per-class Accuracy			
		Happy	Angry	Sad	Neutral
90.06%	90.98%	94.44%	83.49%	87.04%	98.97%

RoBERTa-base

Accuracy	F1 Score	Per-class Accuracy			
		Happy	Angry	Sad	Neutral
91.00%	91.87%	95.29%	85.86%	86.94%	99.35%

RoBERTa-large

[Hyper Parameter Tuning Performance]

Accuracy	F1 Score	Per-class Accuracy			
		Happy	Angry	Sad	Neutral
91.16%	92.07%	95.00%	85.96%	87.24%	99.87%

RoBERTa-large with Hyper Parameter Tuning

[Further Transfer Learning Performance]

Accuracy	F1 Score	감정 별 정확도(%)			
		행복	분노	슬픔	중립
91.32%	92.25%	92.84%	89.45%	88.30%	97.99%

RoBERTa-large with Further Transfer Learning

Inference Youtube Data

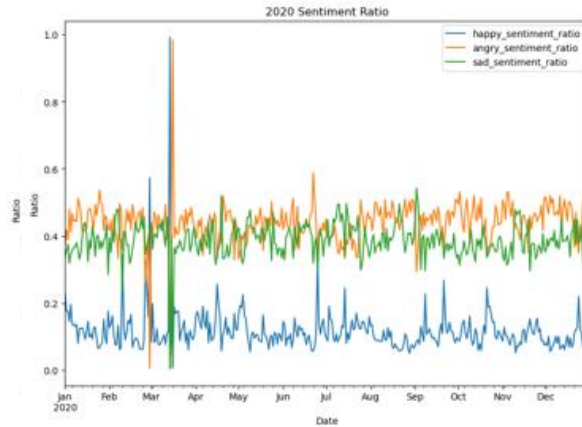
Text	Top_Sentiment	Prob_Happy	Prob_Angry	Prob_Sad	Prob_Neutral	Mental_index_score
한동훈 너무 뻔뻔스럽다. 대한 민국 법이 너무 웃켜. 법인이 큰소리 치고있고만.	happy	0.623844	0.011767	0.356197	0.008191	0.255880
애들아고맙다""물났리나서'내 가살았다'코로나가나를인기 올려주네'그래'그래잘한다'애 들..	happy	0.999581	0.000219	0.000033	0.000167	0.999328
핵발전의 축복! 물부족 국가에 서 물폭탄 국가로. 중국에 핵발 전소 올 연말이면 49기..	happy	0.999403	0.000472	0.000109	0.000016	0.998823

추론 댓글	추론 댓글 수	추론 댓글 비율
중립 댓글	1,500,286	62%
분노 댓글	632,930	26%
슬픔 댓글	173,058	8%
행복 댓글	92,244	4%

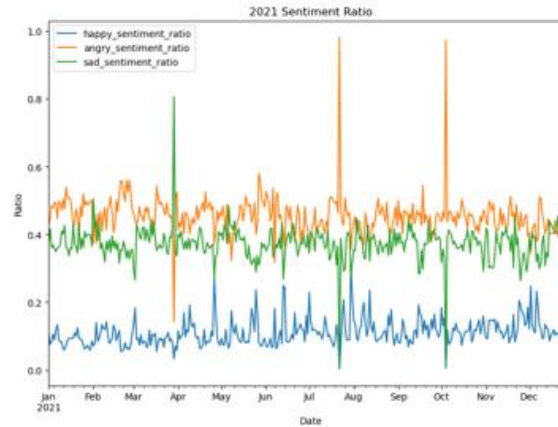
유튜브 댓글 감정 추론 결과

유튜브 댓글 별 감정 및 감정 확률 추론 결과

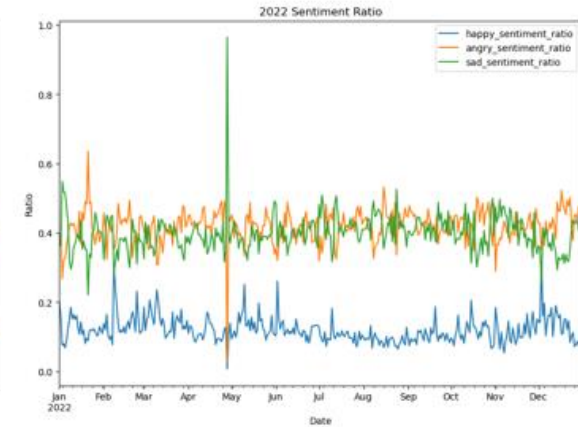
Daily Sentiment Distribution & Mental Health Index



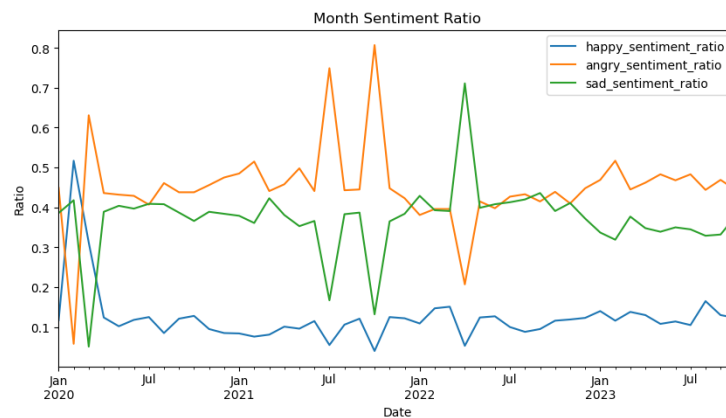
(a) 2020년 일자 별 감정 비율



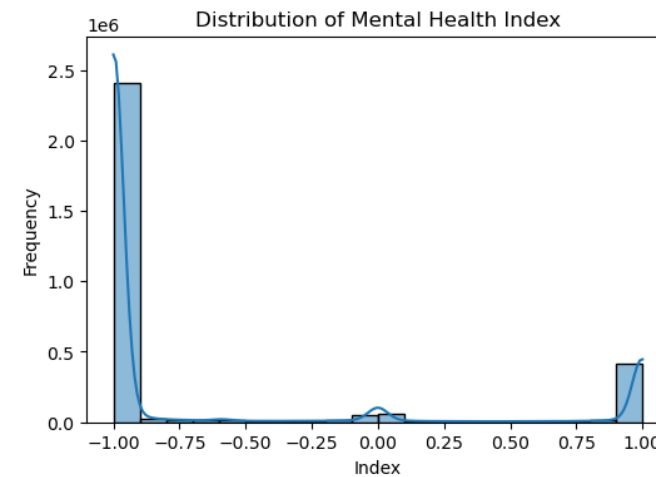
(b) 2021년 일자 별 감정 비율



(c) 2022년 일자 별 감정 비율



월별 감정 비율



정신건강 지수 분포

Conclusion

향후 연구 방향 및 개선 방

[데이터 다양화 및 확대]

- 감정 레이블의 다양성 높여 모델 성능 일반화
- 트위터, 페이스북 등 다른 SNS 기반 감정 레이블 데이터 추가 수집

[실시간 모니터링 시스템 구축]

- 실시간으로 데이터 수집하고 분석할 수 있는 시스템 개발
- 국민들의 정신건강 상태를 실시간으로 모니터링하고 즉각적인 대응 가능하도록 설계

[정신건강 지수 고도화]

- 감정 분포 기반 정신건강 지수 외에도 경제 상황, 사회적 이슈, 의료 관련 변수를 포함하여 정신건강 지수의 예측 정확도 개선

논문 투고 예정

Call for paper

2024 한국인공지능학회 하계학술대회

8월 15일(목)~17일(토) | 부산 벡스코 제1전시장 1층 컨벤션홀

주최
(사)한국인공지능학회

발표분야

- Machine learning applications
computer vision, robotics, signal processing, network, communication, security, natural language processing, neuroscience, climate change, computational biology, etc.
- Learning and Inference for AI
active learning, online learning, continual learning, reinforcement learning, unsupervised, supervised and semi-supervised learning, ranking, clustering, statistical learning theory, causal inference, etc.
- Deep learning
architectures, generative models, optimization, etc.
- Probabilistic inference
Bayesian methods, variational inference, Gaussian Process, graphical models, Monte Carlo methods, etc.
- Social aspects of AI
fairness, accountability, and transparency, privacy, interpretability, etc.

논문제출

- 논문제출 마감 **1차** 2024년 6월 16일(일) ※ 추가 연장 없음
2차 2024년 6월 30일(일)
- 심사결과 통보 2024년 7월 15일(월)
- 논문제출 링크 <https://easychair.org/conferences/?conf=cakia2024>
(국제학회 투고계획을 고려, 저작권 양도는 선택사항임)
- 논문제출 방법 학회홈페이지 참조
- 최우수/우수논문상 채택된 논문들 중 선정, 우수논문 총상금 1,000만원
(마감 시기별로 논문상 차등 배정, 70%(1차):30%(2차))

행사문의

- TEL 070.8830.5920
- Email kaia@aiassociation.kr
- 홈페이지 <http://aiassociation.kr/>

[주최]

-한국인공지능학회

[논문 제출 마감]

-2024년 6월 30일(일)

[발표 분야]

-Machine learning applications

-Learning and Inference for AI

-Deep Learning

-Probabilistic inference

-Social aspects of AI

- [1] Economist Intelligence Unit(EIU), 아시아태평양 지역 정신건강통합지수.
- [2] ETRI. Promising Services Based on AI for Mental Health.
- [3] Adyan Marendra Ramadhani; Hong Soon Goo. Twitter Sentiment Analysis using Deep Learning Methods. IEEE, 2017.
- [4] Devlin, J, Chang, M. W, Lee, K, & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. [arXiv preprint arXiv:1810.04805].
- [5] Devlin, J, Chang, M. W, Lee, K, & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805.
- [6] 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.
- [7] 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.
- [8] Clark, K, Luong, M. T, Le, Q. V, & Manning, C. D. (2020). ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators. arXiv preprint arXiv:2003.10555.
- [9] Lan, Z, Chen, M, Goodman, S, Gimpel, K, Sharma, P, & Soricut, R. (2019). ALBERT: A Lite BERT for Self-supervised Learning of Language Representations. arXiv preprint arXiv:1909.11942.
- [10] Brown, T. B, Mann, B, Ryder, N, Subbiah, M, Kaplan, J, Dhariwal, P, . & Amodei, D. (2020). Language Models are Few-Shot Learners. arXiv preprint arXiv:2005.14165.
- [11] Radford, A, Wu, J, Child, R, Luan, D, Amodei, D, & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI Blog, 1(8), 9.
- [12] Ruder, S, Peters, M. E, Swayamdipta, S, & Wolf, T. (2019). Transfer Learning in Natural Language Processing. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorials (pp. 15-18).
- [13] Howard, J, & Ruder, S. (2018). Universal Language Model Fine-tuning for Text Classification. arXiv preprint arXiv:1801.06146.
- [10] 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.
- [14] Muhammad, A, Wiratunga, N, & Lothian, R. (2016). Contextual sentiment analysis for social media genres. Knowledge-Based Systems, 108, 92-101.
- [15] klue/bert, <https://huggingface.co/klue/roberta-base>

**경청해 주셔서
감사합니다.**