Analysis Of Transitions In Differences Between Frequent Medical-order Sequences For COVID-19

Zitai Zhao¹, <u>Hieu Hanh Le^{1,*}</u>, Yuki Yasumitsu¹, Tomoyoshi Yamazaki¹, Kenji Araki², Haruo Yokota¹

- ¹ Tokyo Institute of Technology
- ² University of Miyazaki Hospital

* Recently moved to Ochanomizu University

² EMR Data Analysis

- EMR has been widely used in many hospitals and the data has been analyzed to improve medical tasks
 - Generate clinical pathways by using SPM techniques



EMR: Electronic Medical Records System, SPM: Sequential Pattern Mining

³ One Institute Data Analysis

- Visualize the generated clinical pathways [Honda, DMAH2019]
 - Sequence Variant (SV) is the extension of a sequence with branches
- Understand the background reasons that led to the branches [Le, ACM Healthcare 2023]
 - Specimen test results, medical background, gender, age



Sequence Variants

4 Multiple Institutes Data Analysis

- Analyzing data from multiple medical institutes is desired
 - Compare the differences in the treatment patterns with those in other hospitals [Li, DEXA2022]
 - A hospital can confirm its characteristic and improve medical practices by referring to the treatment patterns of others



Merged Sequence Variants of the two hospitals' frequent treatments

⁵ Challenges

- New treatments and drugs for a disease have been continually developing and changing
 - E.g.: The treatment methods for COVID-19 at each wave may be different
- Understanding the differences in medical-order patterns and the transition history of medical orders is useful
 - When were there differences?
 - What were the differences?

⁶ Contributions

- Propose methods to analyze differences in frequent medicalorder sequences for different infection waves of COVID-19
 - Proposed three types of sliding window to obtain the temporal characteristics of patterns
- Apply the proposed methods to real COVID-19 data acquired from two medical hospitals
- The proposed methods can be applied to diseases other than COVID-19

7 Proposal Data Preprocessing

- Exclude the **not important** medical orders
 - Inspired by TF/IDF algorithm
 - A high-frequency word appears everywhere is not important
 - High frequency but appear in many sequences
 - E.g.: Common specimen inspection such as percutaneous oxygen saturation measurement

8 Proposal Showing the Differences of Frequent SVs

- Apply SPM to each period's data to extract the frequent sequences
- Merge adjacent periods' frequent sequences to the MSV to show the differences COVID-19



LCSV: Longest Common Sequence Variant, MSV: Merged Sequence Variant

9 How to Indicate When the Differences Occurred?

- Identifying which period to generate the SVs is important
- Adopt the concept of the sliding window which requires choosing an appropriate windows type, window size, etc.



¹⁰ Type 1: Time Interval Window

- The window size and slide unit are defined by a time interval
 - The window size is set to a specific time range, and with that range, the SPM is applied to extract frequent patterns
 - Window size = 4 weeks
 - Slide unit = 1 week



¹¹ Time Interval Window's drawback

 Applying SPM may not extract anything for periods with a small number of sequences



¹² Type 2: Sequence Number Window

- The windows size and slide unit are defined by the number of sequences
 - The sequence data are sorted in time order
 - Try to extract frequent patterns with different timing characteristics by calculating the standard deviation (SD) from the data before and after the peak period
 - Window size = 4 SD
 - Slide unit = 2 SD



¹³ Sequence Number Window's Drawback

- The fixed slide unit may skip the sequences at the beginning of the next infection waves
 - Prevent extracting medical-order patterns in such periods



6/23/23

¹⁴ Type 3: Time-sequence Number Window

- Based on the time interval window, while considering the number of sequences in subsequent periods
 - **default slide unit** is initialized (e.g. half of window size)
 - Calculate the **#sequences** for the time range set from the beginning of the window
 - slide unit = #sequences if #sequences < default slide unit
 - slide unit = default slide unit if others

15 Comparisons of Frequent Patterns between Windows

- Comparing adjacent windows one by one, it is possible to detect when there has been a major switch in treatment
- Sequence Variant Similarity

$$SVS(SV_1, SV_2) = \frac{2|LCSV|}{|SV_1| + |SV_2|}, SVS(SV_1, SV_2) \in [0,1]$$

- |*LCSV*|: number of nodes in LCSV
- $|SV_1|$: number of nodes in SV1
- $|SV_2|$: number of nodes in SV2

¹⁶ Experimental Evaluation

- Apply the proposed methods on medical-order data from the actual EMR data for COVID-19 from a hospital*
 - From the first wave to the fifth wave

COVID-19 patient stats

#patients	329
Max #medical_orders	2,319
Longest days in hospital	190

#patients and #sequences per wave

First (2020.04.01-2020.06.30) 42	19
Second (2020.07.01-2020.10.31) 52 6	53
Third (2020.11.01-2021.02.28)9610	28
Fourth (2021.03.01-2021.06.30) 89	99
Fifth(2021.07.01-2021.09.30) 98 10	34

* which was collaborating organizations in the

"Survey of the Impact of COVID-Infection on Medical Practice and Development of a Predictive Model"

¹⁷ Experimental Contents

- Show the transition history of medical treatments from the first to the fifth wave
- Apply three types of windows
 - Time interval window
 - Sequence number window
 - Omitted due to impossibility of extracting meaningful results
 - Time-sequence number window



*Camostat Mesilate showed antiviral activity against COVID-19



*DT reduced the risk of death in COVID-19 patients with severe disease



*MT is a cough suppressant medicine used for COVID-19 patients with mild disease



22 Experimental Results Time-Sequence Number Window



23 Experimental Results Time-Sequence Number Window



24 Discussion of the Difference between the Two Windows Types

Time interval window

- Discover more frequent change of pattern dosage and a wider variety of medications
- Should be used to capture **detail changes**

• Time-sequence number window

- Discover less frequent change without detail
- Should be used to capture **major changes**

²⁵ Conclusion

- Proposed methods to analyze the differences between frequent medical-order patterns during different periods
- Extracted frequent patterns with different temporal characteristics by a applying and adjusting three types of sliding windows
- Positive comments from medical staff have confirmed the visualization results on actual COVID-19 data

26

THANK YOU!

Analysis Of Transitions In Differences Between Frequent Medical-order Sequences For COVID-19