

연관분석에 기반한 건강검진결과 시각화 방법

김준우[○], 박상찬^{*}

[○]동아대학교 산업경영공학과

^{*}경희대학교 경영대학원

e-mail:kjunwoo@dau.ac.kr[○], sangchanpark@daum.net^{*}

Association Mining based Visualization Method for Health Examination Results

Jun-Woo Kim[○], Sang-Chan Park^{*}

[○]Dept. of Industrial & Management Systems Engineering, Dong-a University

^{*}The Graduate School of Business Administration, Kyung Hee University

● 요약 ●

병원에서 다양한 정보시스템을 도입하면서 환자들과 관련된 방대한 의료 데이터들이 전자적인 형태로 축적되어 왔고, 최근에는 의료진이나 환자에게 적절한 정보를 제공하는데 이러한 데이터를 활용하고자 하는 노력이 이어지고 있다. 그러나 의료 데이터는 분량이 방대하고 전문적인 내용을 다루기 때문에 이에 기반한 정보를 개인 환자에게 제공하는데 있어서는 데이터에 포함된 내용을 사용자의 이해가 편리한 형태로 가공하는 것이 중요하다. 이에 본 논문에서는 연관분석과 관련된 행렬 기반 표현 방법을 기반으로 한 하이브리드 시각화 방법을 개발하고, 이를 건강검진 결과에 적용하는 것을 제안하고자 한다.

키워드: 건강검진(health examination), 시각화(visualization), 연관분석(association mining), 데이터마이닝(data mining), 계층형군집(hierarchical clustering)

I. 서론

최근 대부분의 병원에서 다양한 정보시스템을 도입하면서 환자들의 건강 정보나 의무 기록이 전자적인 데이터로 축적되고 있다. 이러한 데이터는 EMR(electronic medical record)과 같이 병원 내부의 업무 효율성 제고를 위해 사용될 수도 있고, 최근 등장한 PHR(personal health record) 개념과 같이 개인의 편의와 건강 관리를 위해 환자에게 제공될 수도 있다[1].

그러나 의료 관련 기록은 오랜 시간 동안 방대한 분량이 축적되고, 전문적인 영역을 다루기 때문에 이를 개인 환자에게 제공하는데 있어서는 데이터에 포함된 내용을 사용자의 이해가 편리한 형태로 가공하는 것이 중요하다. 이에 본 논문에서는 특히 대중들이 많이 경험하는 의료기록인 건강검진 결과에 시각화 방법을 적용하는 방안을 제안하고, 향후 활용 방안에 대해 모색해보고자 한다.

남기게 되고, 검진결과는 표준화된 양식으로 수검자에게 전달된다. 그러나 검진결과는 발견된 증상 위주의 비교적 단순한 형태로 작성되며, 앞으로 주의해야할 사항들에 대한 개인별 피드백은 상대적으로 미흡하다. 따라서 사용자가 이해하기 쉬운 형태로 이러한 부분을 제공하는 자동화된 방법이 요구된다.

건강검진 결과 항목의 상당수는 특정 증상이나 질환의 유무를 의미하는 단순한 이진(binary) 변수 형태이나, 항목의 개수가 많을 경우 그 내용을 한 눈에 파악하기 어렵다. 한편, 최근 이렇게 방대한 양의 데이터를 다루기 위한 방법으로 인간의 시각적 인지 능력을 활용하고자 하는 시각화(visualization) 기법의 활용이 많아지고 있고, 특히 이진 변수로 구성된 거래 데이터(transaction data)에 대한 연관규칙 탐사와 관련된 연구가 활발하다[2][3]. 이에 본 논문에서는 연관규칙 시각화 기법을 건강검진 결과 표현에 사용하는 방법을 제안하고자 한다.

II. 연구 배경

건강검진은 질병의 조기발견과 예방을 목적으로 하며, 직장이나 학교에서의 단체검진 또는 생애주기검진 등의 형태로 대중들이 흔히 경험하게 되는 의료 서비스이다. 아울러, 건강검진을 받는 수검자는 여러 영역을 순회하면서 측정, 검사, 설문 등의 방법으로 기록을

III. 건강검진결과 시각화 방법

그림 1은 제안하는 시각화 방법을 보여준다. 행렬의 행과 열은 각각 이진 변수를 의미하며 각 셀은 행 변수를 전향, 열 변수를 후향으로 하는 연관 규칙의 평가지표(신뢰도, Lift 등)에 따라 색깔

을 부여한다. 이는 연관규칙 시각화를 위해 사용되는 기존의 행렬 기반 표현과 동일하나, 본 논문에서는 건강검진결과에 적용하기 위하여 몇 가지 변형을 가하였다.

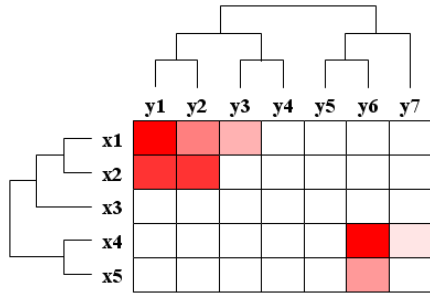


그림 1. 행렬 기반 하이브리드 시각화 방법
fig. 1. Matrix-based Hybrid Visualization Method

먼저, 행 변수들은 현재 수검자에게서 발견된 항목들을 배치하고, 열 변수에는 미발견 항목들을 배치하여 향후 주의해야할 항목들을 시각적으로 표현하고자 하였고, 둘째로는 행과 열 변수들에 각각 계층군집을 적용하여 얻은 계통도(dendrogram)를 함께 표시하였다. 계층군집은 여러 항목들 간의 전체적인 상관관계를 한 눈에 보여주는 방법으로 그 결과는 그림 1에서의 같은 계통도로 정리할 수 있다[4]. 따라서 이러한 정보는 당장 발견된 항목과 직접적인 관련이 적더라도 다른 항목을 통해 간접적인 관련이 있는 항목들을 보여줄 수 있다. 이외에도 계통도는 행과 열 변수들을 적절히 정렬하는 목적으로도 사용되어 행렬의 셀 중 높은 지표를 갖는 것들이 비교적 인접한 곳에 배치되도록 하는 효과를 갖는다.

IV. 결 론

데이터베이스 등의 도입으로 인해 기술적인 측면에서 의료기록의 저장과 접근이 용이해졌으나, 이를 사용자가 편리하게 이용할 수 있도록 표현하는 방법에 대한 연구가 추가적으로 필요한 실정이다.

본 논문은 특히 대중들이 흔히 접하는 건강검진결과에 대한 시각화 방법을 제안하였으며, 추후 실제 건강검진결과에 이를 적용하고 변수들의 정렬이나 각 셀의 지표 산출을 위한 세부적인 방법들을 보완해나갈 예정이다. 아울러, 본 논문에서는 건강검진결과와 같은 비교적 정형화된 데이터만을 가정하였으나, 앞으로는 EMR이나 PHR 등 보다 확장된 데이터를 이용하여 의료진이나 일반 개인들에 대해서도 이러한 시각화를 통해 유용한 정보를 제공하기 위한 방법도 필요할 것으로 생각된다.

감사의 글

이 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임 (NRF-2012R1A1A1044834)

참고문헌

- [1] S.-Y. Shin, and C.K. Chung, "A Future of Medical Information System: Establishment of Hospital-oriented Personal Health Record," Journal of Korean Medical Association, Vol.52, No.11, pp.1115-1121, 2009.
- [2] M.C.F. de Oliveira, and H. Levkowitz, "From Visual Fata Exploration to Visual Data Mining: A Survey," IEEE Transactions on Visualization and Computer Graphics, Vol.9, No.3, pp.378-394, 2003.
- [3] J.W. Kim, and H.K. Kang, "Visual Exploration based Approach for Extracting the Interesting Association Rules," Journal of the Korea Society of Computer and Information, Vol.18, No.9, pp.177-187, 2013.
- [4] P.-N. Tan, M. Steinbach, and V. Kumar, "Introduction to Data Mining," Addison-Wesley, 2005.