

1. **과제의 필요성**
   1. **Abstract**

IT 기술은 나날이 진화하고 있지만 한국 정치인의 홍보 수단은 2000년대에 머물고 있다. 이미 미국에서는 2012년 대선 당시 버락 오바마 대통령 캠프가 빅데이터 분석을 이용해 유권자 맞춤형 공약 홍보 전략을 실시했으며 그 전략은 오늘날에도 사용되고 있다. 본 졸업작품에서는 오바마 캠프에서 활용한 선거 전략과 유사하게 유권자의 소셜 네트워를 분석하여, 더 자세히는 텍스트 마이닝 기법, 소셜네트워크 분석 기법 등을 이용하여 유권자 성향을 파악한 후, 이에 맞는 정치인의 공약을 추천해주는 기능을 제공하고자 한다. 본 졸업작품을 통해 공약을 효과적으로 전달함으로써 투표율 증가의 효과를 기대해볼 수 있다. 또한 본 졸업 작품을 활용하여, 다양한 홍보물 발송을 통해 사용자에게 맞는 공약을 직접 전달하는 시스템 개발 역시 기대할 수 있다.

* 1. **연구배경**

바야흐로 IoT 시대가 도래하면서, 빅데이터 기술은 4차 산업혁명의 핵심인 인공지능의 필수 요소로써 많은 각광을 받고 있다. 인공지능의 주요 기법인 머신 러닝이 빅데이터 기술을 활용하고 있으며, 빅데이터 기술력이 곧 인공지능 기술력이라 할 수 있을 만큼 그 중요성이 날이 갈수록 커지고 있다. 미국과 영국 정부는 빅데이터 기술을 통해 각각 연간 300조 원과 33조 원의 정부 예산 비용 절감이 예상된다면서 향후 빅데이터 기술에 막대한 투자할 계획이라 발표했고, 한국 정부 또한 향후 빅데이터 전문가를 대거 양성하겠다 발표한 상태다.

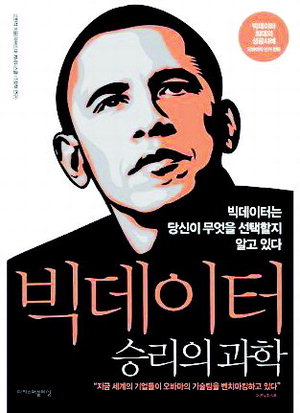
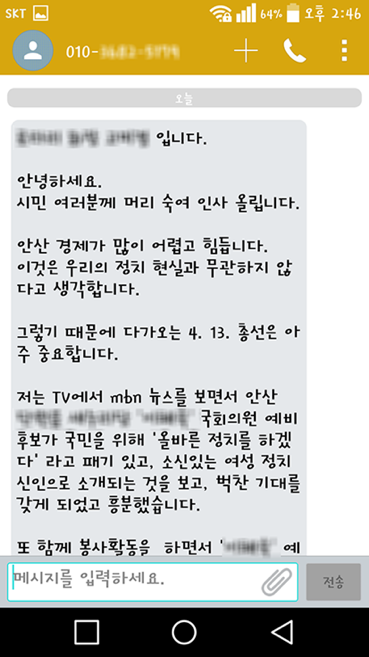
2012년 미국 대선 당시, 오바마 대통령 캠프는 유권자를 다양한 데이터 마이닝 기법을 통해 매우 상세한 데이터 셋으로 분류하였다. 그 후 이들의 신분 및 성향 등을 분석한 후 자신의 공약 중 특정 유권자에게 어떠한 공약이 가장 적절한지를 도출한 다음 홍보를 함으로써 오바마 미국 대통령이 재선되는 데 지대한 영향을 끼쳤다. 이 사례는 빅데이터 활용의 대표적인 선례로 꼽혔으며 이후 미국 정치인들이 빅데이터를 적극적으로 활용하는 밑거름이 되었다.

그림 2 총선/대선 기간의

무차별 문자 전송

그림 1 『빅데이터: 승리의 과학』, 오바마 대선 캠프의 빅데이터

전략은 책으로까지 소개되었다.

본 작품의 목적은 미국 정치에서 대대적으로 사용되고 있는 빅데이터 기술을 한국 정치에도 사용하는 것이 어떤가 하는 생각에서 출발하였다. 대선/총선 기간만 되면 “수원의 자랑! 수원의 아들! 김ㅇㅇ 오늘도 열심히 뛰겠습니다!”와 같은 식상한 문자가 유권자들 휴대폰에 무차별적으로 전송된다.

하지만 이런 문자는 유권자의 시선을 끌지 못하고 정당 후보들에게도 별다른 도움을 주지 못한다. 오히려 식상한 텍스트를 복사 및 붙여넣기 형식으로 유권자에게 보내는 과정에서도 불필요한 시간, 재정적인 낭비가 발생할 것이다. 유권자들 또한 자신에 맞는 정치인의 공약을 알지 못해 정치에 대한 무관심으로 이어질 수 있다. 국력은 투표에서 나온다 할 정도로 투표율은 해당 국가 국력에 중요한 요소 중 하나이다. 그러므로 비효율적인 홍보 방식을 철폐하고 국민의 정치에 대한 관심을 증대시키기 위해, 나아가 국가의 미래를 위해 빅데이터를 적극 활용하여 대선/총선 후보들의 공약을 맞춤형 방식으로 도출하는 기술을 국가 차원에서도 적극적으로 활용해야 할 것이다.

1. **선행연구 및 기술현황**
   1. **선행연구**

2012년 미국 대선에서 버락 오바마 대통령이 마이크로타케팅을 활용한 성공적인 선거 전략을 선보인 사례가 있다. 오바마 캠프는 SNS를 통해 유권자들의 성향을 파악하여 개인 맞춤형 공약을 제공하는 등 빅데이터를 활용한 전략을 사용하였다. 유권자가 오바마 캠프 페이스북 페이지에 ‘좋아요’를 누르게 되면 캠프는 유권자의 페이스북 정보에 접근할 수 있게 된다. 이렇게 개인 SNS 정보를 통해 해당 유권자의 성향에 대해 일일이 분석을 하였다. 또한 유권자가 캠프 홍보 홈페이지에 접속하면, 마우스의 클릭 기록을 모두 데이터화하여 유권자의 관심사를 분석하였다.

오바마 캠프는 이렇게 분석한 유권자의 성향을 개인 맞춤형 메일을 보내며 홍보를 하는 데에 활용하였다. 예를 들어 유권자 성향에 맞춰 개를 좋아한다고 분석된 유권자에게는 오바마가 키우는 개에 관련된 소식을 제공하고, 환경문제를 중요시 생각하는 유권자에게는 풍력산업에 관한 공약을 제공하였다. 이런 오바마 캠프이 전략을 알아챈 다니엘 싱커 교수는 심층분석전문 언론사와 함께 오바마 캠프가 보낸 메일이 어떻게 다른 지 조사해보았다. 조사 결과, 약 2만 여 통의 메일 중 800개의 메시지와 1500가지의 개인별 변형이 있다는 사실을 알게 되었다. 이렇듯 개인 맞춤형 전략을 통해 유권자가 관심이 있는 분야에 대한 메일을 보내는 효율적인 전략을 사용하였다.

* 1. **기술현황**

다음과 같은 빅데이터 관련 기술들이 활용되고 있다.

* + 1. **데이터 마이닝(Data Mining)**

데이터 마이닝 기법은 빅데이터 관련 기술 중 하나로 수많은 데이터를 분석하여 상관관계를 도출하여 내고, 미래에 대한 예측을 할 수 있게 해주는 기술이다. 이 기술은 다른 말로 KDD(Knowledge-Discovery in Database)라고 불린다. 이 용어 역시 해당 기술의 핵심이 대용량 데이터베이스로부터 새로운 패턴을 찾아 내는 것이라는 것을 나타낸다. 데이터 마이닝은 상관관계를 도출하기 위해 통계학의 패턴 인식 기법 등 다양한 기법을 사용한다. 데이터 마이닝의 기법으로는 탐색적 자료분석, OLAP, 신경망, SOM 등의 방법이 있다.

* + 1. **텍스트 마이닝(Text Mining)**

텍스트 마이닝이란 자연어로 구성된 대규모의 비정형 텍스트에서 의미 있는 정보를 추출하는, 자연어 처리기술에 기반한 기법을 말하며, 큰 의미로 텍스트 분석과 같은 말이라고 할 수 있다. 텍스트 분석은 다양한 분야의 텍스트를 분석하는 머신 러닝 기술을 말한다. 텍스트 분석은 오랜 시간 동안 연구되어 온 분야이지만, 자연어를 처리해야 하는 기술이기 때문에, 언어의 복잡성으로 인해 아직까지도 연구가 활발히 이루어지고 있는 분야이다. 텍스트 마이닝 기법의 주요 연구 분야로는 문서 분류, 문서 군집, 정보 추출, 문서 요약 등 다양한 분야가 존재한다.

본 졸업작품 팀이 제안하는 시스템은 SNS에서 생산되는 데이터를 분석해야 한다. SNS의 데이터들은 비정형 데이터이기 때문에 데이터 분석을 할 때 텍스트 마이닝을 활용할 수 있다.

* + 1. **하둡(Hadoop)**

하둡이란 빅데이터를 처리하기 위한 분산파일 시스템이다. 아파치 루씬의 창시자 더그 커팅이 개발하여 2005년에 공개하였다. 하둡은 맵리듀스를 사용하여 빅데이터를 빠른 시간 안에 처리할 수 있다. 하둡은 오픈소스라는 장점과 정형 데이터뿐만 아니라 비정형 데이터 역시 빠른 시간 내에 처리할 수 있다는 장점이 있어 널리 쓰이고 있다.

1. **작품/논문 전체 진행계획 및 구성**

**[표 1] 논문 월별 진행계획**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **월별**  **내용** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** |
| **제안서 작성 및  사전 조사** |  |  |  |  |  |  |  |  |  |
| **관련 기술 조사** |  |  |  |  |  |  |  |  |  |
| **SNS 데이터 수집** |  |  |  |  |  |  |  |  |  |
| **빅데이터 기술을 활용한 사용자**  **분석** |  |  |  |  | |  |  |  |  |
| **맞춤형 공약  시스템 구축** |  |  |  |  |  | | |  |  |
| **중간 보고서 작성** |  |  |  |  |  |  |  |  |  |
| **디버깅 및  결론 도출** |  |  |  |  |  |  |  | | |
| **최종 보고서 작성** |  |  |  |  |  |  |  |  |  |

본 졸업작품 팀에서 제안하는 시스템은 오바마 대통령이 활용한 선거 전략 중 SNS를 활용하여 유권자들에게 개인 맞춤형 공약을 제공한다는 전략을 활용하고자 한다. 개인 SNS 분석을 통해 유권자들이 관심 있을 만할 공약을 맞춤형으로 제공하는 시스템을 구축하여 우리나라에서도 효율적인 공약 홍보 시스템을 제공하고자 한다.

* 1. **제안서 작성 및 주제 관련 조사**

졸업 작품을 진행하기 전 제안서 작성 및 작품 진행에 관련된 사전 조사를 한다. 추후에 자세하게 진행할 SNS 데이터 수집 방법 연구, 공약 카테고리 선정, 데이터 마이닝 알고리즘 연구 등에 대해 사전 조사를 진행한다.

* 1. **관련 기술 조사**

관련 기술 조사는 사전 조사 시에 다룬 내용을 바탕으로 각각의 항목들에 대해 세부적인 조사 및 연구를 진행한다.

* 1. **SNS 데이터 수집**

빅데이터의 기술을 활용하여 SNS 데이터를 수집한다. 트위터와 페이스북에서 제공하는 API 중 목적에 맞는 API를 선정 및 활용하여 데이터를 수집한다.

* 1. **빅데이터 기술을 활용한 사용자 분석**

수집된 데이터에 빅데이터 기술을 활용하여 사용자를 분석한다. 다양한 빅데이터 기술 중 목적에 맞는 기술을 선정하여 사용자의 연령대, 성향, 관심사 등을 분석한다.

* 1. **맞춤형 공약 시스템 구축**

사용자 분석 데이터를 바탕으로 해당 사용자가 관심 있을 만한 공약을 매칭시켜주는 맞춤형 공약 시스템을 구축한다.

* 1. **디버깅 및 결론 도출**

시스템 동작 시 발견되는 오류를 수정하여 시스템이 정상적으로 동작할 수 있게 한다. 또한 졸업 작품 진행 결과에 대해 결론을 도출하며 futurework를 제안한다.

1. **기대효과 및 개선방향**
   1. **기대효과**

본 작품을 통해 기대되는 효과는 바로 투표율의 증가, 장기적으로는 국민의 정치에 대한 관심 향상이다. 서론에서 서술하였듯이 정치인들의 무차별 문자 폭격은 국민의 정치에 대한 관심을 떨어뜨리는 역효과만 내고 있다. 하지만 국민 개개인의 정보 및 성향을 분석하여 맞춤형 공약을 홍보한다면 국민은 분명 자신이 지지해야 할 정치인을 선택할 것이며 이는 정치에 대한 국민 관심 증대로 이어질 것이다. 지역, 정치 성향만으로 정치인을 선택하는 것이 아닌 자신을 위한 정치인을 선택하는 것이다.

예컨대 임산부로 판단되는 유권자에겐 정치인의 육아 관련 정책을 주로 홍보하고, 취업 준비생으로 판단되는 유권자에겐 취업 관련 정책을 주로 홍보함으로써 국민이 자신의 미래를 책임질 만한 정치인을 선택하게끔 하는 것이다.

* 1. **개선방향**

본 작품의 개선 방향으로는 사용자 데이터의 수집 범위를 더 확대하고 데이터 셋을 더욱 세분화하는 것이다. 향후 소셜 네트워크뿐이 아닌 다른 분야, 예컨대 정부에서 제공하는 유권자 정보를 바탕으로 군집을 더욱 세분화하고 마지막에는 개개인의 성향을 정확히 파악해 정치인의 공약을 일인 맞춤형으로 매칭시키는 것이다.

또한 사용자에게 맞는 공약을 그저 도출하는 것이 아닌 문자 등 홍보물 발송 서비스 개발도 본 작품의 개선방향 중 하나이다. 안드로이드와 ios의 현 정책으로는 푸시 알림을 사용자 휴대전화로 보낼 시 까다로운 절차를 밟아야 한다. 이러한 한계 때문에 현재는 그저 데이터를 도출하는 것에 그치고 있다. 이를 해결하기 위해 문자 푸시 알림 서비스와 카카오톡 푸시 알림 서비스를 추가하여, 전화번호가 확보된 사용자에게 맞춤형 공약을 메시지로 발송하는 방안이 있다.

1. **기타**
   1. **팀원간의 역할분담**

팀원간의 역할분담은 아래 [표 2]에 나타내었다.

**[표 2] 역할분담**

|  |  |
| --- | --- |
| **성명** | **역할** |
| **전지윤** | 분석 알고리즘 구현 및 시스템 구축 |
| **두건** | SNS 데이터 수집, 데이터베이스 구축 |
| **공통과제** | 기술 조사, 공약 조사 |

* 1. **진행계획**

진행계획은 아래 [표 3]에 나타내었다.

**[표 3] 진행계획**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **월별**  **내용** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** |
| **연구노트 작성** |  | | | | | | | | |
| **작품 제안서**  **서약서** |  |  |  |  |  |  |  |  |  |
| **작품 중간보고서** |  |  |  |  |  |  |  |  |  |
| **작품 최종보고서** |  |  |  |  |  |  |  |  |  |
| **작품 발표회** |  |  |  |  |  |  |  |  |  |

1. **참고문헌**

[1] 윤미영, “주유국의 빅데이터 추진전략 분석 및 시사점”, 한국정보화진흥원, 9월 2013.

[2] 고한석 (2013), 『빅데이터: 승리의 과학』, 이지스퍼블리싱

[3] 송규봉, 박용재 <http://publish.gisutd.com/?p=3287>

[4] 데이터 마이닝 <https://en.wikipedia.org/wiki/Data_mining>

[5] 텍스트 마이닝 <https://en.wikipedia.org/wiki/Text_mining>