

# Mining Sequential Patterns of Suffered Diseases before Getting Stevens-Johnson Syndrome in Taiwan

**Shih-Chung Chou** †

Department of Industrial Management

National Taiwan University of Science & Technology (Taiwan Tech), Taipei, Taiwan

Tel: (+886) 2-2733-3141 ext. 7656, Email: [D10401002@mail.ntust.edu.tw](mailto:D10401002@mail.ntust.edu.tw)

**Chao Ou-Yang**

Department of Industrial Management

National Taiwan University of Science & Technology (Taiwan Tech), Taipei, Taiwan

Tel: (+886) 2-2737-6340, Email: [ouyang@mail.ntust.edu.tw](mailto:ouyang@mail.ntust.edu.tw)

**Yeh-Chun Juan**

Department of Industrial Engineering and Management

Ming Chi University of Technology, New Taipei City, Taiwan

Tel: (+886) 2-2908-9899, Email: [ycjuan@mail.mcut.edu.tw](mailto:ycjuan@mail.mcut.edu.tw)

**Abstract.** Stevens-Johnson syndrome (SJS) is a rare and life-threatening disease, mainly due to medication reactions. Many drugs may cause SJS and specialized physicians may use them in various disease therapies. However, many physicians do not be aware of the risk of drug-induced SJS and may inadvertently prescribe the SJS-risky drugs to patients. Therefore, this research aimed at identifying the sequential patterns of suffered diseases in the two months before patients got SJS. This research has retrieved about 700 SJS patients' medical records from Taiwan National Health Insurance Research Database. A sequential pattern mining algorithm, Generalized Sequential Patterns (GSP), was applied to the retrieved dataset. The research results can remind specialized physicians to pay attention to SJS in related disease therapies.

**Keywords:** Sequential pattern mining, Stevens-Johnson Syndrome, Prescription behavior.

## 1. INTRODUCTION

Stevens-Johnson syndrome (SJS) is a rare and life-threatening disease which is induced by cutaneous drug reaction, involving skin and mucous membranes (Harr and French, 2010; Raucci et al., 2013). Many physicians do not be aware of the risk of drug-induced SJS. Therefore, identifying the sequential patterns of suffered diseases before patients got SJS may be the first and most important task for related specialized physicians.

In 1995, Agrawal and Srikant (1995) introduced the sequential pattern mining problem which focuses on the analysis of sequencing or time-related processes, such as scientific experiments, natural disasters, disease treatments, and DNA sequences, etc (Jian et al., 2001). They also proposed an algorithm, AprioriAll, to solve the problem. Later, they presented another algorithm, Generalized Sequential

Patterns (GSP), which is much faster than AprioriAll (Srikant and Agrawal, 1996). Therefore, this research used GSP to discover the disease-got patterns before getting SJS from Taiwan National Health Insurance Research Database.

## 2. METHODOLOGY

This research retrieved one-year diagnosis and prescription data of SJS patients from Taiwan National Health Insurance Research Database. Since the highest risk of induction of SJS occurs during the 2 months of treatments before getting SJS (Roujeau et al., 1995), this research further extracted the suffered diseases in the 2 months before patients got SJS.

The input data is a set of sequences of suffered diseases (disease-sequences). A sequential pattern, therefore, consists of a list of sets of diseases. Finally, GSP algorithm was

used to find the sequential patterns of suffered diseases before getting SJS with a minimum support of 10%.

### 3. RESULT ANALYSIS

This research collected 545 SJS patients and their 2779 medical records before getting SJS. These SJS patients include 266 males and 279 females with a mean age of 41 years old.

Table 1 shows the number of SJS patients for different number of times of outpatient visits before getting SJS, i.e. the sequential pattern length. Since the minimum support was set at 10%, i.e. 55 SJS patients, this research only focused on the sequential pattern mining of the first four SJS patient groups, i.e. the groups with 1~4 times of outpatient visits before getting SJS.

Table 2 shows the mined sequential patterns of suffered diseases before getting SJS. Obviously, respiratory system disease (80) always appears in the two months before patients got SJS, no matter how long the sequential patterns are. But, for the SJS patients who received only one outpatient service before getting SJS, diseases about respiratory system (80), skin and subcutaneous tissue (120), signs, symptoms and ill-defined conditions (160), digestive system (90), circulatory system (70) and nervous system and sense organs (60) frequently appear before getting SJS.

Table 1: Number of SJS patients with different number of times of outpatient visits.

Number of times of outpatient visits (Sequential pattern length)	Number of SJS patients
1	129
2	115
3	107
4	71
5	44
6	20
7	22
8	11
9	8
10	6
11	5
12	3
13	2
15	1
19	1
Total	545

Table 2: The mined sequential patterns of suffered diseases before getting SJS.

Sequential pattern length = 1	Support > 129*10% $\cong$ 13
[80]	51 (39.53%)
[120]	30 (23.26%)
[160]	23 (17.83%)
[90]	17 (13.18%)
[70]	15 (11.63%)
[60]	14 (10.85%)
Sequential pattern length = 2	Support > 115*10% $\cong$ 12
[80 80]	23 (20.00%)
[80 120]	14 (12.17%)
Sequential pattern length = 3	Support > 107*10% $\cong$ 11
[80 80 80]	14 (13.08%)
Sequential pattern length = 4	Support > 71*10% $\cong$ 7
[80 80 80 80]	13 (18.31%)

- (60: Nervous System and Sense Organs Disease)
- (70: Circulatory System Disease)
- (80: Respiratory System Disease)
- (90: Digestive System Disease)
- (120: Skin and Subcutaneous Tissue Disease)
- (160: Signs, Symptoms and Ill-Defined Conditions)

### 4. CONCLUSIONS

This research has identified the sequential patterns of suffered diseases in two months before patients got SJS in Taiwan. According to the research results, respiratory system disease (80) is the one that most frequently appears in all sequential patterns. Our findings can remind specialized physicians to pay attention to SJS in related disease therapies, as well as to provide information on safe alternative for further treatment of the patient. In the future, we will continue to discover the taken drugs, then can assist specialized physicians in avoiding prescribing the SJS-risky drugs to patients in specific disease therapy.

### REFERENCES

- Agrawal, R., and Srikant, R. (1995) Mining sequential patterns. *Proceedings of the Eleventh International Conference on Data Engineering*, Taipei, Taiwan, 3-14.
- Jian, Pei, Han Jiawei, B. Mortazavi-Asl, H. Pinto, Chen Qiming, U. Dayal, and Hsu Mei-Chun. 2001. PrefixSpan: mining sequential patterns efficiently by prefix-projected pattern growth. *Proceedings 17th International Conference on Data Engineering*, Heidelberg, Germany, 215-224.
- Harr, T., & French, L. E. (2010). Toxic epidermal necrolysis and Stevens-Johnson syndrome. *Orphanet Journal of Rare Diseases*, 5, 1-11.
- Rauci, U., Rossi, R., Da Cas, R., Rafaniello, C., Mores, N., et al. (2013). Steven-Johnson Syndrome Associated

with Drugs and Vaccines in Children: A Case-Control Study. *PLoS ONE*, **8**(7), e68231.

Roujeau, J.-C., Kelly, J. P., Naldi, L., Rzany, B., Stern, R. S., Anderson, T., Auquier, A., Bastuji-Garin, S., Correia, O., Locati, F., Mockenhaupt, M., Paoletti, C., Shapiro, S., Shear, N., Schöpf, E. & Kaufman, D. W. (1995). Medication use and the risk of Stevens–Johnson Syndrome or Toxic Epidermal Necrolysis, *New England Journal of Medicine*, **333**, 1600-1608.

Srikant, R., and Agrawal, R. (1996) Mining sequential patterns: generalizations and performance improvements. *Proceedings of the 5th International Conference on Extending Database Technology: Advances in Database Technology*, Avignon, France, 1-17.