

# 국민건강정보 기반 진료과목 예측

안신영\*<sup>1</sup>, 이유경\*<sup>2</sup>, 박명호\*\*<sup>3</sup>, 변정용\*\*<sup>4</sup>

\*동국대학교 경주캠퍼스 컴퓨터과학과

\*\*동국대학교 경주캠퍼스 컴퓨터공학과

e-mail : {<sup>1</sup>anm3638, <sup>2</sup>leey, <sup>4</sup>byunjy}@dongguk.ac.kr, <sup>3</sup>myunghopark@gmail.com

## National Health Data based Medical Treatment Prediction

Shinyoung Ahn\*, Yookyung Lee\*, Minghao Piao\*\*, Jeongyong Byun\*\*

\*Dept. of Computer Science, Dongguk University at Gyeongju

\*\*Dept. of Computer Engineering, Dongguk University at Gyeongju

### 요 약

생활수준의 향상 및 소비자들의 건강에 대한 관심의 증가로 인해 자신의 건강에 대해서 스스로 결정하고자 하는 요구가 점차 증가하고 있다. IT 와 의료기술의 발달은 이를 가능하게 하였으며 각종 의료정보를 기반으로 하는 질병진단에 대한 연구가 많이 진행되고 있다. 본 논문에서는 국민건강정보 기반 진료과목 예측에 대한 연구를 진행하여 소비자 스스로 진료과목을 선택하는데 도움을 주고자 한다.

### 1. 서론

최근 빅데이터가 여러 분야에서 화제의 중심이 되면서 보건의료 분야에서도 이에 대한 관심이 높아지고 있다. 인간의 평균 기대수명이 점차 높아지고 건강관리와 질병 예방에 대한 수요가 증가함에 따라 보건의료 분야에서의 빅데이터 활용은 더 이상 선택이 아닌 필수가 되고 있다. 특히 소비자를 중심으로 하는 소비자 맞춤형 및 주도형 서비스에 대한 연구가 많이 진행되고 있다.

맞춤형 서비스는 개인의 의료정보를 바탕으로 건강에 대한 정보를 제공하고 개인별 맞춤 건강관리 지침을 제공받는 일체 서비스를 말한다. 이러한 서비스를 목적으로 비만, 대사증후군, 뇌졸중, 심혈관 질환, 암 예측 등에 대한 많은 선행연구들이 수행되고 있다[1]. 하지만 국내에서는 한국인을 위한 데이터의 부재로 다양한 시도가 진행되지 못하고 있는 상황이다.

국민건강정보데이터[2]의 개방은 다양한 시도를 할 수 있는 기반이 되고있다. 따라서 본 논문에서는 국민건강정보를 기반으로 하는 소비자 스스로 진료과목을 선택할 수 있도록 도움을 주는 예측 모델을 만들어 보이고자 한다.

### 2. 국민건강정보데이터 기반 진료과목 예측 모델링

본 연구에서는 국민건강보험공단 빅데이터 운영실에서 정부 3.0 정책 기조에 맞추어 개방한 공공개방데이터[3]를 사용하여 데이터 마이닝 기반의 진료과목 예측 모델링을 진행하였다.

국민건강정보데이터는 2002년부터 2013년 기간에 해당하는 국민건강보험가입자 100만명의 진료내

역정보, 의약품처방정보 및 건강검진정보로 이루어져 있다. <표 1>는 데이터의 구성을 보여준다.

<표 1> 국민건강정보데이터 구성

	수진자 기본정보: 성, 연령, 거주지 시도코드
진료내역정보 (가입자일련번호, 진료내역 일련번호 필수)	진료상세정보: 주상병, 부상병, 요양일수, 입·내원일수, 총 처방일수 요양급여 청구 심사 결과: 요양급여비용, 심사청구 결과 결정된 보험자와 수진자가 각각 부담해야 할 금액 등
의약품처방정보 (가입자 일련번호, 의약품처방내역 일련번호 필수)	수진자 기본정보: 성, 연령, 거주지 시도코드 의약품 처방 상세정보: 처방일련번호, 요양개시일자, 의약품 일반성분명코드, 1회·1일 투약량, 총 투약일수 단가, 금액 등
건강검진정보 (가입자일련번호 필수)	수진자 기본정보: 성, 연령, 거주지 시도코드 건강검진결과 및 문진정보: 신체, 몸무게, 허리둘레 등 신체사이즈 정보, 혈압, 혈당, 콜레스테롤, 요단백, 감마지피티, 시력, 청력, 구강검사, 음주, 흡연 등

(그림 1)는 국민건강정보 기반 진료과목 예측 프레임워크를 보여준다.

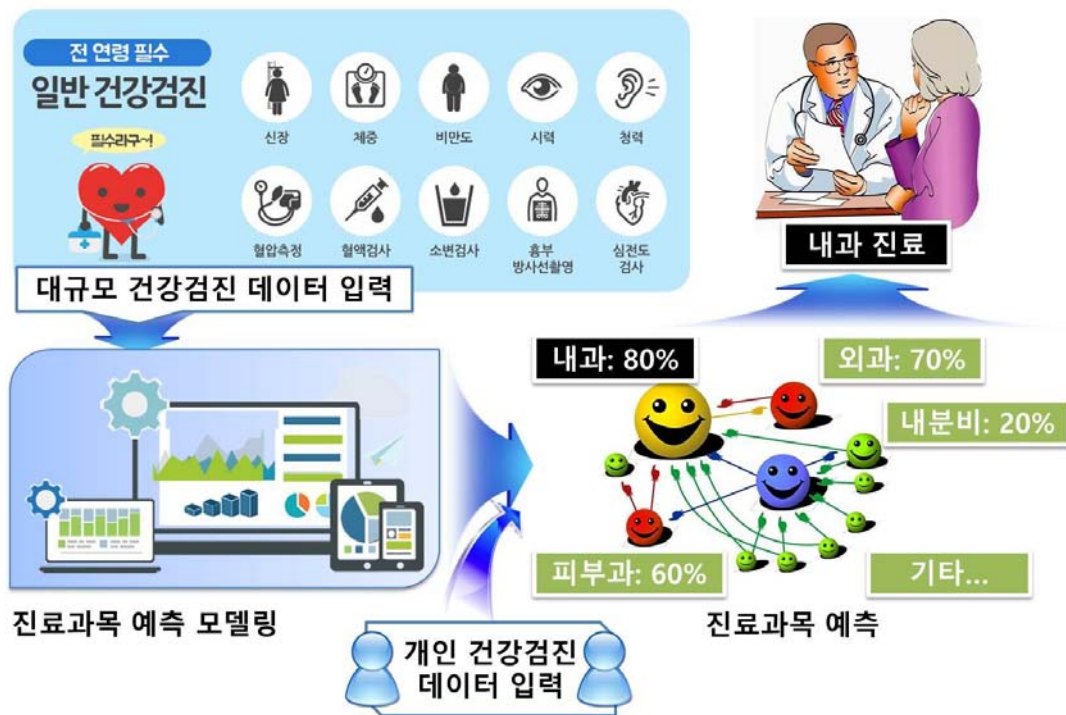
- I. 진료과목 예측 모델링을 위한 훈련 데이터로 2013년 건강검진정보 및 진료내역정보를 가입자일련번호로 통합을 진행하였다. 결측치가 포함된 데이터는 전체 레코드를 삭제하였으며, 가입자 별 연간 최대 방문한 진료과목을 훈련 및 예측 대상으로 정하였다.
- II. 훈련 알고리즘으로는 SVM(Support Vector Machine)을 사용하였으며 [4] 다항식 커널(Polynomial kernel)을 사용하여 훈련데이터를 기반으로 예측 모델을 생성하였다.

최종 602 개의 훈련데이터를 추출하였다. 해당 진료과목은 모두 18 개로 내과, 소아청소년과, 정형외과, 산부인과, 이비인후과가 전체 훈련 데이터의 80%로서 해당 진료과목은 모두 95% 이상의 정확도를 보여주었으며, 전체 평균 95.51%의 정확도를 보여주었다. 다만 훈련 데이터가 적은 일부 진료과목은 전혀 예측을 하지 못하였다.

단일 연도의 훈련 데이터를 이용한 테스트로 진료과목별 데이터 분포가 일정하지 못하여 일부 진료과목에 대해서는 예측이 전혀 진행되지 못하였다. 따라서 향후 전체 데이터에 대한 통합 및 정제를 통하여 전체 진료과목에 적합한 모델링을 진행할 예정이다.

### 3. 실험결과 및 향후 연구

건강검진정보 및 진료내역정보를 통합 및 정제하여



(그림 1) 국민건강정보 기반 진료과목 예측 프레임워크

#### 참고문헌

- [1] 조경희, 박영민, 지수혜, 추정은, 임현선. “개인별 맞춤형 통합 건강관리 프로그램 연구 및 개발,” 2014.
- [2] 정부 3.0 국민체험마당, 마이헬스뱅크 서비스, [http://news.jtbc.joins.com/article/ArticlePrint.aspx?news\\_id=NB11277183](http://news.jtbc.joins.com/article/ArticlePrint.aspx?news_id=NB11277183)
- [3] 국민건강보험공단 빅데이터 운영실. “국민건강정보데이터 사용자 매뉴얼[ver 1.0],” 2016
- [4] Jiawei Han, Micheline Karnber, Jian Pei “데이터마이닝 개념과 기법”, 2nd Ed. 에이콘출판사