**SW프로젝트 요약서**

|  |  |
| --- | --- |
| **프로젝트 기간** | 2022.09.01 ~ 2023.05.31 **(총 9개월)** |
| **프로젝트 팀원** | **이규민**(컴퓨터소프트웨어학부, 4학년), **박경하**(컴퓨터소프트웨어학부, 4학년) |
| **지도교수** | **김태욱** 교수님 |
| **프로젝트 명** | Pretrained-Language Model을 활용한 MBTI 모델 개발 및 WEB 서비스 제공 |
| **프로젝트 내용** | **- 프로젝트의 이름**  Pretrained-Language Model을 활용한 MBTI 모델 개발 및 WEB 서비스 제공  - **프로젝트의 목적**  기존의 MBTI 응답구조를 벗어나, 사용자들이 작성한 텍스트를 바탕으로 자연어 처리를 통해 MBTI를 예측하여 검사 답안의 자율성을 높이고자 하였습니다. 또한, 기존 언어모델과 대조하여 새로운 방법을 시도해보면서 각각 성능에 얼마나 영향을 미치는지 실험해보고, 성능을 높이고자 하였습니다.  - **프로젝트의 수행 내용**  1. 자연어 처리 기본개념 조사 및 공부  2. (MBTI Personality Prediction for Fictional Characters Using Movie Scripts) 등의 연구 사례 조사  3. Dataset 조사 (kaggle, Aihub 등)  4. 데모 실험 및 실제 모델 구현  5. Web UI 및 UX 제작  6. Web server 및 API 구현  - **프로젝트의 결과**  최종 모델은 T5기반으로, XGBoost을 Classifer로 사용하였으며,  MBTI을 Binary Classification으로서 측정하였을 때, 각 53~75%의 정확도를 보이고 있습니다.  해당 모델을 최종 모델로 채택하여 Django 프레임워크를 통해 Web 서버를 구축하였고, SQLite3 DB에 사용자의 input을 저장하여 계산되는 형태로 모델을 사용할 수 있는 웹 서비스가 제공됩니다. |
| **기대효과**  **및 개선방향** | * 모델을 클라이언트에서 API 형태로 제공할 수 있는 완성된 결과물을 도출할 수 있다는 점과 자연어처리 기술을 통해 개발한 모델을 사용자들이 편리하게 접근할 수 있는 기회를 제공할 수 있다는 점에서 학술적, 산업적인 의의를 두고 있습니다. |

|  |  |
| --- | --- |
| **SW프로젝트 결과보고서** | |
| **프로젝트명** | Pretrained-Language Model을 활용한 MBTI 모델 개발 및 WEB 서비스 제공 |
| **프로젝트**  **요약** | 본 프로젝트에서는 사용자가 작성한 텍스트를 바탕으로 MBTI을 예측하여 기존의 언어모델을 바탕으로 새로운 방법을 시도해보며 각 성능의 영향을 비교하고, 정확성을 높이고자 하였습니다. 또한, 프로젝트에서 새로 학습시킨 모델을 웹 클라이언트에서 이용할 수 있도록 웹 서버를 구현합니다.  본 프로젝트에서는 T5 기반의 모델을 이용하였으며, Binary Classification을 기준으로 정확도를 측정하였을 때 약 55%의 성능을 보입니다. 이는 웹서버와 연동되어 가상서버를 통해 MBTI을 검사해볼 수 있도록 합니다. |
| **프로젝트**  **기간** | 2022.09.01 ~ 2023.05.31 (총 9개월) |
| **산출물** | 졸업 작품 ( O ), 졸업 논문 ( ) |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **학과** | **학번** | **학년** | **이름** | **연락처** |
| 컴퓨터소프트웨어 | 2018008304 | 4 | 박경하 | r[udgk030@hanyang.ac.kr](mailto:udgk030@hanyang.ac.kr)  010-7577-2376 |
| 컴퓨터소프트웨어 | 2017030082 | 4 | 이규민 | [ringgo5@naver.com](mailto:ringgo5@naver.com)  010-7242-9983 |

**목 차**

1. 프로젝트 개요

1.1 프로젝트 목적 및 배경

1.2 프로젝트 최종 목표

2. 프로젝트 내용

3. 프로젝트의 기술적 내용

4. 프로젝트의 역할 분담

4.1 개별 임무 분담

4.2 개발 일정

5. 결론 및 기대효과

1. 프로젝트 개요

**1.1 프로젝트 목적 및 동기**

최근 성격 유형 지표인 Myers-Briggs Type Indicator (MBTI)가 유행하고 있습니다. 시중의 MBTI 검사는 대부분 5종 1택의 형태로 제공되고 있습니다. 그러나, 객관식으로서는 사용자가 생각하는 자신의 성격을 정확하게 표현하기 어렵다는 한계점이 있습니다. 본 프로젝트에서는 사용자가 작성한 텍스트를 바탕으로 MBTI을 예측하여 기존의 언어모델을 바탕으로 새로운 방법을 시도해보며 각 성능의 영향을 비교하며 정확성을 높이는 데에 기여하고자 합니다. 또한, 텍스트 기반의 MBTI 예측이 적절한 성과를 보일 수 있다면 빅데이터를 통한 딥 러닝을 적용하여 모델의 성능을 우수하게 높일 수 있을 것으로 보며, 그 시발점을 연구하고자 합니다.

**1.2 프로젝트 목표**

위와 같은 단점을 극복하고자 자연어처리를 바탕으로 텍스트를 통해 답안의 자율성을 높이면서 정확성을 높이는 검사를 연구합니다.

또한, 이렇게 진행된 모델은 누구나 쉽게 체험해볼 수 있도록 UI형태로 결과물을 도출하도록 합니다.

토대가 되는 모델은 'MBTI Personality Prediction for Fictional Characters Using Movie Scripts' 논문이며, 이를 바탕으로 다양한 연구 사례 및 데이터 셋을 조사합니다.

또한, 여러가지 모델을 적용하며 여러 번의 실험을 거치고, 가장 적합한 모델을 이끌어냅니다.

Web Server을 따로 구현하여, 채택된 최종 모델을 연동하고 클라이언트에 제공하여 사용자가 UI를 통해 즐길 수 있도록 웹페이지를 구현합니다.

2. 프로젝트 내용

**2.1. <MBTI Personality Prediction for Fictional Characters Using Movie Scripts> 리뷰**

텍스트, 문서, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명텍스트, 스크린샷, 폰트, 문서이(가) 표시된 사진

자동 생성된 설명

**2.2. Dataset 조사**

- Kaggle 데이터셋 (https://www.kaggle.com/datasets/datasnaek/mbti-type)

텍스트, 문서, 스크린샷, 영수증이(가) 표시된 사진

자동 생성된 설명

- AIHUB 한국어 드라마 대본 데이터셋

텍스트, 스크린샷, 문서, 대수학이(가) 표시된 사진

자동 생성된 설명텍스트, 영수증, 폰트, 화이트이(가) 표시된 사진

자동 생성된 설명

기반이 되는 논문에서는 BERT classifier을 사용하였으며, Large-Scale 데이터 셋을 구축하고, 비언어적인 표현과 언어적 표현을 자동처리하는 Movie script parser을 개발하였다는 것이 특징입니다. 각 지표의 성능을 높이기 위해 XGBoost을 활용해 Decision Tree 4개를 만들어 MBTI을 예측하는 방향으로 모델을 설계하였습니다. 최종적으로, 구현한 모델의 특정 Classifier는 해당 논문에 비해 좋은 성능을 보였습니다.

3. 프로젝트의 기술적 내용

**3.1. 모델 구축 관련 기술적 내용**

- SW 및 HW 개발환경 소개

- 적용기술에 대한 소개 및 적용 방법

- 어려웠던 점과 그 해결과정 등 기술

-

**3.2. Web 구축 관련 기술적 내용**

개발환경은 MAC OS, 사용한 소프트웨어 툴은 VSCode입니다. 구현된 모델을 활용하여 웹 서비스로 제공하기 위해 DB와 Server가 요구되었고, 따라서 MongoDB을 이용하여 Database 시스템 구축 및 Django 프레임워크를 통한 중앙서버를 구축하였습니다. 사용자가 메인페이지에서 MBTI 검사하기 버튼을 누르면, 웹서버와 DB가 연동되며, 사용자가 input을 입력한 뒤 폼을 이동할 때마다 클라이언트에서 서버로 'GET' 리퀘스트를 보내 DB에 정상적으로 데이터를 저장하도록 하였습니다. 질문페이지에서는 데이터의 혼동을 방지하기 위해 SPA 프레임워크로 개발하였습니다. 마지막 문항까지 모든 답변이 저장되면 Input array을 구현된 모델에 인자로 넘겨 MBTI을 예측한 뒤 result 페이지로 라우팅합니다.

웹 서버를 구현하고, RESTful API을 도전해보는 경험 속에서 모르는 부분에 대해 공부하고 자잘한 오류를 해결하는 시행착오를 반복했는데, 오류를 만나면 검색을 통해 해결했고, 더불어 새로운 개념, 잊혀져 가고 있던 개념들에 대해서 다시 공부할 수 있었습니다. 웹에서 서버와 클라이언트가 데이터를 송수신 하는 과정과 동작원리를 다시 공부하면서 해결방법을 탐구해나갔습니다.

4. 프로젝트의 역할 분담

4.1 개별 임무 분담

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **번호** | **학과** | **학번** | **학년** | **이름** | **담당업무** |
| 1 | 컴퓨터소프트웨어 | 2017030082 | 4 | 이규민 | 데이터 조사 및 모델개발 |
| 2 | 컴퓨터소프트웨어 | 2018008304 | 4 | 박경하 | 데이터 조사 및 웹개발 |

4.2 개발 일정

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **구분** | **추진내용** | **수행기간** | | | | | | | | | |
| **9월** | **10월** | **11월** | **12월** | **1월** | **2월** | **3월** | **4월** | **5월** | **6월** | |
| 계획 | 사전 연구 조사(기본 개념 공부) |  |  |  |  |  |  |  |  |  |  | |
| 분석 | 논문 조사, 데이터셋 조사 |  |  |  |  |  |  |  |  |  |  | |
| 설계 | 주제 확정, 디자인 설계 |  |  |  |  |  |  |  |  |  |  | |
| 개발 | 모델 개발 |  |  |  |  |  |  |  |  |  |  | |
| 웹 개발 |  |  |  |  |  |  |  |  |  |  | |
| 테스트 | 모델 성능 테스트 |  |  |  |  |  |  |  |  |  |  | |
| 종료 | 실험 및 결과 정리 |  |  |  |  |  |  |  |  |  |  | |

5. 결론 및 기대효과

모델을 클라이언트에서 API 형태로 제공할 수 있는 완성된 결과물을 도출할 수 있다는 점과 자연어처리 기술을 통해 개발한 모델을 사용자들이 편리하게 접근할 수 있는 기회를 제공할 수 있다는 점에서 학술적, 산업적인 의의를 두고 있습니다.

시중의 MBTI 검사기와는 달리 서술형 형태이므로 유저들에게 비교적 신선함을 제공해볼 수 있는 기회가 되지 않을까, 그런 차원에서 다양한 경험의 폭을 늘려보는 좋은 기회를 제공해볼 수 있다고 생각합니다.

또한, 체계적으로 실험을 계획하고 모델을 구현하기 위해 공부 및 개발하는 과정에서 개인의 능력을 성장할 수 있는 좋은 기회였다고 생각합니다.

웹 서버는 차후 Node.js 프레임워크 툴을 이용한 SPA로 재구현해 볼 계획이며, 다른 모델을 기반으로 했을 때의 성능을 비교하며 더 효과적인 성능을 보이는 방안이 있는지 부가적으로 탐구해볼 계획입니다.