

氏名（本籍）	Alwis Nazir（インドネシア）
学位の種類	博士（医科学）
学位授与番号	甲第 27 号
学位授与日付	平成 26 年 9 月 30 日
専攻	医療情報学専攻
学位論文題目	医療におけるデータマイニングの応用 (Application of Data Mining in Healthcare)
学位論文審査委員	（主査）教授 丹 羽 雅 之 （副査）教授 赤 尾 幸 博 （副査）准教授 一 宮 尚 志 （副査）教授 紀 ノ 定 保 臣

### 論文内容の要旨

Data mining is the process to mine a series of valuable information from a database. These information are obtained by extracting and recognizing the patterns from data contained in the database. Data mining is very useful to get information from huge datasets. It seems natural that we can get more good information as the size of dataset increases. However, large dataset often include strong noise, unexpected errors, and complex structure, which make the analysis difficult. Data mining has been developed to overcome these difficulties.

One place that data mining will be useful is healthcare. The data such as the reports of adverse effects caused by drugs, and records of annual health checkups are very large, but very rarely processed to obtain useful information. The adverse event report, if we processed properly, will give us very useful information such as the relationship between an adverse events with a symptom of a disease. Likewise, the annual health checkups of data, when processed properly, will be very helpful for the prediction of a person's health in the future. In my research, I focused on two data, the adverse event reports and annual health checkup data, and applied two methods of data mining, the network analysis and hidden Markov models.

The aim of the first research was to identify the symptoms that would suggest a high suicide risk of depression. To achieve this task, we applied the network analysis of the data obtained from the US Food and Drug Administration Adverse Event Reporting System (FAERS) of selective-serotonin reuptake inhibitors. Using FAERS reports from 1997 to the second quarter of 2012, we constructed the co-occurrence network of adverse events. From this network, we extracted the events that were strongly connected to suicidal events (suicidal attempts, suicidal ideation, suicidal behavior, and complete suicide) by means of the community detection method. Using this method, we succeeded in obtaining a list of suicide-related adverse events. Owing to the randomness inherent in the algorithms of community detection, we found that the obtained list differed according to each trial of analysis. However, the lists we derived show considerable efficiency in identifying suicidal events. The

network analysis appears to be a promising method for identifying signals of suicide.

The aim of our second research is to find out the probability of change in the health risks based on annual time series data of health checkup and to determine the level of risk and the progression in health conditions especially for persons with hypertension. For this purpose, we made hidden Markov model analysis of the health checkup data between 2002 and 2007 which include 912,765 records from 279,904 participants, provided by the medical center in Gifu prefecture, Japan. From this dataset, we extracted the data of people with hypertension, i.e. systolic blood pressure (SBP) above 140 mmHg or diastolic blood pressure (DBP) values above 90 mmHg. For the person with hypertension who have a 4-6 year time series of data, we carried out the hidden Markov model analysis using the following test values: the body mass index (BMI), systolic blood pressure (SBP), diastolic blood pressure (DBP), hematocrit (Ht), platelets (Pla), glutamic oxaloacetic transaminase (GOT), glutamic pyruvic transaminase (GPT), total cholesterol (T.Chol), neutral fat (N) and blood sugar (BS). Because the health condition is strongly dependent on age, we divide the data in several age groups, namely 30's, 40's, 50's and 60's, and carried out the analysis for each group. We evaluated the obtained Markov model by comparison with the classification by professional medical staff. We succeeded to cluster the data in 6 groups. Almost in all states in each age group, the average value of BMI, T.Chol and N are out of the normal range. In age group 30's, the model has 4 different levels of risk. In age group 40's, 50's and 60's, the model has 3, 2, and 2 different levels of risk, respectively. Taking the transition probability into account, we found the risk in the future may differs even if the current risk is same. Using hidden Markov model we succeeded to find out the probability of change in the health risks based on annual time series data of health checkup.

From these results, we conclude that data mining methods such as the network analysis and hidden Markov model is useful to process the data contained in the healthcare centers and to derive the valuable information.

### 論文審査結果の要旨

我が国では少子化、高齢化、国民医療費の高騰といった社会的な課題に対する解決が喫緊の課題となっている。また、政府はデータヘルス計画を推進し、検診データとレセプトデータの突合を通して医療費の削減に効果的な方策を探るなど、健康医療分野での効果的なデータ活用に対する関心が高まっている。

申請者 **Alwis Nazir** は、健康医療分野の大規模データを対象に、広義のデータマイニング手法、狭義のネットワーク分析手法とHMM(隠れマルコフ手法)を用いて社会的な課題を解決するために効果的であると思われる手法を提案し、その有効性を明らかにした。

1) FAERS (FDA Adverse Event Reporting System) が有する抗うつ薬(SSRI)服用患者の有害事象(自殺企図)に関係すると思われる症状群(Agitation、Fatigue、Crying等)をネットワーク分析手法の一種であるコミュニティ検出法を用いて描出した。また、これら症状群の個数が増加するほど有害事象リスクが上昇することを明らかにした。

2) 大規模時系列検診データ(20代、30代、40代、50代、60代の被験者)を対象に、HMMを用いて加齢とと

もに健康リスクがどのような状態を遷移するかを検討した。その結果、健康状態は6つのグループに分類できること、状態の遷移確率がもっとも高いのは自己ループ(同じ状態が続くこと)であることを明らかにした。また、HMMは事前の仮定を必要とせず、健康状態の遷移を可視化する効果的な手法であることも明らかにした。

以上の結果より、申請者 **Alwis Nazir** は学術的には健康医療分野の大規模データを効果的に活用する手法を構築したこと、この成果は社会的な課題の解決に大きく貢献すると期待できることより、学位論文として高く評価できる内容であると判断した。

### 最終試験結果の要旨

申請者**Alwis Nazir**は、公聴会において以下の論文リストにある3編の研究成果に基づいて報告した。審査の結果、学位論文およびこれを構成する以下の公表済み学術論文は、博士(医科学)の学位を授与するにふさわしい内容であることを確認した。

また、学位論文に記載されている内容について、その根幹を成す健康医療分野の大規模データの特徴やこれを効果的に分析するためのネットワーク分析手法やHMM(隠れマルコフモデル)手法等に対する質問にも適切に回答したことより、最終試験に合格したと判断した。

### 論文リスト

1. A. Nazir, T. Ichinomiya, N. Miyamura, Y. Sekiya, Y. Kinosada : “ Identification of suicide-related events through network analysis of adverse event reports ” , Drug Safety, 2014, 37(7). 【IF=2.620】

### 参考資料

2. R. Kawamoto, A. Nazir, A. Kameyama, T. Ichinomiya, K. Yamamoto, S. Tamura, M. Yamamoto, S. Hayamizu, Y. Kinosada: “ Hidden Markov model for analyzing time-series health checkup data. ”  
Stud Health Technol Inform. 2013; 192:491-5.
- 3 . Alwis Nazir, Ryouhei Kawamoto, Keiko Yamamoto, Satoshi Tamura, Takashi Ichinomiya, Satoru Hayamizu, Yasutomi Kinosada : “ Time-series analysis of health checkup data using Hidden-Markov Model ” , Annual Symposium American Medical Information Association 2013, 16 - 20 November 2013, Washington DC, USA (Poster Presentation)