Challenges of Trustable AI and Added-Value on Health
B. Séroussi et al. (Eds.)

doi:10.3233/SHTI220502

CDS-Compare: A Web Application for Machine Learning Assisted Curation of Clinical Order Sets

Zackary FALLS[a], Steven H BROWN[b,c], Naveed SHAH[a], Frank LeHOUILLIER[a], Randeep BADWAL[a], Kendria HALL[a], Gillian FRANKLIN[a], William H KELLY[a], Andrew HOLDAWAY[b], Diane MONTELLA[b], Linda WEDEMEYER[b], Sarita KENI[b], Eric ROSE[b,d], Jonathan R. NEBEKER[b,e] and Peter L ELKIN[a,b,f,g,1]

[a] *Department of Biomedical Informatics, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, USA*
[b] *Department of Veterans Affairs, Office of Health Informatics, USA*
[c] *Department of Biomedical Informatics, Vanderbilt University Medical Center, USA*
[d] *Department of Biomedical Informatics and Medical Education, University of Washington, USA*
[e] *Department of Medicine, University of Utah, USA*
[f] *Department of Veterans Affairs, WNY VA, Bioinformatics Laboratory, USA*
[g] *Faculty of Engineering, University of Southern Denmark, Denmark*

Abstract. Order sets that adhere to disease-specific guidelines have been shown to increase clinician efficiency and patient safety but curating these order sets, particularly for consistency across multiple sites, is difficult and time consuming. We created software called CDS-Compare to alleviate the burden on expert reviewers in rapidly and effectively curating large databases of order sets. We applied our clustering-based software to a database of NLP-processed order sets extracted from VA's Electronic Health Record, then had subject-matter experts review the web application version of our software for clustering validity.

Keywords. Order sets, Database Curation, Clinical Decision Support, Machine Learning

1. Introduction

Clinical order sets, collections of orders for a clinical scenario, are one aspect of clinical decision support that has been shown to improve clinician efficiency and patient safety [1,2]. Each clinical site may have a unique aggregation of orders into an order set for any given condition or situation, and certain sites may use different order sets with more frequency due to population factors, such as prevalence of certain diseases [2]. The use of order sets has also been shown to reduce the variation in patient care for a single site; however, order sets vary across different sites, so variation in care exists from site to site [2]. If the goal is to adopt guideline-based order sets, then they should be standardized across multiple sites [3]. This is especially true for large hospital networks like the US

[1] Corresponding Author, Peter L. Elkin, Buffalo, NY, USA; E-mail: elkinp@buffalo.edu.

Department of Veterans Affairs (VA) and its nationwide network of medical facilities. However, creating order sets is time and resource intensive [2]. If we can reduce the difficulty in order set curation within and between facilities, then we can take steps towards a system-wide database of patient-centered order sets that adhere to condition-specific guidelines, ultimately reducing unwanted variation in patient care across all VA medical facilities [4,5]. We further anticipate that this ability will help smooth VA's transition from 128 legacy CPRS installations to a single Cerner Millennium instance shared with the US Department of Defense by reducing multi-site order sets to a single, well-curated dataset (https://www.ehrm.va.gov/resources/factsheet) [6].

This study describes the clustering performance of a clustering method in the pragmatic setting of a web application, CDS-Compare, that will enable experts to curate clinical order set databases. We have created clinically relevant clusters using a hierarchical clustering method and an effective Natural Language Processing (NLP) algorithm to create features for each order set based upon standardized terminologies. This methodology was applied to a VA order set dataset of medication orders and embedded within a user-friendly web application for end-user experts to access and rapidly curate the dataset. To evaluate our software's performance and usability, we asked ten clinicians to demo our web application, assess the clustering efficacy by comparing inter- and intra-cluster order sets, and then fill out a questionnaire with pointed questions regarding the usability of the interface.

2. Methods

2.1. Order sets extraction and featurization

The database of order sets used in this study were extracted from the VA's Electronic Health Record, a component of the VA's Electronic Health Record Modernization initiative. The dataset used herein consisted of 1,293 order sets with an average number of orders/per order set of 12.6 (minimum = 1, maximum = 141). This dataset was a subset of prescription medication order sets and was not the complete set of order sets across all VA medical sites. The complete dataset will contain both medications and procedures.

Each order set consists of one or numerous medication orders,. We wanted a way to compare all of the order sets by the orders which they contained, therefore we applied High-Definition Natural Language Processing (HD-NLP), a well-tested, high-speed system that processes clinical data and extracts clinically relevant terms in a standardized form, e.g., SOLOR, SNOMED-CT, RxNorm, and LOINC [7,8]. Processing of the study order sets with HD-NLP resulted in 860 unique SNOMED-CT, RxNorm, and LOINC codes across the 1,293 order sets present in our dataset, and every order was mapped to at least one code. We used the 860 unique codes extracted with HD-NLP as features to represent each of the order sets as a bit vector of length 860 with each bit representing the presence or absence of a code.

2.2. Hierarchical clustering

We leveraged an agglomerative hierarchical clustering algorithm to enable easier and more rapid curation of order sets. We determined the optimal cosine distance threshold using the elbow heuristic method which results in the clusters that best fit the data. After computing clusters for the order sets, we wanted to provide the end-users with a useful

visualization of the order sets and their corresponding clusters. We used the dimensionality reduction method t-SNE (t-distributed stochastic neighbor embedding) to reduce the 860-dimension vectors to two dimensions allowing for easier visualization.

2.3. Web application

The last stage of our software development was the creation of the web application interface, where users will take advantage of the clustering method employed to curate the dataset of order sets more easily. The first component of the software, a table containing all order sets, provides the user with the specific name of each order set and the cluster to which it belongs so the user can use both pieces of information to isolate order sets or groups of order sets that may need to be assessed together. Second is the t-SNE plot (Figure 1) , which is embedded into the web interface as a 2D visualization of the order sets and their clusters to allow the user to select specific order sets for view in the comparison tables. All or user-selected order sets are populated in the plot with the marker color-coded based upon the cluster to which the order set belongs. This enables users to easily identify order sets that are most similar by color and relative location to one another.

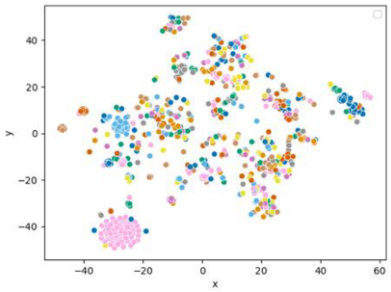

Figure 1. t-SNE plot for visualization of clusters of order sets where each color represents a unique cluster.

Lastly, the side-by-side comparison tables are where users can view two order sets at a time with all the orders and pertinent information corresponding to those order sets. This component is critical for users to determine if there is redundancy in order sets or large overlap where two order sets can be merged to create more robust order sets for clinical use. HD-NLP extracted codes used for clustering are not presented to the user.

Ten clinicians assessed the application clustering performance in CDS-Compare by selecting five order set pairs where both order sets in each pair belonged to the same cluster, and determined the similarity between the two order sets in each pair (intra-cluster comparison) based on the orders (prescription medications). The clinicians assessed each order set pair using three mutually exclusive choices: “same” - order set pairs were 100% identical; “different” - order set pairs were unique with minimal to no overlap in orders; and “overlapping” - order set pairs had substantially similar orders. The same ten clinicians then repeated this procedure on five order set pairs where any two order sets in each selected pair belonged to different clusters (inter-cluster comparison). We state that the clustering works well if order sets are the same or highly overlapping intra-cluster significantly more often than order sets in different clusters.

3. Results

The ten reviewers in all analyzed 50 intra-cluster order set pairs and 50 inter-cluster order set pairs. The cluster validity results showed 66% of the intra-cluster order set pairs assessed had overlap and 22% had the same orders. The inter-cluster assessment showed 72% of order set pairs were different and 28% had overlap. The intra-cluster analysis showed 44/50 were either the same or overlapping, while the inter-cluster analysis showed 14/50 order set pairs were overlapping ($p<0.001$; Pearson Chi-Square test).

We asked the reviewers to rate aspects of their subjective experience with the web application by indicating the degree to which they agreed or disagreed with 10 statements using a Likert scale of 1 to 5 (1: strongly disagree, 2: disagree, 3: neutral, 4: agree, 5: strongly agree); results are aggregated in Table 1. They found that order set contents were easy to view (average score of 4.40), they were able to make judgments about order set content (4.40), and the web application was both easy to access (4.40) and visually pleasing (4.50). Further development of CDS-Compare is needed for clinical understanding (4.00) and consistency across reviewers (3.90) .

Table 1. Questions and corresponding results from our Likert scale questionnaire.

Questions	Average score	Median score	Minimum score	Maximum score
The system was easy to access?	4.40	5	2	5
The system was easy to use?	3.90	4	2	5
You could view the contents of the order sets?	4.40	5	2	5
You could make judgments about the content of the order sets?	4.50	4.5	4	5
The system was visually pleasing?	4.50	5	3	5
The ability to focus on certain topics was helpful?	4.10	4	3	5
The system was easy to understand from a clinical perspective?	4.00	4	2	5
The system had the functionality I needed?	4.10	4	2	5
The system in my opinion could be used consistently across reviewers?	3.90	4	2	5
I enjoyed using the system?	4.30	5	2	5

4. Discussion

Analysis of the cluster validity assessments show that the software is clustering the order sets in a way that effectively groups order sets that are the same or contain highly overlapping orders into the same cluster. Moreover, the order sets in different clusters were significantly less likely to be overlapping. These results support the use of this application to reduce the difficulty for experts to select and remove redundant order sets in the database. Use of this application by experts will result in a highly curated set of order sets that will reduce variation in patient care across VA medical facilities.

With regards to the usability of the web interface, the overall responses were very positive. Many aspects of the web interface were well liked by the reviewers, such as ease of access and ability to make judgments about the order sets. We still need to further develop components of the interface to enhance clinical understanding of the order sets for users and improve potential inter-reviewer consistency. Constructive comments provided by the reviewers also alluded to some of the needed improvements, such as more clinically relevant variables for the orders and clearly written instructions on use of the application. We continue to improve the interface using clinician feedback.

5. Conclusion

Based on the results obtained from our reviewer assessment of the clustering and web interface for CDS-Compare, we have a promising web application for curation and accurate clustering of large order set databases. The clustering method used within the application is effective at grouping very similar or exact order sets, while separating those that are completely different or have limited overlap in orders. The web interface also received strong support from the reviewers and the constructive criticism will help guide us in our development of a user-friendly and effective order set curation software.

Our future work will focus on continued development of the software to improve the usability based upon feedback from the reviewers. Additional functionality to enable users to create a separate curated database by selecting, removing, merging, and modifying existing order sets in the database will be considered, as well as features to aid the decision making. Lastly, a comprehensive inter-rater reliability study will be conducted on the full order set database with the goal of documenting output reliability. We believe the combination of NLP and clustering methodologies has great potential for the curation of Clinical Decision Support artifacts in ways that can reduce unnecessary variation within and among clinical sites in the VA network and other medical facilities. In so doing we believe we can improve care provided to Veterans.

Acknowledgements

This work has been supported in part by grants from NIH NLM T15LM012495, NIAAA R21AA026954, R33AA0226954 and NCATS UL1TR001412. This study was funded in part by the Department of Veterans Affairs.

References

[1] Kuperman GJ, Gibson RF. Computer physician order entry: benefits, costs, and issues. Annals of internal medicine. 2003 Jul 1;139(1):31-9.

[2] Wright A, Sittig DF, Carpenter JD, Krall MA, Pang JE, Middleton B. Order sets in computerized physician order entry systems: an analysis of seven sites. AMIA Annu Symp Proc. 2010 Nov 13;2010:892-6.

[3] Ozdas A, Speroff T, Waitman LR, Ozbolt J, Butler J, Miller RA. Integrating "best of care" protocols into clinicians' workflow via care provider order entry: impact on quality-of-care indicators for acute myocardial infarction. J Am Med Inform Assoc. 2006 Mar-Apr;13(2):188-96.

[4] Hulse NC, Lee J, Borgeson T. Visualization of Order Set Creation and Usage Patterns in Early Implementation Phases of an Electronic Health Record. AMIA Annu Symp Proc. 2017 Feb 10;2016:657-666.

[5] Muniga ET, Walroth TA, Washburn NC. The impact of changes to an electronic admission order set on prescribing and clinical outcomes in the intensive care unit. Applied clinical informatics. 2020 Jan;11(01):182-9.

[6] Brown SH, Lincoln MJ, Groen PJ, Kolodner RM. VistA--U.S. Department of Veterans Affairs national-scale HIS. *Int J Med Inf.* 2003;69(2–3):135–56.

[7] Resnick MP, LeHouillier F, Brown SH, Campbell KE, Montella D, Elkin PL. Automated Modeling of Clinical Narrative with High Definition Natural Language Processing Using Solor and Analysis Normal Form. Stud Health Technol Inform. 2021 Nov 18;287:89-93.

[8] Elkin PL, Mullin S, Mardekian J, Crowner C, Sakilay S, Sinha S, Brady G, Wright M, Nolen K, Trainer J, Koppel R, Schlegel D, Kaushik S, Zhao J, Song B, Anand E. Using Artificial Intelligence With Natural Language Processing to Combine Electronic Health Record's Structured and Free Text Data to Identify Nonvalvular Atrial Fibrillation to Decrease Strokes and Death: Evaluation and Case-Control Study. J Med Internet Res. 2021 Nov 9;23(11):e28946.