Standardized Drug and Pharmacological Class Network Construction

Qian Zhu[a], **Guoqian Jiang**[a], **Liwei Wang**[b], and **Christopher G. Chute**[a]

[a]Department of Health Sciences Research, Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, USA

[b]School of Public Health, Jilin University, Changchun, Jilin, China

Abstract

Dozens of drug terminologies and resources capture the drug and/or drug class information, ranging from their coverage and adequacy of representation. No transformative ways are available to link them together in a standard way, which hinders data integration and data representation for drug-related clinical and translational studies. In this paper, we introduce our preliminary work for building a standardized drug and drug class network that integrates multiple drug terminological resources, using Anatomical Therapeutic Chemical (ATC) and National Drug File Reference Terminology (NDF-RT) as network backbone, and expanding with RxNorm and Structured Product Label (SPL). In total, the network consists of 39,728 drugs and drug classes. Meanwhile, we calculated and compared structure similarity for each drug / drug class pair from ATC and NDF-RT, and analyzed constructed drug class network from chemical structure perspective.

Methods

In this paper, we introduce a drug and drug class network built by using multiple drug terminological resources: ATC, NDF-RT, RxNorm, and SPL.

Mapping NDF-RT to ATC

To map NDF-RT to ATC via UMLS[1], we translated NUI, which is NDF-RT Numerical Unique Identifier, and ATC name to UMLS CUI (UMLS concept unique identifier) by invoking NLM RxNav RESTful API [2] and NCBO annotator [3] respectively. Mappings between NDF-RT and ATC have been utilized to build drug network backbone.

Mapping NDF-RT to RxNorm and UMLS

RxCUIs were retrieved from RxNorm RXNCONSO table that was preloaded into our local MySQL database for NUIs, followed by UMLS CUI retrieval by invoking NLM RxNav RESTful API [1] with each NUI as an input parameter.

Calculating structure similarity

To analyze and expand the drug and drug class network from chemical structure perspective, we calculated the structure similarity among the drugs from ATC and NDF-RT by using the Chemistry Development Kit (CDK) [4], and grouped them using the score of structure similarity.

Integrating RxNorm and SPL mappings

We extracted existing RxNorm and SPL mappings with NDF-RT from RxNorm RXNCONSO table directly. The network has been expanded from NDF-RT nodes with

mappings between RxNorm and SPL. We extracted SPL identifier (setId) from RXNREL table and saved for future SPL relevant information integration.

Results and discussion

In this study, we successfully built a drug and drug class network based on 39,728 concepts from ATC and NDF-RT. All concepts were mapped to UMLS and represented as UMLS CUIs accordingly. We also integrated RxNorm and SPL, and structure similarity calculation to extend the network. In total, 3,607 ATC names comprising 3,152 drugs and 455 drug classes were mapped to UMLS CUIs by RxNorm and NDF-RT from NCBO BioPortal ontologies. 99.2% NDF-RT concepts had their corresponding UMLS CUIs identified. In total, 3,850 unique mappings were generated, including 2,015 chemical/ingredients, 1,826 Generic Ingredient Combinations and 1 VA class. The mappings between RXNORM, NDF-RT and MTHSPL resulted in 5,838 unique RxNorm concepts with 36,408 NDF-RT concepts and 41,188 SPL labels. The mappings mostly fall into two main categories according to term types defined by RxNorm, 3,056 are Semantic Clinical Drugs and 1,543 are Ingredients.

Conclusion

We successfully integrated NDF-RT, ATC, RxNorm and SPL and built a drug and drug class network using standardized identifier for representing drug and drug class entities. In addition, the network was expanded from chemical structure perspective by similarity calculation. More other drug terminological resources and drug interaction information will be integrated in the future study.

Acknowledgments

References

1. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res. 2004; 32:267–270.

2. [Accessed by Dec. 2, 2012] NDF-RT RESTful API. http://rxnav.nlm.nih.gov/NdfrtRestAPI.html#label:r24

3. Jonquet, C.; Shah, N.; Musen, M. [Accessed by Dec. 2, 2012] The Open Biomedical Annotator; AMIA Summit on Translational Bioinformatics. 2009. p. 56-60.The NCBO Annotator web service: http://www.bioontology.org/annotator-service

4. Steinbeck C, Han Y, Kuhn S, Horlacher O, Luttmann E, Willighagen E. The Chemistry Development Kit (CDK): an open-source Java library for chemo- and bioinformatics. J Chem Inf Comput Sci. 2003; 43(2):493–500. [PubMed: 12653513]