

Clustering of chief complaint

Jeonghoon Lee, MD¹, Won Chul Cha, MD^{1,2,3}

¹Department of Digital Health, SAIHST, Sungkyunkwan University, Seoul, Korea; ²Department of Emergency Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea; ³Health Information and Strategy Center, Samsung Medical Center, Seoul, Korea

Is this the first time you have submitted your work to be displayed at any OHDSI Symposium?

Yes No

Abstract

Objectives: To grouping CCs with patient's final diagnosis through application of clustering method. *Methods:* 5,789 different CCs were used during the study period, which were represented by Unified Medical Language System (UMLS) code and the number of unique UMLS codes were 1,753. To develop CC clusters, we selected diagnosis that are coded by the International Classification of Disease (ICD) codes. To grouping of CC, we applied T-distributed Stochastic neighbor embedding (T-sne) technique and K-means clustering method.

Results: The most common CC was 'abdominal pain' that recorded 55,625 times and the second most common CC was 'fever' that used 42,371 times. We manipulated many parameters for T-sne and k-means and decided to make 30 clusters.

Conclusions: We made 30 clusters of Chief complaints with unsupervised clustering method and most clusters consist of relational CCs

Introduction

Chief complaint (CC) is a starting point of diagnosis and management of patients visited Emergency Room (ER) for physicians. Since CC itself contains an important information about patients, especially in ER, it can navigate physician's diagnosis flow¹. However, same CCs can be presented with different words because patients usually use variety expression to describe the main reason for visiting ER. Diversity of CC works as barrier for clinical, research area. There are some efforts to developing standard of CC². and finding useful CC categorization for detecting an outbreak in syndromic surveillance³, defining some patient groups^{4,5}. Nevertheless, availability of some sort of standardized CC is still limited for specific object. In this circumstance, the objective of this study is to grouping CCs with patient's final diagnosis through application of clustering method.

Methods.

There were 734,476 ER visit records stored from 2008 to 2018 in Clinical Data Warehouse (CDW) of Samsung medical center, Korea. We removed 1,723 visit record which have no information about CC and included only patients who were older than 18 years old. Total of 654,271 cases were available for analysis. 5,789 different CCs were used during the study period, which were represented by Unified

Medical Language System (UMLS) code and the number of unique UMLS codes were 1,753. To develop CC clusters, we selected diagnosis that are coded by the International Classification of Disease (ICD) codes. To grouping of CC, first, we applied T-distributed Stochastic neighbor embedding (T-sne) technique to reduce dimensionality of CC dataset. After that CCs were clustered by means of K-means clustering method.

Results

During study period, patients who visited ER complained of 1,753 unique CCs in 5,789 different expressions. And the number of probable UMLS codes in Samsung medical center is 36,701. The most common CC was 'abdominal pain' that recorded 55,625 times and the second most common CC was 'fever' that used 42,371 times. We manipulated many parameters for T-sne and k-means and decided to make 30 clusters. Table 1. shows top 5 clusters that total frequency of contained CCs is high with comprised top 20 high frequency CCs. This unsupervised CC clustering method is not a perfect way because some CCs were unacceptably included specific cluster. In spite of this result, the rest part of clusters seem like reasonably constructed with well related CCs. And the results is considered that CCs of each cluster reflect attributes of patients who suffered specific disease. For example, many of CCs that consist of 5th cluster in table1 are related with hepatobiliary system. Interestingly CCs associated with mental status are also contained in this cluster which means these CCs can be connected to 'hepatic encephalopathy'. From this result, we found probability of using unsupervised method to CC clustering.

Conclusion

In this study, we made 30 clusters of Chief complaints with unsupervised clustering method and most clusters consist of relational CCs. On account of diversity in expression, the effort of CC clustering could be helpful to clinical and research area. Because this way is not perfect, we are going to include more patient information and expert opinion to develop more reasonable and applicable CC clusters.

Table 1. Top 5 most frequently used cluster group.

No	Cluster feature	% of visit	Number of Included CC	Top 20 Included CC (Total frequency)
1	Fever related symptom	21.87%	44	Fever(42,317), dyspnea(25,319), cough(4,655), Influenza(4,157), Fever with chills(3,436), COUGHED SPUTUM(3,229), Sore Throat(3,000), DYSPNEA ACUTE(2,933), Hemoptysis(2,849), General weakness(2,131), Cough with fever(1,992), Myalgia(1,476), FEVER HIGH(1,129), Hyperventilation(980), Fatigue(934), Common Cold(925), for medication(860), PO INTAKE POOR(844), DYSPNEA PROGRESSIVE(640), Chills(568)
2	Abdominal symptom	17.68%	69	Abdominal pain(55,625), Diarrhea(5,719), Flank Pain(4,824), Nausea and vomiting NOS(4,646), Epigastric pain(3,800), Left Flank Pain(2,005), Right Flank Pain(1,715), Abdominal discomfort(1,609), ABDOMEN PAIN DIFFUSE(823), ABDOMEN RLQ PAIN(719), ABDOMEN EPIGASTRIUM PAIN(700), Burning epigastric pain(693), Percutaneous insertion of nephrostomy tube(519), Maintenance of drainage tube of kidney(421), Lower abdominal pain(355), Dyspepsia(315), Diarrhea and vomiting, symptom(302), Abdominal Colic(252), Inguinal pain(240), Acute gastroenteritis(156)
3	Trauma related head and facial symptom	5.62%	91	Syncope(4,599), Neck pain(3,298), Trauma history(2,254), Traffic accidents(2,079), Facial laceration(1,952), Laceration(1,664), Scalp laceration(1,365), Facial Pain(920), injuries(777), Slipping(623), Laceration of forehead(598), Facial Injuries(537), Brain Concussion(535), Contusion of face(521), Brain Contusion(516), Contusion of scalp(502), Contusions(462), CHIN LACERATION(451), Loss of consciousness(436), Fall, NOS(393)
4	Neurologic symptom	4.94%	68	Dizziness(21,606), Vertigo(694), Numbness(679), HAND PARESTHESIA TINGLING(360), Gait abnormality, NOS(336), Tingling feet/hands(222), Weakness of distal arms and legs(147), Gait abnormality(141), Dysequilibrium(118), aphagia(105), Positional Vertigo(99), FINGER WEAKNESS(84), Ataxia(70), Paresthesia(69), LOWER EXTREMITY PARESTHESIA TINGLING(63), WEAKNESS TRANSIENT(63), Numbness in leg(62), Sensory Disorders(61), Dysesthesia(61), Presyncope(57)
5	Hepatobiliary symptom	4.81%	34	Ascites(3,534), Abnormal mental state(3,395), Melena(3,378), Hematochezia(3,335), Jaundice, obstructive(2,418), Hematemesis(2,124), Abdominal distention(1,550), Abdominal distension symptom(747), Drainage Tube Care(559), Confusion(341), Irrigation of cholecystostomy and other biliary tube(259), Disorientation(218), Hyperbilirubinemia(183), Icterus(181), Percutaneous transhepatic insertion of biliary drain(177), Abdominal distension, gaseous(147), Paracentesis(116), Hepatic Encephalopathy(106), Abdominal paracentesis(104), Bilirubin total increased(75)

References

1. Richard T Griffey, MD, MPH, Jesse M. Pines, MD, Heather L. Farley, MD, Michael P Phelan, MD, Christopher Beach, MD, Jeremiah D Schuur, MD, MHA, and Arjun K. Venkatesh, MD, MBA. Chief complaint-based performance measures: a new focus for acute care quality measurement. *Ann Emerg Med.* 2015;65(4):387–395
2. Aronsky D, Kendall D, Merkeley K, James BC, Haug PJ. A comprehensive set of coded chief complaints for the emergency department. *Acad Emerg Med.* 2001; 8:980–9
3. Philip Brown, Sylvia Halasz, Colin Goodall, Dennis G. Cochrane, Peter Milano, John R. Allegra. The ngram chief complaint classifier: a novel method of automatically creating chief complaint classifiers based on international classification of disease groups. *J. Biomed. Inform.* 2010;43:268–272
4. Allison J. Beitel, MD, Karen L. Olson, PhD, Ben Y Reis, PhD, Kenneth D. Mandl, MD, MPH. Use of emergency department chief complaint and diagnostic codes for identifying respiratory illness in a pediatric population. *Pediatric Emergency Care.* 2004;20(6):355-360.
5. Marc H. Gorelick, MD, MSCE, Elizabeth R. Alpern, MD, MSCE, Evaline A. Alessandrini, MD, MSCE. A System for Grouping Presenting Complaints: The Pediatric Emergency Reason for Visit Clusters. *ACADEMIC EMERGENCY MEDICINE* 2005; 12:723–731