MEDINFO 2013
C.U. Lehmann et al. (Eds.)

doi:10.3233/978-1-61499-289-9-1101

Standardizing Drug Adverse Event Reporting Data

Liwei Wang[a,b]; Guoqian Jiang[b]; Dingcheng Li[b]; Hongfang Liu[b]

[a] *School of Public Health, Jilin University, Changchun, Jilin, China;*
[b] *Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA*

Abstract

Normalizing data in the Adverse Event Reporting System (AERS), an FDA database, would improve the mining capacity of AERS for drug safety signal detection. In this study, we aim to normalize AERS and build a publicly available normalized Adverse drug events (ADE) data source. The drug information in AERS is normalized to RxNorm, a standard terminology source for medication. Drug class information is then obtained from the National Drug File - Reference Terminology (NDF-RT). Adverse drug events (ADE) are aggregated through mapping with the PT (Preferred Term) and SOC (System Organ Class) codes of MedDRA. Our study yields an aggregated knowledge-enhanced AERS data mining set (AERS-DM). The AERS-DM could provide more perspectives to mine AERS database for drug safety signal detection and could be used by research community in the data mining field.

Keywords:

Adverse drug events, adverse event reporting system (AERS), data normalization, data mining, knowledge discovery

Introduction

The US Food and Drug Administration (FDA) Adverse Event Reporting System (AERS) is one of the main resources in post-marketed adverse drug events (ADE) detection based on data mining techniques. However, most of existing data mining studies on the AERS were carried out for a small number of drugs [1], and few studies focused on large-scale mining or on detecting the etiology of ADE signals in terms of mechanism of action, physiologic effect, or molecular structure. We realize that potential of the AERS has not been fully utilized, and one of main reasons for this is the lack of standardization among drug names. There have been some attempts in drug name normalization in the AERS, but typically it is either unclear how the normalization was conducted or the normalization was attempted only for a small number of drugs [2]. In this study, we aim to produce an open-source AERS data mining set (AERS-DM), which is normalized and aggregated with a number of standard biomedical ontologies, including RxNorm, the National Drug File-Reference Terminology (NDF-RT), and the Medical Dictionary for Regulatory Activities (MedDRA).

Methods

The method contains three steps: de-duplication, drug normalization, and data aggregation. In the de-duplication step, redundant reports are removed. In the normalization step, MedEx, an natural language processing tool, is applied to normalize AERS drugs to RxNorm codes. During aggregation, adverse events are aggregated according to MedDRA SOC and PT codes, and NDF-RT–based classification information for those drugs is obtained from RxNorm. Two tables are formed; one stores the normalized Drug-ADE information and the other stores the aggregated information of Drug-ADE. The data in the two tables can be connected through the RxNorm codes.

Results

After de-duplicating reports, according to the FDA's recommended method, the number of AERS records is reduced to 2,643,979 from the original 3,874,965. The number of unique "verbatim" drug names is reduced to 1,517,811 from the original 1,700,925. For drug name normalization, 1,125,045 of 1,517,811 (74%) AERS unique drug names were normalized to 14,489 unique RxNorm concepts, of which 10,221 (71%) were classified to 9 classes in NDF-RT. For the ADE normalization, we mapped 14,740 existing MedDRA PT terms in the AERS to MedDRA codes, accounting for 76% of 19,294 total MedDRA PT terms. These MedDRA PT codes were then mapped to 26 MedDRA SOCs. The AERS-DM includes two tables, as discussed above, and they can be downloaded from the website
http://informatics.mayo.edu/adepedia/index.php/Download.
There are 37,029,228 Drug-ADE records after de-duplication. The number of unique pairs between RxNorm concepts and MedDRA PT codes is 4,639,613, and between RxNorm concepts and SOC codes after ADE aggregation, 205,725.

Conclusion

In this study, we leveraged three biomedical ontologies—RxNorm, NDF-RT, and MedDRA—for normalizing and aggregating the AERS data and produced an AERS-DM. With the normalized codes and aggregated features, the AERS-DM would be useful for the research community in the data mining field. We will demonstrate the usefulness of the AERS-DM with case studies and continue to refine and optimize the AERS-DM and update it periodically in the future.

References

[1] Harpaz, R., et al., Biclustering of adverse drug events in the FDA's spontaneous reporting system. Clinical Pharmacology & Therapeutics, 2010. 89(2): p. 243-250.

[2] Vilar S, Harpaz R, Chase HS, Costanzi S, Rabadan R, Friedman C. Facilitating adverse drug event detection in pharmacovigilance databases using molecular structure similarity: application to rhabdomyolysis. Journal of the American Medical Informatics Association. 2011;18(Suppl 1):i73-i80.