Artificial Intelligence in the Life Sciences 3 (2023) 100067

Contents lists available at ScienceDirect

Artificial Intelligence in the Life Sciences

journal homepage: www.elsevier.com/locate/ailsci

Editorial

Data science and data analytics in life science research

Jürgen Bajorath

Department of Life Science Informatics and Data Science, B-IT, LIMES Program Unit Chemical Biology and Medicinal Chemistry, Rheinische Friedrich-Wilhelms-Universität, Friedrich-Hirzebruch-Allee 5/6, D-53115 Bonn, Germany.

ARTICLE INFO

Keywords:
Data science
data analytics
artificial intelligence
big data
interdisciplinary research

Data science (DS) [1,2] and data analytics (DA) [3] are closely related yet self-contained fields, which is also reflected by the availability of different educational programs in DS and DA [4]. Both DS and DA concentrate on exploratory data analysis. Common tasks include the implementation and application of software tools, algorithms, and statistical techniques for data processing, organization, integration, and visualization – with the ultimate goal of extracting knowledge from data. A major distinguishing feature is the much stronger focus of DS on data models. DS preferentially extends the identification of data patterns and statistical trends through machine learning and predictive modeling [1,2]. On the other hand, DA typically focuses on tasks such as the curation of data from a given source or application, followed by extraction of insights and knowledge to address specific questions. Hence, DS is generally more algorithm- and prediction-oriented than DA.

Over the past decade, DS and DA have steadily gained in importance due to massively increasing volumes of structured and unstructured data in many areas [5–7]. In addition to data volumes, other 'big data' characteristics challenge DS and DA including the speed of data generation, data variability, resolution, heterogeneity, or confidence [5–7].

Across the life sciences including drug discovery, data heterogeneity is a major issue that complicates interdisciplinary research and requires strong contributions of DS and DA towards data integration [8]. Moreover, life science data have discipline-specific attributes and DS/DA specialists must also have scientific domain knowledge and the ability to operate and communicate at interfaces between different disciplines. Cleary, contributions from DS and DA experts are essential for developing data-driven concepts in life science and drug discovery research [8].

With increasing emphasis on machine learning and artificial intelligence in many areas, DS continues to be on the rise. However, this does not alleviate the need for DA, especially in specific project contexts. Learning from data does not depend on machine learning and data models. However, careful data analysis and curation are pre-requisites for knowledge extraction, regardless of the methods used.

It is emphasized that DS and DA contributions are integral to the scope of *AILSCI*. For example, the *Methods & Protocols* manuscript category offers an excellent forum for the communication of new methodological developments in DS and DA and publication of study or data curation protocols for specific applications. Going forward, contributions that highlight specific aspects of data analysis and knowledge extraction or particular challenges and possible solutions are strongly encouraged. Clearly, machine learning depends on available data and their integrity. The better source data are characterized and understood the more likely predictions will be successful. Hence, DS, DA, and artificial intelligence research must inevitably be considered in context, emphasizing their relevance for *AILSCI*.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

No data was used for the research described in the article.

References

[1] Cao L. Data science: a comprehensive overview. ACM Comput Surv 2017;50:1–42.
[2] Dhar V. Data science and prediction. Commun ACM 2013;56:64–73.

E-mail address: bajorath@bit.uni-bonn.de

https://doi.org/10.1016/j.ailsci.2023.100067
Received 25 February 2023; Accepted 26 February 2023
Available online 27 February 2023

[3] Runkler TA. Data analytics. 3rd Editon. Fachmedien Wiesbaden: Springer; 2020.
[4] Aasheim CL, Williams S, Rutner P, Gardiner A. Data analytics vs. data science: a study of similarities and differences in undergraduate programs based on course descriptions. J Inf Syst Educ 2015;26:103.
[5] Marx V. The big challenges of big data. Nature 2013;498:255–60.
[6] Provost F, Fawcett T. Data science and its relationship to big data and data-driven decision making. Big Data 2013;1:51–9.
[7] Agarwal R, Dhar V. Big data, big science, and analytics: the opportunity and challenge for IS research. Inform Syst Res 2015;25:443–8.
[8] Bajorath J. Foundations of data-driven medicinal chemistry. Future Sci OA 2018;4:FSO320.