

요구사항 분석서



201711649 손소영
201512265 박인우
201814246 김예지
201710304 정창해

목차

1. 개요	1
1.1 프로젝트 기획 배경.....	1
1.2 기술 동향	1
1.2.1 감성 분석	1
1.2.2 추천 시스템	4
1.3 프로젝트 주요 기능 및 특징	5
1.3.1. 일기 작성 및 저장 기능	5
1.3.2. 감정분석 기능	5
1.3.3. 영화 추천 기능.....	6
1.4 조원 구성 및 역할 분담	6
1.5 일정.....	6
2. 기능적 요구사항	7
2.1 Top Level Use Case Diagram	7
2.2 각 기능 별 동작 시나리오(또는 Use Case Document)	7
2.2.1. 정의.....	7
2.2.2. 일기 입력	8
2.2.3. 일기 감정 분석.....	8
2.2.3. 감정 기반 영화 추천	8
3. 비기능적 요구사항	9
3.1 사용 편리성	9
3.2 신뢰성.....	9
3.3 성능.....	9
3.4 이식성.....	10
3.5 유지관리.....	10
3.6 구현 상 제약 사항	10
3.7 법적 제약 사항.....	10

1. 개요

1.1 프로젝트 기획 배경

현대인들은 복잡하고 빠르게 변화하는 사회 속에서 방대한 양의 감정을 소비하며 살아가고 있다. 소비하는 감정이 많음에도 불구하고 그들은 사회적 지위를 위해, 혹은 자신이 더 낫다고 생각하는 가치를 위해 본인의 감정을 억압하며 살아간다. 감정의 억압은 결국 정신 질환을 야기하고, 정신 질환은 더 심각한 사회적 문제를 초래할 수 있다.

감정을 표출할 줄 아는 것은 정신적으로 건강한 사회를 살아가는 데 있어서 중요한 요소이다. 우리가 가장 쉽게 감정을 표출할 수 있는 방법은 일기를 작성하는 것이다. 또한 영화와 같은 문화 콘텐츠를 시청하는 것도 하나의 방법이다. OTT 플랫폼이 활성화된 오늘날 우리는 여가 시간에 영화를 시청할 때 수많은 콘텐츠 중에서 하나를 골라서 시청을 하는데, 이때 선택 당시의 감정이 영화 선정에 영향을 준다.

이러한 점을 고려하여 본 팀은 사용자가 일기를 작성하면, 그 일기를 통해서 사용자의 감정을 분석하고, 그 감정에 맞춰서 영화를 추천하는 서비스를 개발하는 프로젝트를 기획하였다. 이 서비스를 통해 사용자는 자신의 감정을 표출하고, 부정적인 감정은 해소하며 정신 건강에 이로운 효과를 줄 것으로 예상한다.

1.2 기술 동향

감정을 기반으로 영화를 추천한 연구 사례인 ‘감정 온톨로지 기반의 영화 추천 기법¹’, ‘감성 분석을 이용한 협업적 영화 추천 방법²’에서는 영어 문장의 감성을 분석하고 IMDB의 영화 데이터를 기반으로 영화를 추천하였다. 본 팀은 이와 다르게 한글 문장의 감성을 분석하고 영화의 줄거리 데이터를 이용해 영화를 추천하는 시스템을 제작하고자 한다.

1.2.1 감성 분석

감성 분석은 텍스트에 들어있는 단어들을 통해 어떤 감정인지 분석하는 기법이다. 어휘 기반(Lexicon – based Approach)의 감성 분석, 머신 러닝 기반(Machine Learning Approach)의 감성 분석, 어휘 기반 접근 방법과 기계학습 접근 방법이 통합된 (Hybrid Method)의 감성 분석으로 분류된다.

¹ 김옥섭, 이석원.(2015).감정 온톨로지 기반의 영화 추천 기법.멀티미디어학회논문지,18(9),1068-1082.

² 박한샘, Abdel-ilah Zakaria Khiati, 강대현, 권경락, 정인정.(2014). 감성 분석을 이용한 협업적 영화 추천 방법. 한국정보처리학회 학술대회 논문집 21 권 1 호

1.2.1.1 어휘 기반(Lexicon – based Approach)

어휘 기반의 감성분석은 말뭉치 기반(corpus-based)와 사전 기반(dictionary-based)로 나뉜다. 말뭉치 기반(Corpus-based)은 거대한 말뭉치(데이터)를 기반으로 해당 말뭉치에 대한 적절한 감성 단어를 재구축하는 방법이다.

사전 기반(dictionary-based)은 수집된 텍스트를 분석하여 문서 내의 단어들을 통해 사전을 구축한다. 그런 다음 새로운 문서가 주어지면 구축된 사전을 적용시키는 방법이다.

1.2.1.2 머신 러닝 기반(Machine Learning Approach)

머신 러닝 기반의 감정 분석은 대표적으로 지도학습 기법(Supervised Learning), 비지도 학습 기법(Unsupervised Learning), 준지도 학습 기법(Semi-Supervised Learning), 자가 지도 학습(Self-Supervised Learning)이 있다.

지도 학습은 정답이 있는 데이터를 활용하여 컴퓨터를 학습시키는 것이다. 입력 값이 주어지면 입력 값에 대한 Label 의 형태로 학습이 진행된다.

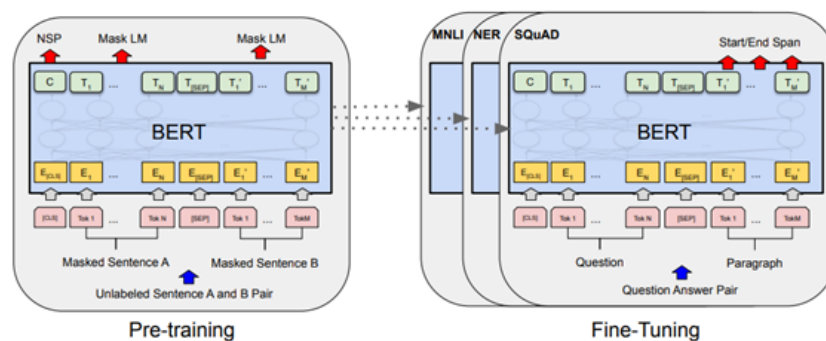
비지도 학습은 데이터에 대한 레이블(Label) 없는 비슷한 데이터들을 군집화 시키면서 학습하는 방법이다. 데이터의 숨겨진 특징이나 구조를 발견하는 데 사용된다.

준지도 학습은 지도 학습과 비지도 학습의 단점들을 보완한 방법이다. 적은 양의 labeled data 와 많은 양의 unlabeled data 를 학습 데이터로 사용한다.

자가 지도 학습은 간단한 문제(pretext task)와 그 문제에 맞는 정답(supervision)을 레이블(label)이 없는 데이터를 기반으로 학습한 후 downstream task(풀고 싶은 문제)에 맞는 정답이 존재하는 데이터에 label 을 생성하는 방식을 수행한다. 자가 지도 학습의 대표적인 예로 BERT 가 있다.

1) BERT (Bidirectional Encoder Representations from Transformers)

BERT 는 질문에 대한 대답, 텍스트 생성, 문장 분류 등과 같은 태스크에서 가장 좋은 성능을 도출해 자연어 처리 분야에 크게 기여한 기술이다.



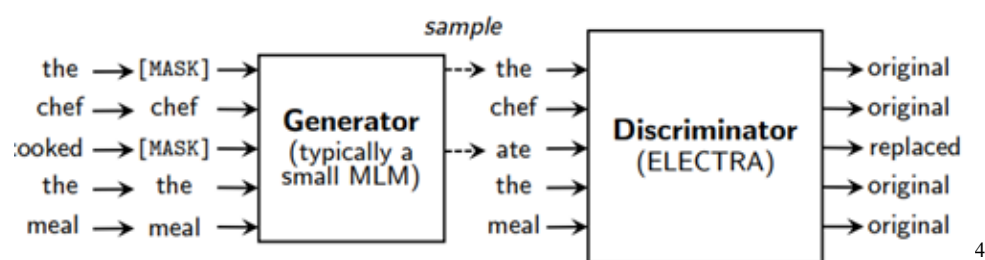
3

³ BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding Jacob Devlin Ming-Wei Chang Kenton Lee Kristina Toutanova (<https://doi.org/10.48550/arXiv.1810.04805>)

그림과 같이 Bert 는 크게 pre-training 단계와 fine-tuning 단계로 구분된다. MLM (Masked language model)과 NSP(Next sentence Prediction)라는 거대한 말뭉치를 기반으로 사전학습(pre-training) 된다. 새로운 태스크가 주어질 경우 BERT 를 처음부터 학습시키는 대신 사전학습 된 BERT 를 사용한다. 원하는 문제를 해결하기 위해 사전 학습된 BERT 를 기반으로 fine-tuning 을 하면 된다. Fine tuning 과정에서는 모든 파라미터(가중치 등)이 변화된다. 그렇기 때문에 보다 훈련 시간이 짧고 적은 데이터셋으로도 좋은 성능을 낼 수 있다는 이점이 있다.

2) ELECTRA (Efficiently Learning an Encoder that Classifies Token Replacements Accurately)

BERT 에서 파생 모델 중 하나이다. BERT 는 NLP 에서 좋은 성능을 내지만 많은 계산량을 요구하는 단점이 있다. 이를 해결하기 위해서 ELECTRA 에서는 사전 학습에 교체한 토큰 탐지(Replaced Token Detection)기법을 사용한다.



생성자(Generator)는 MLM 태스크를 수행한다. 15% 확률로 모든 토큰을 마스크된 토큰으로 교체한다. 마스킹한 이후에 생성기에 입력하고 나서 마스크된 토큰에 대한 예측 결과를 얻는다. 변경한 판별자(Discriminator)에서 토큰이 대체되었는지 아닌지를 판단한다.

1.2.1.3 Hybrid Method

하이브리드 접근 방식은 단일 모델 보다 좋은 효율성을 낼 수 있는 경우 다른 접근법이나 모델들을 조합해서 정확도를 높이는 방법이다

⁴ ELECTRA: PRE-TRAINING TEXT ENCODERS AS DISCRIMINATORS RATHER THAN GENERATORS Kevin Clark, Minh-Thang Luong, Quoc V. Le, Christopher D. Manning (<https://doi.org/10.48550/arXiv.2003.10555>)

1.2.2 추천 시스템

추천 시스템이란 특정 사용자가 관심을 가질 만한 정보를 추천하는 것이다. 추천 시스템의 종류로는 크게 콘텐츠 기반 필터링 (Content-based Filtering)과 협업 필터링 (Collaborative Filtering), 그리고 이 둘을 적절히 혼합한 하이브리드 필터링 (Hybrid Filtering)이 있다.

1.2.1.1 콘텐츠 기반 필터링

콘텐츠 기반 필터링은 사용자 또는 아이템에 대한 프로필 데이터를 통해 기존에 선호한 아이템과 비슷한 유형의 아이템을 추천하거나(Item-based Filtering), 비슷한 유형의 다른 사용자가 선호하는 아이템을 추천하는(User-based Filtering) 방법이다. 이는 다른 유저의 데이터가 필요하지 않으며, 개인의 독특한 취향을 고려한 추천이 가능하고, 새로운 아이템이나 비주류 아이템도 추천이 가능하다는 장점이 있다. 하지만 처음 유입된 사용자는 그의 선호에 대해서 모르기 때문에 추천이 어렵다는 단점이 있다.

1.2.1.2 협업 필터링

협업 필터링이란 많은 사용자로부터 수집한 취향 정보를 기반으로 추천하는 방법이다. 데이터에 어떻게 접근하는지에 따라 메모리 기반 접근법과 모델 기반 접근법으로 나뉜다.

여기서 메모리 기반 접근법은 또 사용자 간의 유사도를 측정하는 사용자 기반 협업 필터링과, 아이템 간의 유사도를 측정하는 아이템 기반 협업 필터링으로 나눌 수 있다. 사용자 기반 협업 필터링은 사용자의 군집을 확인하고, 특정 사용자의 아이템과의 상호작용 정보를 이용하여 같은 군집에 있는 유사한 사용자의 상호작용을 예측하는 것이다. 아이템 기반 협업 필터링은 사용자가 평가한 아이템의 군집을 확인하고, 이 아이템과 유사한 다른 아이템에 대한 사용자의 상호작용을 예측하는 것이다. 메모리 기반 접근법은 사용자-아이템 상호 작용 행렬이 희소하다는 문제가 있다. 행렬이 너무 희소하면 정확한 군집 형성이 힘들기 때문이다.

모델 기반 접근법은 머신 러닝, 데이터 마이닝 기술 기반으로, 모델을 학습시켜 예측하게 만드는 것이 목표이다. 이는 행렬이 희소할 때도 많은 사용자에게 많은 아이템을 추천 가능하다는 장점이 있다. 대표적인 모델 기반 접근법으로는 Matrix Factorization 이 있다. 이는 사용자의 선호도는 몇 개의 hidden factor(embedding)들로 결정될 수 있다는 것에서 출발한 것으로, 사용자-아이템 상호 작용 행렬을 두 개의 저차원 직사각형 행렬의 곱(latent factor 행렬)으로 분해한다.

협업 필터링은 도메인 지식이 필요가 없다는 장점이 있지만, 추가적인 특징(고객 개인정보, 아이템 추가 특징 등)을 포함시키기 어려우며, 충분한 정보가 없다면 추천이 힘들다는 cold-start problem 이 있다는 단점이 있다.

1.2.1.3 하이브리드 필터링

하이브리드 필터링은 콘텐츠 기반 필터링과 협업 필터링을 조합한 새로운 알고리즘으로, 기존의 두 추천 시스템이 갖고 있는 장점을 합하여 단점을 최소화하고자 만들어진 알고리즘이다. 기본적으로는 새로운 콘텐츠들은 콘텐츠 기반 필터링으로 분석을 진행하고, 충분한 데이터가 쌓이면 협업 필터링으로 정확성을 높이는 방식으로 사용된다.

1.2.1.4 딥러닝 기반

딥러닝 기반 추천 시스템은 기존 추천 시스템에서 많이 쓰이던 Matrix Factorization 이 latent factor 간의 곱셈을 선형으로 결합한다는 점에서 사용자-아이템 상호작용 데이터의 복잡한 구조를 알아내기 어렵다는 점과, 새로운 사용자가 나타나면 저차원 공간에서 이를 표현하기 어렵다는 점을 극복하기 위해서 고안된 방법론이다.

딥러닝 기반 추천 시스템은 비선형 변환이 가능하여 더 복잡한 사용자-아이템 상호작용을 파악할 수 있으며, 다차원 정보 처리가 가능하다. 기본적인 MLP, DNN 을 사용하는 알고리즘 외에도 RNN, CNN 등 다양한 딥러닝 알고리즘을 활용한 모델이 고안되었으며, 요즘은 딥러닝 기반 강화학습과 GNN(Graph Neural Network)도 추천 시스템 모델에 사용되는 추세이다.

1.3 프로젝트 주요 기능 및 특징

1.3.1. 일기 작성 및 저장 기능

사용자는 기본적으로 일기를 작성하고 저장할 수 있다. 글자수를 150 자로 제한하여 그 이상의 내용은 작성할 수 없다. 이 외에 사용자의 입력을 받아야하는 기능은 존재하지 않는다.

사용자가 작성한 일기는 앱을 사용하는 기기 내에 자체적으로 저장된다. 앱의 서버에는 사용자의 일기가 전달되지 않는다.

1.3.2. 감정분석 기능

해당 일기를 기반으로 사용자의 감정을 분석한다. 감정을 분석할 수 있는 모델로는 BERT, ELECTRA 등이 있는데, 다양한 모델을 통해 사용자의 감정을 분석하고 가장 신뢰도가 높은 모델을 실사용 할 것이다.

감정 분석모델은 앱 자체에서 보유한다. 따라서 서버에 따로 일기 데이터를 보내주지 않아도 사용자의 감정을 분석하는 것이 가능하다. 서버에 보내는 것은 감정분석의 결과 하나이다.

1.3.3. 영화 추천 기능

분석된 감정 데이터를 기반으로 현재 분위기에 어울리는 영화를 추천한다. 영화 추천데이터는 서버에 존재한다. 따라서 사용자가 서버로 보낸 감정분석 데이터를 기반으로 적절한 영화를 선정해 그 결과를 사용자에게 전달한다.

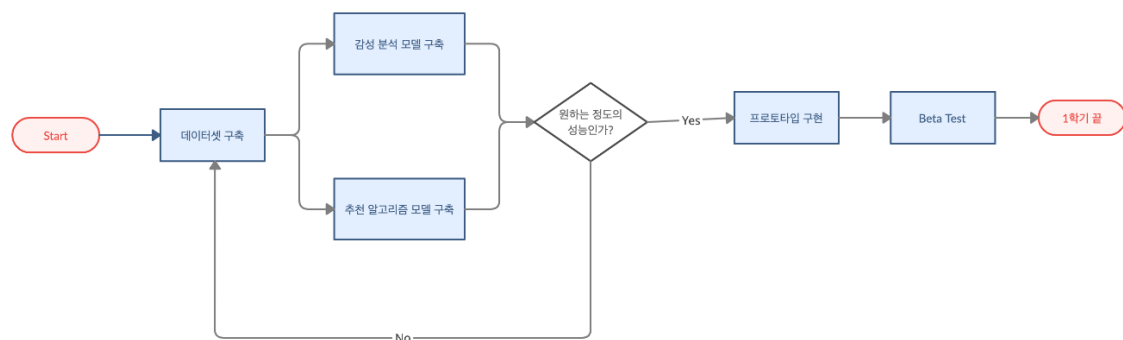
앱에서는 해당 결과를 기반으로 영화추천 결과를 사용자에게 보여준다. 사용자는 추천된 영화 중 보고싶은 영화를 체크할 수 있다. 해당 정보 역시 서버에 전달된다.

1.4 조원 구성 및 역할 분담

정창해	팀장 및 일정관리 감정 분석 기반 추천 시스템 설계 및 구현
손소영	사용자 일기 기반 감정 분석 모델 설계 및 구현
김예지	사용자 일기 기반 감정 분석 모델 설계 및 구현
박인우	감정 분석 기반 추천 시스템 설계 및 구현 테스팅 앱 설계 및 구현

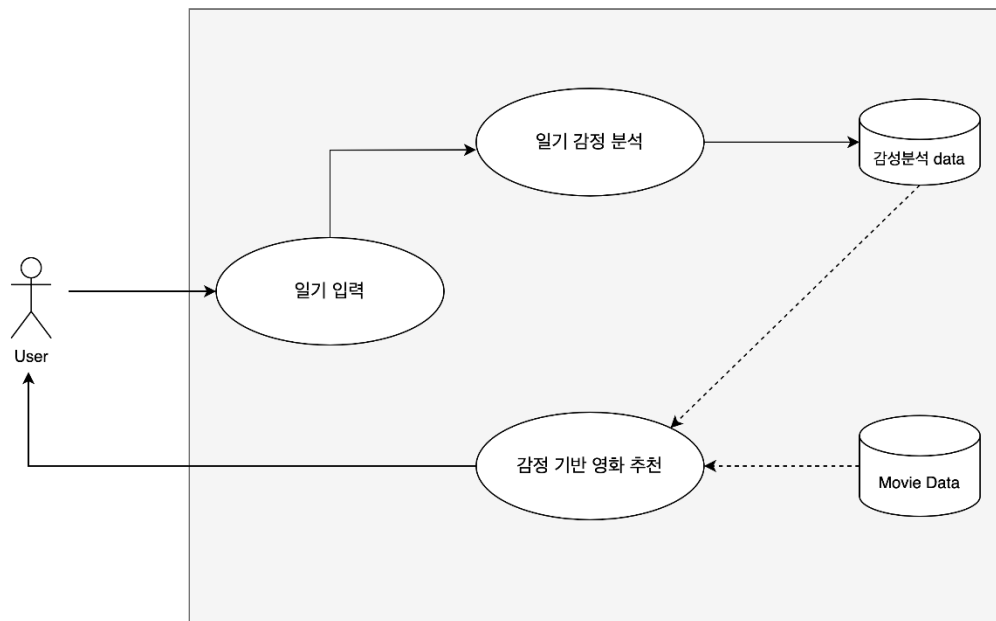
1.5 일정

7 주차 ~ 8 주차 (04/13 ~ 04/26)	기본 모델 구축
9 주차 ~ 11 주차 (04/27 ~ 05/17)	모델 성능 향상을 위한 분석 및 수정
12 주차 ~ 13 주차 (05/18 ~ 05/31)	프로토타입 실현, Beta 테스터 모집



2. 기능적 요구사항

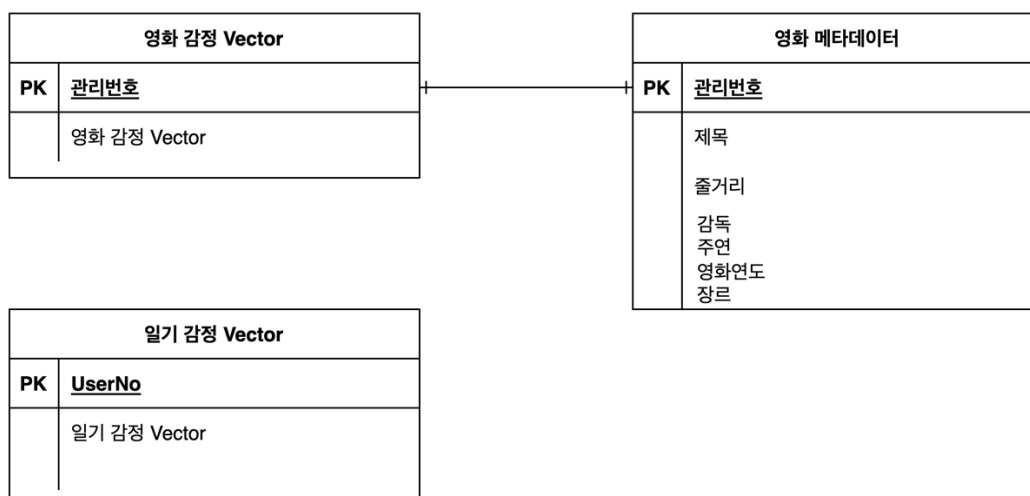
2.1 Top Level Use Case Diagram



2.2 각 기능 별 동작 시나리오(또는 Use Case Document)

2.2.1. 정의

Word	Description	Example / Type
일기 감정 벡터	일기를 분석해 각 감정의 확률이 저장된다.	[0.2,0.4,..0.6, ...] / Vector
영화 감정 벡터	영화 메타데이터를 분석해 각 감정의 확률이 저장된다.	[0.2,0.4,..0.6, ...] / Vector



2.2.2. 일기 입력

Use case name	일기 입력
Participating actors	User
Flow of events	
Actor Actions	System Responses
1. User 가 System 에 일기를 입력한다.	
	2. 일기 데이터를 저장하고 일기감정분석으로 넘어간다.

2.2.3. 일기 감정 분석

Use case name	일기 감정 분석
Participating actors	User
Flow of events	
Actor Actions	System Responses
	1. 저장된 일기 데이터를 감정 분석 모델에 insert 한다.
	2. 일기 데이터의 일기 감정 벡터를 출력한다.
3. User 에게 일기 감정 벡터를 기반으로 감정을 보여준다.	

2.2.3. 감정 기반 영화 추천

Use case name	감정 기반 영화 추천
Participating actors	User
Flow of events	
	1. 영화 메타데이터를 이용해 영화 감정 벡터를 생성한다.
	2. 일기 감정 벡터와 영화 감정 벡터를 추천시스템에 insert 한다.
	3. 일기 감정 벡터와 영화 감정 벡터의 유사도를 기반으로 영화를 추천해준다.
4. User 가 추천된 영화를 확인한다.	

3. 비기능적 요구사항

3.1 사용 편리성

- 사용자에게 익숙한 아이콘을 통해 직관적이고 깔끔한 사용자경험을 제공한다.
- 사용자의 입력을 요구하는 일기 입력 부는 최대 문자열 및 안내문구를 제공하여 사용자가 입력에 어려움을 겪지 않도록 한다.
- 일기에 대한 감정을 차트 형태로 제공하여 사용자가 한 눈에 현재 감정 상태를 확인할 수 있도록 한다.
- 왜 해당 영화가 추천되었는지에 대한 안내문구를 제공하여 해당 영화의 추천 사유를 파악할 수 있도록 한다.

3.2 신뢰성

- 모델 별로 감정 분석 신뢰도를 분석하여 보다 명확한 감정상태를 파악할 수 있는 모델을 선정한다.
- 베타 테스트 기간을 운영하여 추천시스템의 신뢰도를 파악한다.
- 배포 전 충분한 기능 테스트를 통해 사용자 경험에 있어서의 버그를 방지한다.
- 배포 후 사용자가 버그 및 성능 개선 리포트를 할 수 있도록 한다.

3.3 성능

서버에서 데이터를 받아올 때 일정시간(3 초) 이상 서버에서 응답이 없을 경우 서버로부터 데이터를 받아올 수 없음을 알리고 임의의 데이터를 출력한다.

감정 분석 모델의 성능은 정확도 70 을 목표로 한다. 단, 해당 데이터를 기반으로 감정 분석 모델을 제작한 선례가 없기 때문에 확정할 수 없다.

추천시스템은 베타 테스트로부터 평가를 받아 정확도를 계산한다. 감정 분석을 통한 영화 추천 시스템은 영화 리뷰 데이터를 사용한 선례가 존재하지만, 본 팀은 영화와 전혀 무관한 일기 데이터를 감정 분석하여 영화를 추천해주므로 해당 일기와 영화 시청을 매핑할 수 없다. 또한 이렇게 일기를 통해 영화를 추천하는 선례가 존재하지 않으므로 벤치마킹 할 선행 연구 또한 없다. 그렇기 때문에 추천시스템의 성능 평가는 베타 테스트로부터 평가를 받아서 진행된다.

사용자는 추천된 영화 중 보고 싶은 영화만을 체크해 서버로 전송한다. 추천된 영화 중 사용자가 보고싶은 영화가 하나라도 생기는 것을 목표로 한다. 베타 테스트의 평가는 일기의 감성분석의 결과가 어느정도 맞는지 지표 0~5 점, 사용자가 선택한 영화의 개수로 0 개에서 5 개의 점수로 구성된다. 이를 통해 감성분석과 추천시스템의 성능을 평가한다. 최종 평가 방식은 위와 동일하게 진행되며, 팀원 4 명과 전혀 관련 없는 사람으로 구성한다.

3.4 이식성

- iOS 를 사용하는 플랫폼에서 사용할 수 있도록 한다.
- View 의 Layout 이 기기마다 유동적으로 변경되도록 하여 모든 iOS 기기에서 유사한 UI 배치를 확인할 수 있다.
- 웹 및 안드로이드 플랫폼에서는 사용할 수 없다.

3.5 유지관리

- 추후 앱 사용성 개선과 모델의 성능 개선을 위해 Git 으로 지속적인 버전관리를 진행한다.
- 모델의 지속적 발전을 위해 문서 정리 및 코드 공유를 지속적으로 수행한다.
- 앱의 지속적 유지보수를 위해 추후 관리가 용이한 아키텍처로 앱을 설계한다.

3.6 구현 상 제약 사항

- 모델 학습을 위해 GPU 성능이 충분한 서버를 대여한다.
- iOS 앱 테스트를 위해 다양한 iOS 기기를 필요로 한다.

3.7 법적 제약 사항

- 데이터 상업적 이용 금지
- 일기 데이터 무단 사용 금지