

캡슐내시경 동영상으로부터 학습 데이터 레이블링을 위한 정보 추출 기법¹⁾

장현웅*, 임창남*, 박예슬*, 이광재**, 이정원*,
*아주대학교 전자공학과, **아주대학교 의과대학 소화기내과
e-mail: jhw3169@gmail.com, chn0714@naver.com, yeseuly777@gmail.com,
kjl@ajou.ac.kr, jungwony@ajou.ac.kr

Information Extraction Method for Labeling Learning Data from the Capsule Endoscopic Video Images

Hyeon-Woong Jang*, Chang-Nam Lim*, Ye-Seul Park*, Kwang-Jae Lee**,
Jung-Won Lee*

Department of Electrical and Computer Engineering, Ajou University*
Department of Gastroenterology, Ajou University Hospital**

요 약

최근 딥러닝과 머신러닝 기법이 소프트웨어의 성능 향상에 도움이 되는 것이 입증됨에 따라, 의료 영상 진단 보조 소프트웨어를 개발하기 위한 시도가 활발해 지고 있다. 그 중 캡슐내시경은 소장 소화기관을 관찰할 수 있는 초소형 의료기기로, 기존의 내시경 검사와 다르게 이물감이 느껴지지 않고 의료보험 적용으로 최근 들어 널리 이용되고 있다. 일반적으로 캡슐 내시경은 8 시간 동안 소화기간을 촬영하며, 한 번의 검사 결과로 생성된 동영상 데이터 셋은 수 만장의 이미지를 포함하기 때문에, 방대한 양의 이미지들을 효율적으로 관리하기 위한 체계가 필요하다. 특히, 방대한 양의 캡슐내시경 이미지를 학습하는 경우, 수 만장의 이미지 속에서 유의미한 특징(촬영정보, 의사소견, 환자정보, 병변의 위치 및 크기 등)을 추출해내야 하므로 학습 데이터 레이블링을 위한 정보를 정확히 추출해야 하는 작업이 요구된다. 따라서 본 논문에서는 캡슐내시경 영상을 학습할 때, 학습 데이터 레이블 정보를 체계적으로 구축할 수 있게 하는 레이블 정보 추출 기법을 제안하고자 한다. 제안하는 기법은 병원에서 14년간 수집된 총 340명의 캡슐내시경 데이터(약 1,700 만장의 이미지)를 토대로 영상데이터를 구조적으로 분석하여 유의미한 정보를 추출하고 노이즈 데이터를 제거한 뒤, 빅데이터 저장소에 적재할 수 있음을 보였다.

1. 서론

최근 의료 분야에서는 인공지능 기법을 활용한 의료 영상기반의 진단 보조 소프트웨어가 많이 개발되고 있다. 그 중 딥러닝 및 머신러닝 알고리즘을 이용하여 의료 영상을 학습함으로써 의사의 진단을 보조하는 소프트웨어들이 두각을 보이고 있다. 대표적으로는 해외의 IBM Watson의 암진단 보조를 위한 ‘왓슨 포 온콜로지’와 국내에서는 폐결핵이나 유방암 영상 판독 프로그램인 루닛(Lunit)의 Insight, 그리고 뼈나미 예측, 폐와 안구 이상성을 탐지하는 뷰노(Vuno)의 소프트웨어들이 있다[1,2]. 이러한 소프트웨어들은 방대한 양의 의료 영상 데이터를 학습함으로써 진단의 편이성을 제공하고 진단의 정확도를 높이는 의사의 진단 보조 역할을 하고 있다. 의료 영상은 다양한 형태로 존재하는데, 일반적으로는 X-ray, CT, MRI가 알려져 있다. [3]의 연구에서, 의료 영상에 딥러닝을 적용하여 심장, 망막, 디지털 병리, 폐, 유방 등의 분야

에서 진단에 기여한 300여개의 응용을 소개하고 있다. 그러나 캡슐내시경(CE: Capsule Endoscopy) 영상을 이용한 딥러닝 적용에 대한 예는 찾아보기 어렵다. 캡슐내시경은 일반적으로 1초에 2장의 사진을 찍으며, 캡슐 형태의 장치를 삼키고 몸에 부착된 수신 장치를 통해 데이터를 저장한다. 일반적으로 한 번의 검사에 평균 8시간 정도 신체내부에 머무르며, 촬영으로 생성된 이미지의 수가 5만 여장이 넘게 된다. 의료 현장에서 한 환자의 5만여장의 영상을 관독하는 데에만 45분~2시간 가량이 소요될 정도로 집중을 요하는 작업이므로 병변을 탐지하거나 보조하는 소프트웨어 개발이 시급한 실정이다[4].

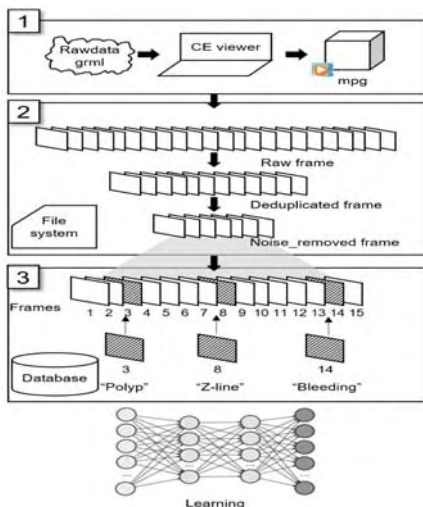
일반적으로 대부분의 의료 영상은 의료 영상 전용 소프트웨어를 통해서만 조회될 수 있다. 그러나 X-Ray, CT, MRI와 같은 영상들은 사진 혹은 3D 형태로 몇 장의 이미지 형태로 관독하므로 특별한 전 처리 과정 없이 딥러닝에 적합한 포맷으로 변형이 가능하다. 하지만 캡슐내시경의 경우에는 동영상 형태로 파일이 생성되기 때문에, 이를 학습에 적합한 이미지 프레임 형태로 변환하는 과정이 필요하며, 각각의 프레임 속에서 중복되거나 노이즈(촬영해상도가 낮거나 조도가 너무 높고 낮은 경우 등)가 심한

1) 본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 대학ICT연구센터육성지원사업의 연구결과로 수행되었음 (IITP-2019-2016-0-00309)

경우를 제거해야 하는 부수적인 작업이 따르게 되어 특별한 정제 과정이 필요하게 된다. 더욱이 5만 여장의 프레임 중에 소장관찰 영역만을 추출해야 하며, 추출한 프레임 셋에서도 전문가의 판독 소견(annotation)과 매핑해야 하는 복잡한 정제 과정이 필요하다. 그림 1은 이와 같은 데이터 정제 작업을 보여준다. 먼저, 원시 캡슐내시경 영상을 전용 뷰어를 이용해서 학습이 가능한 동영상 파일로 추출하는 작업을 진행한다(그림 1-1: 환자별 동영상 추출). 추출한 동영상 파일을 프레임 단위로 변환한 뒤 중복된 영상과 노이즈 영상을 제거하고 파일 시스템에 적재해야 한다(그림 1-2: 데이터 정제). 마지막으로, 추출한 프레임 중에서 의료 전문가의 소견(위장관 위치, 의심병변 등)이 담겨 있는 썸네일(Thumbnail) 파일들을 수집된 프레임의 일련번호와 매칭시키고, 유의미한 소견 정보를 바인딩 하는 과정을 거친다(그림 1-3: 추출 썸네일과 판독 정보 매핑).

데이터 정제를 위한 연구는 많은 분야에서 이루어지고 있다. 최근, [5]에서는 의료 분야에서 활용되고 있는 빅데이터에 인공지능 기술을 접목할 때, 데이터 정제의 중요성을 강조하고 있으며, [6]에서는 실제 데이터 정제 작업을 기반으로 한 데이터 웨어하우스 구축 사례를 소개하고 있다. 활발하게 이루어지고 있지는 않지만, 최근 [7]과 같이 의료 영상(예: CT)을 학습하기 위한 데이터 정제 방법이 소개되고 있는 추세이다. 그러나 현재 캡슐내시경 영상의 경우, 이와 같은 데이터 정제 과정이 이루어지고 있지 않아 개발자가 수동으로 분석해야 하는 상황이다. 특히, 검사 환자의 판독을 위해 45분~2시간 가량 소요되는 의료 현장의 문제점을 해결하기 위해 캡슐내시경 판독 자동화를 위한 진단 보조 소프트웨어 개발이 시급한 실정이다.

따라서 본 논문에서는 캡슐내시경의 학습을 효과적으로 준비하기 위한 학습 데이터 레이블링을 위한 정보 추출 기법을 제안한다. 제안하는 기법은 학습에 필요한 영상 정보를 체계적으로 추출하고 데이터를 정제함으로써, 딥러닝에 적용할 수 있는 전처리 과정을 완성할 수 있다.



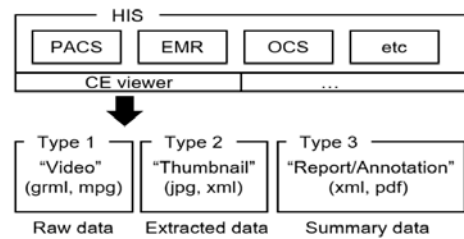
(그림 1) 학습을 위한 캡슐내시경 영상 정제 과정

2. 캡슐내시경 영상 데이터의 구조 분석

2.1 CE 데이터 타입

본 절에서는 캡슐내시경 영상 데이터의 생성 구조에 대해 소개한다. 그림 2와 같이 캡슐내시경 영상은 병원에 구축되어 있는 병원정보시스템(HIS, Hospital Information System) 내의 의료영상저장전송시스템(PACS, Picture Archiving and Communicating System) 및 전자의무기록(EMR, Electronic Medical Record)을 기반으로 획득된다. 획득되는 데이터의 형태는 다음과 같다.

- **Type 1(Raw data):** 영상 전체의 데이터를 가지고 있는 파일이다. 캡슐내시경 검사 후에 수신 장치에 저장된 파일을 Given Imaging사의 Rapid 판독 소프트웨어로 업로드 하면 원시파일인 grml 확장자를 갖는 원시데이터가 생성되며 mpg는 grml 파일에서 동영상 파일로 추출된다.
- **Type 2(Extracted data):** 캡슐내시경 영상 전용 뷰어를 이용하면 썸네일 파일을 추출할 수 있다. 썸네일 파일은 전문가가 의미 있는 프레임들(의사소견, 위장관 위치 구분 등)을 체크한 영상에 대해서만 추출하게 된다.
- **Type 3(Summary data):** 영상에 대한 전체적인 정보가 담겨있는 리포트 파일을 추출할 수 있다. xml이나 pdf의 형태로 추출할 수 있다.



(그림 2) 캡슐내시경 영상 데이터 구조

2.2 CE 데이터의 구조적 분석

본 장에서는 2.1절에서 소개한 세 가지 데이터 타입을 구조적으로 분석한 결과를 제시한다. 데이터 타입 분석 기준은 전 세계 CE 시장의 70% 이상의 점유율을 가지는 Medtronic사의 CE에서 생성된 데이터 파일을 중심으로 분석하였다.

1) 캡슐내시경 동영상

- **캡슐내시경 뷰어 조회용 파일:** 해당 영상의 포맷은 grml 파일이며 캡슐내시경 검사를 한 뒤에 생성되는 데이터이다. 일반적인 동영상 파일과는 다른 Given Imaging사의 전용 파일 형식이므로 학습에 이용할 수 없으나 해당 파일에서 다양한 정보를 추출할 수 있다.
- **학습용 동영상 파일:** 캡슐내시경 뷰어를 통해서 이용자가 설정한 구간의 추출한 동영상 파일이다. 일반적인 mpg 파일의 포맷을 가지기 때문에 해당 영상을 프레임으로 변환시켜 학습에 이용할 수 있다.

2) 캡슐내시경 썸네일

- 캡슐내시경 추출 영상: 전문가가 캡슐내시경 뷰어를 보며 판독한 영상이다. 병변이 의심되는 부위나 의미가 있는 영상들을 체크하고 파일을 추출하면 해당 영상의 파일이 jpg 형태로 나온다.
- 복강내 위치 조회 영상: 해당 영상은 소프트웨어에서 지원하는 기능으로 캡슐내시경 추출 영상을 jpg로 추출할 때, 복강내 위치 조회 영상이 같이 나온다.
- 썸네일 요약 리포트(xml): 썸네일을 추출하는 과정에서 같이 추출되는 리포트 파일이다. 캡슐내시경 검사의 전체적인 정보가 들어있으며, 썸네일에 남긴 의사의 소견이 들어있다.

3) 캡슐내시경 레포트

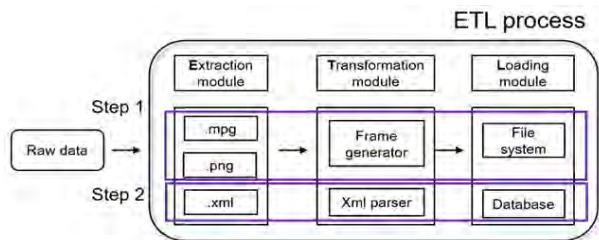
- 해당 파일에는 두 개의 영상과 코멘트가 달린 간단한 레포트가 나온다. 파일의 포맷은 pdf 파일이다.

3. 학습용 CE 데이터 레이블링을 위한 정보 추출 기법

3.1 학습용 CE 데이터 저장소 설계

본 절에서는 2장의 분석 결과를 기반으로 학습용 데이터 저장소를 제안한다. 저장소를 구축하기 위해서는, 그림 3과 같은 ETL(Extract, Transform, Load) 과정이 필요하게 되며, 본 논문에서는 ETL 과정을 파일 시스템으로 구축하는 단계(Step 1)와 데이터베이스를 구축하는 단계(Step 2)로 구분하여 제안한다.

- Step 1 (파일시스템 구축): 동영상 파일들을 프레임 단위로 변환하여 중복 프레임, 노이즈 영상들까지 제거해 파일타입으로 학습용 데이터를 저장한다.
- Step 2 (데이터베이스 구축): xml의 내용을 추출하여 데이터베이스 각 테이블의 요소를 채운다. 데이터베이스는 총 6개의 테이블을 가지고 있으며, 주요 테이블에 대한 설명은 표 1, 2와 같다.



(그림 3) 캡슐내시경 데이터 정제 프로세스

<표 1> thumbnail 테이블 구조

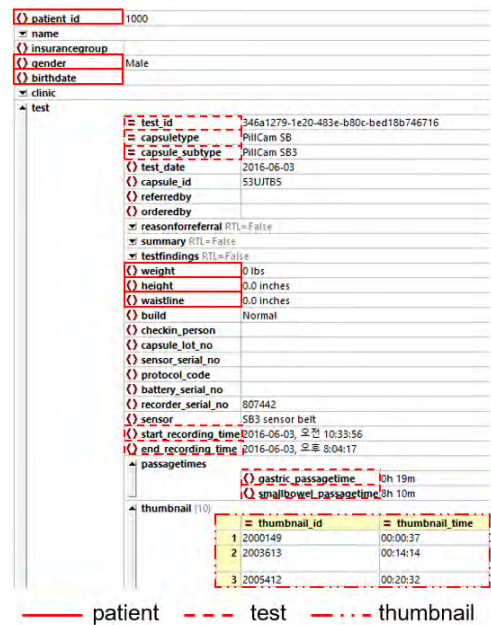
필드명	설명
thumbnailID	썸네일의 고유 ID (Primary Key)
frameID	해당 썸네일과 일치하는 frame의 ID이다. (Foreign Key)
thumb_*	썸네일의 다양한 정보가 들어있다.
comment	전문가가 해당 영상에 코멘트한 내용이다. 레이블 매칭에 쓰인다.

<표 2> frame 테이블 구조

필드명	설명
frameID	mpg 파일에서 frame generator을 통해 추출한 frame의 ID (Primary Key)
mpgID	mpg 테이블에서 외래키로 가져온 값. 어떤 영상인지 알 수 있다.(Foreign Key)
label	무슨 증상을 가지고 있는지 전문가가 코멘트한 정보를 가지고 있다.

3.2 학습용 CE 데이터 레이블링 정보 추출

본 절에서는 앞서 소개한 정보들을 구조적으로 분석하여 학습 레이블 정보를 추출한다. 캡슐내시경 판독시 생성된 모든 의료 정보는 XML 형태로 저장되어 있다. 그러나 현재는 이와 같은 정보들이 세부적으로 분석되지 않아 학습을 위한 데이터 레이블 정보를 개발자가 직접 분석해야 하는 어려움이 따른다. 따라서 본 연구에서는 이와 같은 레코드에서 학습에 필요한 주요 정보를 그림 4와 같은 xml 데이터를 3 가지 그룹으로 분류하여 추출하였다.



(그림 4) 판독 정보를 포함한 XML 파일

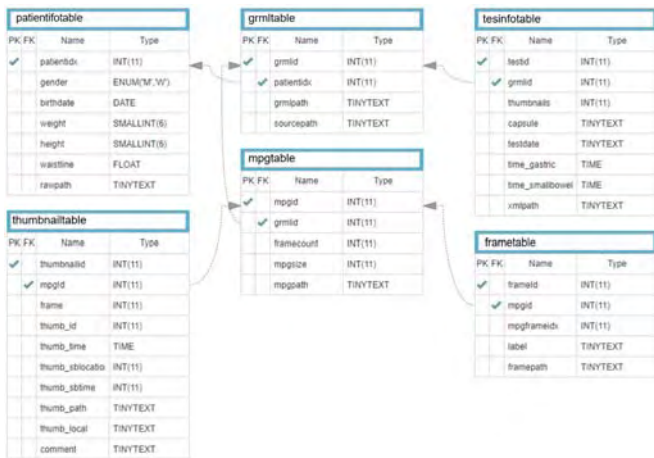
먼저, 환자를 식별하기 위한 정보를 추출한다(예: 환자 ID, 환자 몸무게, 환자 성별, 환자 출생일 등). 그러나 환자의 개인정보를 보호하기 위해, 이와 같은 정보들은 익명화되어 수집되며 수집된 데이터를 구별할 수 있는 수준의 환자 정보(예: 초기화된 환자 ID)만이 분석된다. 다음은 캡슐내시경 검사 정보에 관한 내용으로 검사 ID, 검사 날짜 등이 해당된다. 마지막으로 학습을 위해 필수적으로 분석되어야 하는 썸네일에 대한 정보이다. 이 부분은 의사가 직접 소견(Annotation) 혹은 코멘트(Comment)를 달아 놓은 유의미한 데이터에 대한 정보이다. 구체적으로는 썸네일 ID, 썸네일 시간과 썸네일과 바인딩 되어야 하는 코멘트 등의 정보가 존재한다. 이와 같이 분류된 태그들은 각

정보에 해당하는 데이터베이스의 테이블로 매칭되며, 매칭된 데이터베이스의 일부를 예시하면 그림 5와 같다. 또한 구축된 전체 데이터베이스에 대한 스키마 구조는 그림 6과 같다. 이와 같은 데이터베이스는 최종적으로 학습을 위한 기반으로 활용될 수 있다.

patientidx	gender	birthdate	weight	height	waistline
1	W	1953-04-30	55	160	28
2	M	1930-07-04	65	175	30
3	M	1970-12-12	78	183	32
4	W	1940-11-11	55	155	29
5	M	1948-10-01	65	166	30
6	W	1953-06-17	60	158	31
7	W	1967-05-30	70	171	32

frameid	mpgid	mpgframeidx	label	framepath
1	1	1	first image	(NULL)
2	1	2	(NULL)	(NULL)
3	1	3	Normal gastric mucosa	(NULL)
4	1	4	Normal duodenal mucosa	(NULL)
5	1	5	Angiodysplasia 의심	(NULL)
6	1	6	(NULL)	(NULL)

(그림 5) 구현된 patientinfo, frame 테이블



(그림 6) 데이터베이스 관계도

3.3 구축 결과: 캡슐내시경 영상을 위한 빅데이터 저장소

수신 장치에 저장된 grml파일로부터 프레임단위의 이미지를 추출하고, 픽셀별 RGB 값을 비교하여, 중복 데이터를 정제하였다. 이 경우 5장의 프레임이 규칙적으로 중복되는 것을 확인했으며, 해당 프레임을 제거하면 5장중 1장의 프레임만 남게 된다.

<표 3> 데이터 저장소 구축 결과

데이터 수집 기간	2005년~2019년
총 환자수	396명
유효 검사수	340건 (56건의 검사 결과는 검사실패 혹은 업로드 실패인 경우)
검사당 생성된 이미지수	건당 5만장~19만장
총 이미지 수	340건 * 평균12만장 = 약 4,080 만장
정제된 이미지 수	4,080만장 * 20% = 약 816 만장

총 396명의 환자가 있었으며, 그 중 검사실패 혹은 업로드 할 수 없는 검사를 제외하면 340명의 유효한 환자 데이터를 확보했다. 데이터 정제 과정을 통해 학습이 가능한 약 816만장의 이미지를 확보하였다.

4. 결론 및 향후 연구

본 논문에서는 효과적인 영상 학습을 수행하기 위한 데이터 전처리 방법으로서, 레이블 정보를 추출하기 위한 방법을 제안하였다. 제안하는 방식은 캡슐내시경 영상 데이터 타입과 구조에 대한 분석을 통해, 레이블링에 활용될 수 있는 정보를 분류하고 이를 추출하는 방식이다. 이와 같은 방식은 하나의 데이터 셋이 방대한 프레임 형태로 존재하는 캡슐내시경 영상 데이터에 있어, 유의미한 데이터를 선별해내는 작업을 수행할 수 있으며 더 나아가 다양한 목적으로 캡슐내시경 영상 학습을 수행할 수 있게 한다. 향후에는 구축된 학습용 데이터 저장소를 기반으로 위장관 교차점 인식을 통한 캡슐내시경의 소장내 위치 추적을 위한 딥러닝을 진행할 계획이다.

참고문헌

- [1] Eui Jin Hwang, et al. "Development and Validation of a Deep Learning-Based Automatic Detection Algorithm for Active Pulmonary Tuberculosis on Chest Radiographs", *Clinical Infectious Diseases*, doi: 10.1093/cid/ciy967, 2018.11
- [2] Kim Jeong Rye, et al. "Computerized bone age estimation using deep learning based program: evaluation of the accuracy and efficiency." *American Journal of Roentgenology*, Vol. 209, No. 6, pp. 1374-1380, 2017
- [3] Litjens, Geert, et al. "A survey on deep learning in medical image analysis." *Medical image analysis*, Vol. 42, pp. 60-88, 2017
- [4] 임윤정, 문정섭, 장동경, 장병익, 전훈재, 최명규, "대한 소화기내시경학회: 캡슐내시경의 가이드라인", *대한소화기내시경학회지*, Vol. 37, pp. 393-402, 2008
- [5] Benke, Kurt, and Geza Benke. "Artificial Intelligence and Big Data in Public Health." *International journal of environmental research and public health*, Vol. 15, No. 12 pp. 2796, 2018
- [6] Sebaa, Abderrazak, et al. "Medical big data warehouse: architecture and system design, a case study: improving healthcare resources distribution." *Journal of medical systems*, Vol. 42, No. 4, pp. 59, 2018
- [7] Tachibana, Rie, et al. "Electronic cleansing in CT colonography using a generative adversarial network.", *Medical Imaging 2019: Imaging Informatics for Healthcare, Research, and Applications*, Vol. 10954, doi: 10.1117/12.2512466, 2019