OCR과 패턴분석 알고리즘을 활용한 인공지능 기반 기록 자동화 서비스 제안

황윤영
한국폴리텍I대학 서울강서캠퍼스 데이터분석과
e-mail:duddl425@naver.com

Proposal Record Automation Service Based on AI by Using OCR and Pattern Analysis Algorithm

Yun-Young Hwang
Department of Data Analysis, Campus of Seoul Gangseo, Korea Polytechnics

요 약

제안하는 서비스는 OCR(Optical Character Recognition, 광학문자인식)과 딥러닝 패턴분석 알고리즘을 활용하여 문서를 효율적으로 관리하는 서비스로 필기를 많이 하는 사용자를 위한 기능을 제공한다. 최근 다양한 분야에서의 머신러닝 기반의 OCR의 활용이 증가했지만 기존의 애플리케이션은 패턴 분석 알고리즘과 통계 기반의 OCR을 혼합하여 사용하기 때문에 필기체에 대한 인식률이 높지 않다. 이에 본 논문에서는 OCR과 패턴분석 알고리즘을 활용하여 필기체에 대한 높은 인식률을 제공하는 서비스를 제안한다.

1. 서론

2019년 3월, 한국인터넷진흥원 발표에 따르면 OCR을 활용한 이미지 검색 기능을 도입함으로써 이미지에 포함된 문자를 추출해 개인정보를 확인하는 방법으로 보안 문제를 해결할 수 있다고 한다. 또한 머신러닝 기반의 OCR을 활용한 명함인식 애플리케이션인 '리멤버'는 2014년 출시 이후 가입자 수가 2017년에 200만 명을 넘어섰다. 이와 같이 다양한 분야에서의 머신러닝 기반의 OCR 활용도 및 사용자수가 증가하고 있다.

미국의 골드만삭스는 인공지능 솔루션을 도입하여 자연어 분석을 통해 경제지표, 실적 등을 학습하고 자동으로 문서를 생성하는 시스템을 구축했다. BC카드는 데이터 분석, 시각자료 작성 등을 머신러닝 기반으로 자동화하는 시스템을 구축하는 등 RPA(Robotic Process Automation, 로보틱 처리 자동화)를 활용한 문서 자동화 시스템이 발달하고 있다.

또한, 메모 애플리케이션에 대한 사용자의 관심이 증가하고 있다. 미국 모바일 앱 통계 툴 및 구글의 통계에 따르면, 메모 애플리케이션인 솜노트의 경우 2016년 5월 실행량이 약 700만 회로 집계되었다. 같은 유형의 애플리케이션인 에버노트의 가입자 수는 2016년 2억 명을 넘어섰으며, 전체 사용자의 76% 이상이 미국 이외의 글로벌 사용자이고 한국의 이용자 수는 450만 명 이상으로 집계되었다.

한국 교육 개발원에 따르면, 4년제 대학 재학생 2천여 명 중 노트 필기를 '자주 한다'는 응답이 75%로 많았다. 또한 취업 경력 관리 포털 스카우트에 따르면 대학생과 직장인들의 38%가 정보를 저장하는 방식으로 메모를 선호한다고 밝혔다.

직장인과 학생의 경우 많은 양의 필기를 다시 옮겨 적어 문서화 하는 것이 번거롭고, 수기로 작성한 문서의 경우 보관이 불편함이 있다. 이에 이미지를 인식해 AI를 통한 문서 자동화 서비스가 필요하며 필기로 작성한 문서를 체계적으로 관리하는 기능을 제공하는 서비스가 필요하다.

그러나 국내외에서 제공하는 기존의 애플리케이션의 경우 패턴 분석 알고리즘과 통계 기반의 OCR을 혼합하여 사용하기 때문에 필기체에 대한 인식률이 낮다. 따라서 머신러닝 기반의 OCR을 활용하여 필기체에 대한 높은 인식률을 제공하는 서비스가 필요하다. 이에 본 논문에서는 위의 요구사항을 만족하는 서비스의 구성과 기능에 대해 설명한다.

본 논문의 구성은 다음과 같다. 2장에서는 제안하는 서비스를 구축하는데 필요한 기술을 분석한다. 3장에서는 제안하는 서비스의 주요 기능과 기능 구성도에 대하여 살펴보고, 마지막 4장에서 기대효과 및 활용분야에 대하여 제시한다.

2. 적용 기술 및 시스템 구성도

가. OpenCV

OpenCV(Open Source Computer Vision)는 실시간 컴퓨터 비전을 목적으로 한 프로그래밍 라이브러리이다. 인텔이 개발하여 실시간 이미지 프로세싱에 중점을 두었다. 인텔CPU에서 사용되는 경우 속도의 향상을 볼 수 있는 IPP(Intel Performance Primitives)를 지원한다. 이 라이브러리는 윈도우, 리눅스 등에서 사용 가능한 크로스 플랫폼이며 오픈소스 BSD 허가서 하에서 무료로 사용할 수 있다. OpenCV는 TensorFlow, Torch, PyTorch 및 Caffe의 딥러닝 프레임워크를 지원한다. OpenCV의 기계학습 알고리즘은 이미지 파일에서 문자를 추출하여 텍스트로 변환한다[1].

나. 딥러닝 분류 알고리즘

딥러닝에서 분류(Classification)은 지도 학습의 일종으로, 기존에 존재하는 데이터와 카테고리와의 관계를 학습하여 새로 관측된 데이터의 카테고리를 판별하는 문제이다. 분류는 데이터에 대한 가정을 먼저 해야하고, 최적의 가설을 찾는 과정을 거친다. 따라서 각각의 서로 다른 분류 알고리즘들은 서로 다른 인수를 가지고 있으며, 해당 인수를 가장 잘 만족하는 기능 매개 변수를 계산하는 과정이다. 제안하는 서비스에서는 Google ML Engine을 이용하여 딥러닝 분류 알고리즘을 학습하고 학습된 모델을 Serverless 방식으로 구동한다[2].

다. Google Docs API

Google Docs API는 Google에서 제공하는 웹 기반 서비스로, 워드 프로세서, 스프레드시트, 프레젠테이션, 그림 등의 기능이 제공된다. 문서는 WYSIWYG 방식으로 문서 작성이 가능하며 지원하는 문서 포맷은 마이크로소프트 워드(.doc), 서식있는 텍스트(.rtf), 오픈도큐먼트(.odt), PDF(.pdf), 텍스트(.txt), HTML(.html)이다. 스프레드시트를 지원하는 포맷은 마이크로소프트 엑셀(.xls), CSV(.csv), 오픈도큐먼트(.odt), PDF(.pdf), 텍스트(.txt), HTML(.html)이다. 지원하는 크기는 1만행, 256열, 10만개 셀, 40시트 이내이며 동시에 열 수 있는 스프레드시트는 11개이다. 프레젠테이션은 ZIP 포맷으로 압축하여 다운로드가 가능하다. 지원하는 문서 포맷은 마이크로소프트 파워포인트(.ppt)이다. Google Docs API는 프로그래밍 방식으로 문서를 읽고 쓸 수 있어 프로세스를 자동화하고 문서를 대량으로 만들 수 있다[3].

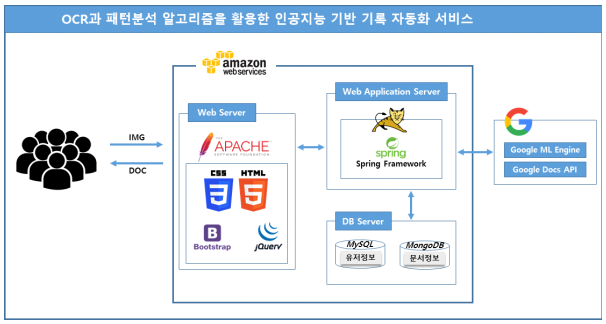
라. MongoDB

몽고DB는 크로스 플랫폼 도큐먼트 지향 데이터베이스 시스템이다. NoSQL 데이터베이스로 분류되는 몽고DB는 JSON과 같은 동적 스키마형 도큐먼트들을 선호함에 따라 전통적인 테이블 기반 관계형 데이터베이스 구조의 사용을 삼간다. 이렇게 함으로써 특정한 종류의 애플리케이션을 더 쉽고 더 빠르게 데이터 통합을 가능케 한다. 아페로 GPL과 아파치 라이선스를 결합하여 공개된 몽고DB는 프리 오픈 소스 소프트웨어이다[4].

마. Spring Framework

Spring Framework는 자바 플랫폼을 위한 오픈소스 애플리케이션 프레임워크로서 간단히 스프링이라고도 한다. 동적인 웹 사이트를 개발하기 위한 여러 가지 서비스를 제공하고 있다. 대한민국 공공기관의 웹 서비스 개발 시 사용을 권장하고 있는 전자정부 표준프레임워크의 기반 기술로서 쓰이고 있다. 제안하는 서비스에서는 Spring Framework는 사용자에게 Restful 웹 서비스를 제공한다[5].

(그림 1)은 제안하는 서비스에 적용되는 기술을 종합하여 나타낸 시스템 구성도를 나타낸다.



(그림 1) 제안하는 서비스의 시스템 구성도

3. 주요 기능

제안하는 서비스는 사용자가 업로드한 사진을 분석하여 텍스트를 추출해 문서 파일로 변환하고 클라우드 서버에 저장하고 자유롭게 관리하고 다운로드하는 기능을 제공한다. 사진에서 텍스트를 추출하는 기능은 오픈소스로 제공되는 OpenCV를 활용하였고, Google ML Engine을 활용하여 문서에 포함된 사진, 도형과 필기체를 추출한다. 추출된 사진, 도형, 필기체를 객체로 저장하여 문서파일로 변환하는 기능은 Google Docs API를 활용하여 Google Drive 클라우드 서버에 저장한다. 문서 정보를 NoSQL인 MongoDB에 저장하여 MySQL과 함께 유기적으로 관리한다.

<표 1>은 제안하는 서비스의 주요 기능을 요약적으로 보여준다.

<표 1> 제안하는 시스템의 주요 기능 요약

구분	기능	설명
S/W	머신러닝을 활용한 문자인식 및 텍스트 변환 기능	- 머신러닝 기반의 OCR을 이용하여 이미지 파일에서 문자를 인식하여 추출함 - 추출된 문자를 컴퓨터가 읽을 수 있는 텍스트로 변환하여 객체로 저장함
S/W	인공지능 기반의 패턴인식을 활용한 도형인식 기능	- 딥러닝 패턴인식을 활용하여 이미지 파일에서 도형 등의 문서삽입물을 인식하고 추출함 - 추출된 문서삽입물을 객체로 저장함
S/W	Google Docs API를 활용한 문서편집 기능	- 저장된 텍스트 객체와 문서삽입물 객체를 불러와 문서파일을 편집하고 저장함

4. 기대효과 및 활용분야

제안하는 서비스는 딥러닝 기반의 OCR을 활용해 한글의 인식률을 높이고 도형의 인식을 가능하게 한다. 기존 OCR 활용 애플리케이션의 단점을 보완하여 필기체의 인식률을 높이고 패턴 인식을 통한 도형 및 사진의 인식으로 OCR의 활용성이 증가할 것으로 기대한다. 또한 Google Docs API를 활용하여 클라우드에 문서를 저장함으로써 브라우저 별도의 문서 편집기를 사용하지 않고 브라우저 환경에서의 문서 편집이 가능하다. 온라인으로 저장된 문서 파일을 열고 편집하는 행위가 가능해져 문서에 대한 관리가 편리하다.

제안하는 서비스는 NoSQL을 활용하여 데이터를 통합적으로 관리해 문서파일을 더 쉽고 더 빠르게 저장하고 불러올 수 있다. 이로 인해 사진을 업로드 하는 것으로 쉽게 문서 파일이 생성되기 때문에 필기를 다시 정리할 때 소요되는 시간과 인력을 감축시킬 수 있으며 업무 및 학업의 효율성이 증가한다.

제안하는 서비스는 웹 브라우저 환경에서 문서를 사진으로 찍어 업로드하면 문서파일로 클라우드에 저장되기 때문에 시간과 장소에 구애 받지 않고 문서의 열람 및 편집이 가능하다. 또한 필기체, 도형과 사진 등의 문서 삽입물의 인식이 가능해 문서를 관리하는 데 편리함을 도모한다. 따라서 학생들에게 수업 필기 내용 정리에 간편함을 제공해 학업의 효율성을 증진 시킬 수 있으며 직장인들은 서류나 회의록 등의 번거로운 자료 정리 방식을 자동화할 수 있어 업무 부담을 덜 수 있다.

최근 활발하게 연구가 진행 중인 RPA 분야와 인공지능분야와 혼합하여 사용할 경우 제안하는 서비스의 사용도가 더욱 높아질 것으로 기대한다.

5. 참고문헌

[1] https://opencv.org
[2] Lee, Moohyun, “Obstacle Classification Method using Multi Feature Comparison Based on Single 2D LiDAR”, Journal of Institute of Control, Robotics and Systems Engineering, vol. 22, 2016.
[3] https://developers.google.com/docs/api
[4] H.J. Park, “A Study about Performance Evaluation of Various NoSQL Databases,” Journal of Korea Institute of Information, Electronics, and Communication Technology, Vol. 9, No. 3, pp. 298-305, 2016
[5] https://spring.io

본 논문은 과학기술정보통신부 정보통신창의인재양성사업의 지원을 통해 수행한 ICT멘토링 프로젝트 결과물입니다.