

종합설계1

졸업 프로젝트 - 설문조사 분석 보고서

주제 : 프라이버시 보호 딥러닝 서비스 개발

[사그램 조]

201402433 조승현 201402392 이상화

201704144 김수민 201704145 김주희

● 설문 개요

본 설문조사는 15명의 응답자를 대상으로 사그램조가 구현한 프로토타입에 대한 사용자들의 의견을 분석해보기 위하여 기획되었습니다. 이를 통해 문제점이나 개선해야 할 점을 참고하여 해결책을 찾아 보 완해보고자 합니다.

- 설문 응답자 : 15명
- 설문 기간 : 2020.05.28 – 2020.05.29 (2일간)
- 설문 방법 : Google 설문지

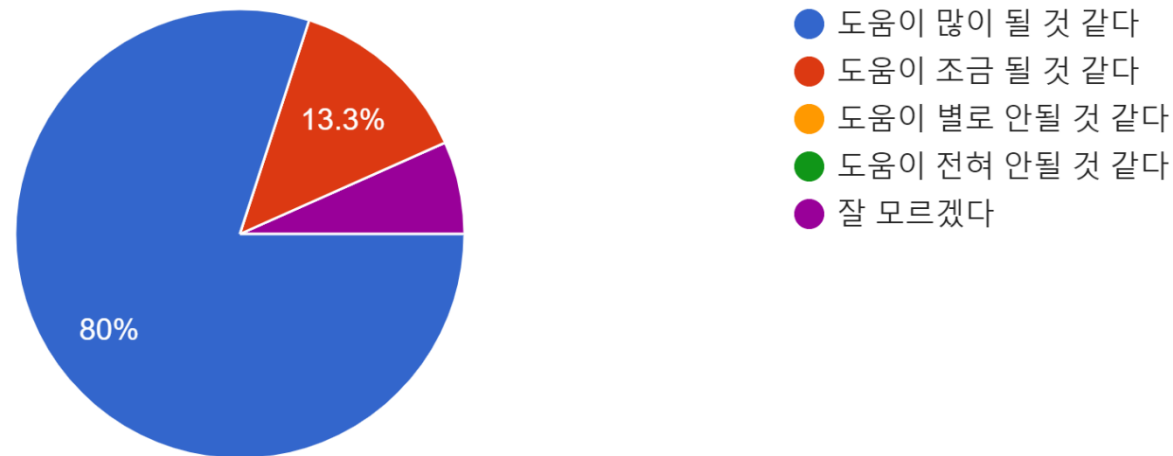
● 설문 문항

- 해당 모델이 사용자의 데이터를 분석하여 적절한 상품을 추천해 주는 어플리케이션을 개발하는데 얼마나 도움이 된다고 생각하시나요?
(도움이 많이 될 것 같다/도움이 조금 될 것 같다/도움이 별로 안될 것 같다/도움이 전혀 안될 것 같다/ 잘 모르겠다)
- 해당 모델의 문제점(부족한 점)이 있다면 무엇인가요?
- 해당 모델로 구현된 서비스를 사용할 의향이 있으십니까?

● 설문 결과

1. 해당 모델이 사용자의 데이터를 분석하여 적절한 상품을 추천해 주는 어플리케이션을 개발하는데 얼마나 도움이 된다고 생각하시나요?

응답 15개



15명의 응답자 중, 도움이 많이 될 것 같다 12명/도움이 조금 될 것 같다 2명/잘 모르겠다 1명

⇒ 도움이 될 것 같다는 응답 (93.3%) 多 !!

2. 해당 모델의 문제점(부족한 점)이 있다면 무엇인가요?

- 데이터 분류시에 특징을 조금 더 디테일하게 분류할 수 있으면 좋겠다.
- 오류가 자주 발생할 것 같다.
- 사용하는 고객 수가 많아지면 데이터의 양도 증가할 것인데, 이 증가에 비례하게 계산비용도 증가한다면 하나의 프로그램에 데이터가 많을 때 정확성이 떨어지는 문제점이 발생하지 않을까 하는 생각이 든다.
- 오차를 줄이기위해 3~4단계를 반복한다는 점이 비효율적이라고 생각한다. 만약 데이터가 많아진다면 더욱 번거로운 일이 될 것으로 예상된다. 한 두번 정도의 실행으로 오차값을 줄일 수 있는 방법을 구현하는 것도 좋을 것 같다.
- 데이터의 차원 자체가 너무 커서 쉽지 않을 것 같다.
- 영상을 통해서 어떤 데이터 셋을 이용하여 테스트를 한 것인지 모르겠지만 기존의 학습데이터 셋을 그대로 넣어서 딥러닝 한 경우 학습데이터가 아닌 테스트 데이터를 넣어서 딥러닝 했을 때 오차율이 지금보다 크게 나올 것 같다.

이 외의 응답은 “문제점(부족한 점)이 없다”

3. 해당 모델로 구현된 서비스를 사용할 의향이 있으십니까?

- 학습된 데이터로 최적의 상품을 추천해준다는 점에서 사용할 의향이 있다.
- 개인정보는 보호하고, 시간을 단축하여 필요한 상품을 가입할 수 있을 것 같아 사용할 의향이 있다.
- 요즘 사람들이 평균적으로 10분에 한번 핸드폰을 본다는 결과가 있듯 핸드폰 사용시간이 늘어나면서 편리하게 이용할 수 있는 모바일 금융서비스의 이용률도 증가하였다. 이 서비스를 역시 편리하게 사용할 수 있을 것 같아 사용할 의향이 있다.
- 금융 관련 정보나 지식이 부족한 분들에게 좋을 것 같다. 아무래도 금융 서비스 이다보니 민감한 정보가 유출되는 걱정을 덜 수 있어서 좋은 모델이라고 생각한다.
- 금융 상품을 선택할 때 내 소비패턴을 판단하여 결정하기까지 쉽지 않았는데 딥러닝을 통해 분석해준다면 보다 빠르게 적절한 판단을 할 수 있을 것 같다.
- 데이터를 분류하여 추천해주는 과정 속에 정보가 유출될 수도 있고 악용될 수도 있을 것이라는 불안감을 가질 수도 있지만, 데이터를 암호화하여 추천된 상품이라 하면 안심되고 해당 서비스에 대한 신뢰감도 생길 것 같다.
- 정확도가 좀 더 높아지고 개인정보 보호만 확실해 진다면 사용할 것 같다.
- 평소에 나에게 맞는 상품을 얻기가 쉽지 않았고 개인정보보호도 중요시 여겨왔기 때문에 서비스를 잘 이용할 것 같다.

● 생각해 볼만한 점

미니배치 K-평균 군집화(Mini-batch K-means Clustering)

K-Means 방법은 중심위치와 모든 데이터 사이의 거리를 계산해야 하기 때문에 데이터의 양이 많아지면 계산량도 늘어나게 된다. 이처럼 데이터의 수가 많은 경우에는 데이터를 미니배치(Mini-batch) 크기만큼 무작위로 분리하여 K-Means를 수행하는 '미니배치 K-Means'로 계산량을 줄일 수 있다. 계산량의 감소로 속도 역시 훨씬 빨라진다. 사이킷런의 cluster 서브 패키지에서 제공하는 MiniBatchKMeans를 사용할 수 있다.

차원 축소(Dimensionality Reduction) – PCA, SVD

사용할 데이터와 feature가 많으면 비지도 학습의 경우 어떤 feature를 사용해야 할지 상당히 난감하다. 이렇듯 고차원 데이터에서 아무런 처리 없이 비지도 학습인 군집 분석을 시도하면 성능이 저하될 수 있다. 따라서 차원 축소를 수행해 주어야 하는데, 차원을 축소해주는 기법인 PCA나 SVD를 이용할 수 있다.

PCA (주성분 분석) : 여러 변수간에 존재하는 상관관계를 이용하여 이를 대표하는 주성분을 추출하여 줄

SVD (특이값 분해) : 임의의 고유값 분해를 직사각형 행렬에 대해 일반화하는 방법