Februrary 2017

Data Science and Symbolic AI: synergies, challenges and opportunities

Robert HOEHNDORF [a,b], Núria QUERALT-ROSINACH [c]
[a] *Computational Bioscience Research Center, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia*
[b] *Computer, Electrical and Mathematical Sciences & Engineering Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia*
[c] *Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, USA*

Abstract. Symbolic approaches to artificial intelligence represent things within a domain of knowledge through physical symbols, combine symbols into symbol expressions and structures, and manipulate symbols and symbol expressions and structures through inference processes. While a large part of Data Science relies on statistics and applies statistical approaches to artificial intelligence, there is an increasing potential for successfully applying symbolic approaches as well. Symbolic representations and symbolic inference are close to human cognitive representations and therefore comprehensible and interpretable; they are widely used to represent data and metadata, and their specific semantic content must be taken into account for analysis of such information; and human communication largely relies on symbols, making symbolic representations a crucial part in the analysis of natural language. Here we discuss the role symbolic representations and inference can play in Data Science, highlight the research challenges from the perspective of the data scientist, and argue that symbolic methods should become a crucial component of the data scientists' toolbox.

Keywords. symbolic AI, machine learning, statistics, empirical science

1. Introduction

1.1. What is Data Science?

The observation of and collection of data about natural processes to obtain practical knowledge about the world has been crucial for our survival as a species, as well as to satisfy our curiosity and to understand the world in which we are living. The detection of regularities such as the daily movement of the sun resulted in the development of calendars, predicting the migratory movements of animals, to develop methods that allow to control nature for a more productive agriculture, and in stellar almanacs that aid in navigation. The regularities upon which these discoveries were based derive from a careful observation and collection of records of astronomic events by ancient cultures. Astronomy, considered the first science or system of knowledge of natural phenomena, led to the development of mathematics in Mesopotamia, China, or India. In the Middle

East, Egypt and Mesopotamia expanded and used mathematics for the description of astronomic phenomena as an intellectual play, and generated large volumes of data about stellar phenomena [8]. Therefore, could we consider ancient Babylonians or Egyptians as the first, or early, data scientists?

Science, as a way of studying and understanding the physical world, is closely tied to data. Experiments to test hypothesis are the starting point of data generation. The so-called *data life cycle* mainly consists of collecting, processing, analyzing, preserving, giving access and re-using the data. These steps require decisions and tasks on management, storage, (meta)data description, interpretation, archival, publishing, distribution, and revision, among others. Decisions and actions undertaken during the experiment and its design will affect the rest of the data cycle, and therefore the generation of data (through experiments or other means) should also be included in the data life cycle.

Recent advancements in science and technology have led to an explosion of our ability to generate and collect data, and led to the era of *Big Data*. Data is now "big" in volume as well as in heterogeneity (including different representation formats such as digitized text, audio, video, web logs, transactions, time series, or genome sequences), and complexity (from multiple sources and about different phenomena spanning several levels of granularity, possibly incomplete, unstructured, and of uncertain provenance and quality). Large amounts of complex data are not only generated in empirical science but data collection and generation now penetrates our whole life: mobile phones, Internet of Things, social interactions and communication patterns, bank transaction, personal fitness trackers, and many more. Often, data is collected first and retained to solve specific questions whenever they arise.

Data Science has as its subject matter the extraction of knowledge from data. While data has been analyzed and knowledge extracted for millenia, the rise of "Big" data has led to the emergence of Data Science as its own discipline that studies how to translate data through analytical algorithms typically taken from statistics, machine learning or data mining, and turning it into knowledge. Data Science also encompasses the study of principles and methods to store, process and communicate with data throughout its life cycle, and starts just after data has been acquired. While the data acquisition process from experiments is, arguably, not a part of Data Science, capture and analysis of (meta)data about the measurement and data generation process falls in the realm of Data Science. In addition, to the analysis, Data Science studies how to store data, and methods such as "content-aware" compression algorithms (e.g., for genomic data [46]) also fall in the subject matter of Data Science. We consider Data Science as an emerging discipline at the intersection of fields in science, technology, and humanities, drawing upon methods from social science, statistics, information theory, computer science, medicine, biology, energy, finance, meteorology, particle physics, astrophysics, healthcare and more (see Figure 1).

1.2. Artificial Intelligence

Data scientists have to deal with large and complex datasets and work with data coming from diverse scientific areas. Artificial intelligence (AI), i.e., machines and algorithms that exhibit intelligent behavior, plays a significant role in Data Science. Intelligent machines can help to collect, store, search, process and reason over both data and knowledge. There are two main approaches to AI: statistical and symbolic. For a long time,

a dominant approach to AI was based on symbolic representations and treating "intelligence" or intelligent behavior primarily as symbol manipulation. In a physical symbol systems [35], physical entities (tokens, symbols) stand for, or denote, entities, are combined with other symbols to form complex symbol structures, and are manipulated by processes. Arguably, human communication occurs through symbols (words and sentences), and human thought – on a cognitive level – also occurs symbolically, so that symbolic AI resembles human cognitive behavior. Symbolic approaches are useful to *represent* theories or laws in a way that is meaningful to the symbol system and can be meaningful to humans; they are also useful in producing new symbols through symbol manipulation or inference rules. An alternative (or complementary) approach to AI are statistical approaches in which intelligence is taken as an emergent behavior of a system. Prominently, connectionist systems [34], in particular artificial neural networks [44], have gained influence in the past decade with computational and methodological advances driving new applications [32]. Statistical approaches are useful in *learning* patterns or regularities from data, and as such have a natural application within Data Science. Advancements in computational power, data storage, and parallelization, in combination to methodological advances in applying machine learning algorithms, are contributing to the uptake of statistical approaches in recent years [32], and these approaches have moved areas such as visual processing, object recognition in images, video labeling by sensory systems, and speech recognition by Natural Language Processing (NLP) significantly forward.

On the other hand, a large number of knowledge bases, knowledge graphs and ontologies are generated to explicitly capture the knowledge within a domain. Reasoning over these knowledge bases allows consistency checking (i.e., detecting contradictions between facts or statements), classification (i.e., generating taxonomies), and other forms of deductive inference (i.e., revealing new, implicit knowledge given a set of facts). In discovering knowledge from data, the knowledge about the problem domain and additional constraints that a solution will have to satisfy can significantly improve the chances of finding a good solution or determining whether a solution exists at all. Integrative approaches to Data Science utilize data analysis methods together with analysis of structured knowledge. Knowledge-based methods can also be used to combine data from different domains, different phenomena, or different modes of representation, and *link* data together to form a linked Web of data [11]. In Data Science, methods that exploit the semantics of knowledge graphs and Semantic Web technologies [7] as a way to add background knowledge to machine learning models have started to emerge. For example, neural tensor networks [50] and neural theorem provers [43] can be applied to learn representation of words and concepts in an unsupervised way while utilizing formalized background knowledge available in a knowledge base, and knowledge graph embeddings [36] generate low-dimensional representations of entities in a knowledge grpah.

The Life Sciences have been one of the key drivers behind progress in artificial intelligence, and the vastly increasing volume and complexity of data in biology is one of the drivers in Data Science as well. In computational biology, Semantic Web technologies such as knowledge graphs and ontologies are especially widely applied to represent and integrate data [49,10,28], and similarly, Life Sciences are a prime application area for novel machine learning methods [3,41]. Here, we start from the perspective of Life Science researchers and explore the role of symbolic representations in Data Science.

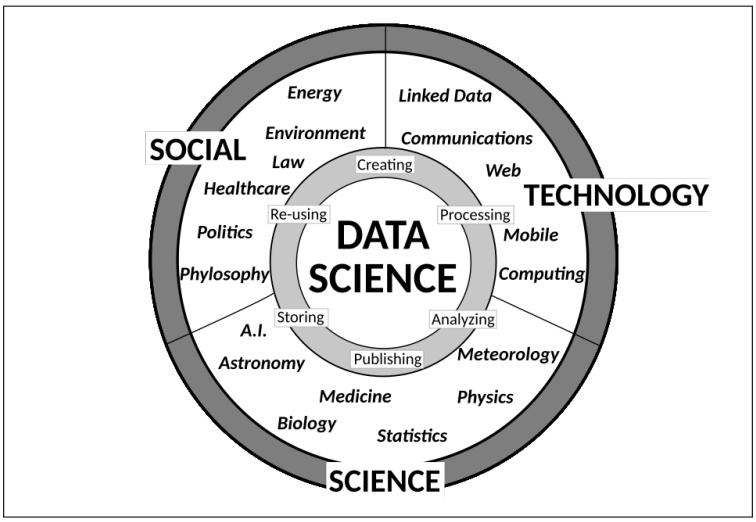

Figure 1. Data Science as the core intersection of other disciplines.

2. Knowledge as data

Not all data that a data scientist will be faced with consists of raw, unstructured measurements. In many cases, data comes as structured, symbolic representation with (formal) semantics attached, i.e., the knowledge within a domain. In these cases, the aim of the data scientist is to apply the methods of Data Science to *knowledge* about a domain itself. This can be the case when analyzing natural language text, but more so in the analysis of structured data coming from databases and knowledge bases. In particular in biology and biomedicine, ontologies, symbolic representations of a conceptualization of a domain [18], are widely used to annotate, integrate and analyze data [24]. Many bio-ontologies are specified in the Web Ontology Language (OWL) [17] and come with a model-theoretic semantics and syntactic inference rules [25]. Similarly, the Resource Description Framework (RDF) [6] is used very widely in biology and biomedicine, but also in many other domains, and gives rise to rule-based inference. It is a challenge for Data Science (and machine learning) to utilize the explicit semantics and inference rules of a structured, semantically represented data set in data analysis. This challenge is even more pronounced as Data Science often relies heavily on statistics, but does not generally emphasize symbolic representations and the role of symbol manipulation.

Several approaches that apply machine learning techniques to structured information, in particular to graphs, have been developed [40,36,38,13]. Of particular interest to symbolic AI are *knowledge graphs*. A knowledge graph consists of entities and concepts represented as nodes, and edges of different types that connect these nodes. From the perspective of Data Science and machine learning, knowledge graphs are a form of directed graph with both edge and node labels. Knowledge graphs store information about entities and their interrelations, and can be used to improve search, generate explanations, or plans to move a system from one state to another. To analyze and learn from knowledge graphs, approaches were developed based on word embeddings generated through

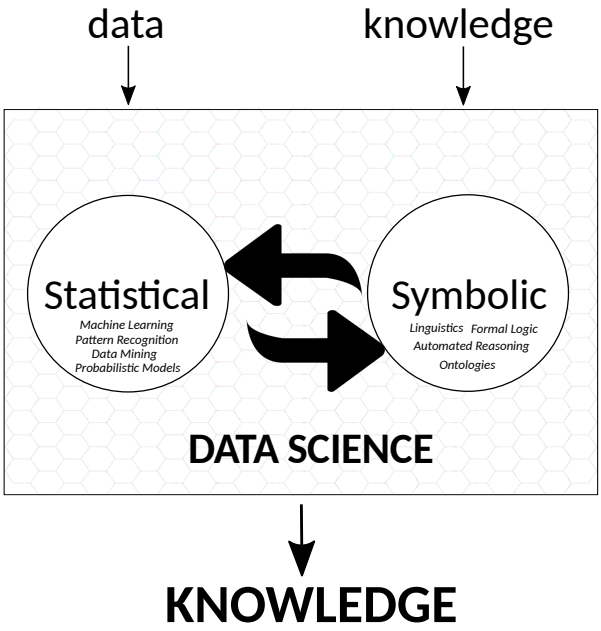

Figure 2. Data Science as a discipline that transforms data into knowledge. We explicitly mark "knowledge" as an input – i.e., subject matter – of Data Science in addition to "data"; knowledge can be used as background knowledge about the problem domain, to determine if an interpretation of data is consistent with certain assumptions, or Data Science can treat knowledge as data for its analyses.

random walks [42], methods based on tensor factorization [29,13,37], and a wide range of semantic similarity measures [20]. Major applications of these approaches are link prediction (i.e., predicting missing edges between the entities in a knowledge graph), clustering, or similarity-based analysis and recommendation.

In the Semantic Web [7], knowledge graphs consist of two parts: RDF-structured data (domain knowledge) and OWL-structured ontologies (conceptual knowledge), or, in Description Logic terminology, ABox and TBox, respectively [4]. While qualitative domain data can be represented in the form of an RDF graph, knowledge is usually expressed through different types of axioms, only some of which (such as subclass axioms) may give rise to a graph structure [48]. For example, by stating that a biological function F_1 is a subclass of F_2 and asserting that protein P has the function F_1, it is already inherent in the semantics of OWL that P also has the function F_2. Representing these axioms and inferences within a graph is not trivial, as axioms can be arbitrarily complex (within the constraints of the language). One option is to consider axiom schemata with two free variables for nodes within the knowledge graph as defining an edge type, and using an automated reasoner to determine whether the axiom can be inferred from the knowledge graph. For example, a *part-of* edge could be defined through the axiom pattern `?X SubClassOf: part-of some ?Y`, and for all pairs of entities (e_1, e_2) within a knowledge graph $\mathscr{KG}$, an edge labeled *part-of* added between e_1 and e_2 if $\mathscr{KG} \models (e_1 \text{ SubClassOf: part-of some } e_2)$. The OBO Flatfile Format and many biomedical ontologies utilize this definition pattern for edges [23,48]. For model-theoretic languages, it is also possible to analyze model structures instead of syntactically induced graph representations. While there are usually infinitely many models of arbitrary cardi-

nality [47], it is possible to focus on special (canonical) models in some languages such as the Description Logics *ALC*. These model structures can then be analyzed instead of syntactically formed graphs, and for example used to define semantic similarity measures [12]. Ultimately, there is a need to build more hybrid algorithms that specifically account for the semantics of (graph-)structured data and distinguishes it from axioms expressed in formal, model-theoretic languages such as Description Logics or first-order logic.

3. Turning data into knowledge

In the opposite direction, methods from Data Science can also be used to directly generate symbolic representations. For example, statistical approaches in Data Science can learn to recognize patterns, and recently there has been a big success in pattern recognition and unsupervised feature learning using neural networks [32]. Feature learning (or deep learning) methods can identify patterns and regularities within a domain and thereby learn the "conceptualizations" of a domain. An explicit, formal representations of a conceptualization is an ontology [19,18], and it is an enticing possibility to use methods from Data Science and statistical AI to automatically learn these formal representations. This problem is closely related to the symbol grounding problem, i.e., the problem of how symbols obtain their meaning [21]. Traditional approaches to learning formal representations of concepts from a set of facts include inductive logic programming [9] or rule learning methods [2,33] which find axioms that characterize the data. Recent feature learning techniques generate distributed representations [22] that represent regularities in raw data implicitly and can be used to identify instances of a pattern in data. But, they are not yet symbolic representations (in particular, they are neither directly interpretable nor can they be combined to form more complex representations). Hence, there is a gap remaining between the distributed representations generated by neural networks and symbolic representations (and reasoning) [5]; closing this gap remains a major challenge for statistical AI and Data Science.

Recent approaches towards solving these challenges include representing symbol manipulation as operations performed by neural network [43,52], thereby enabling symbolic inference with distributed representations grounded in domain data. Notably, in biology and biomedicine, where large volumes of experimental data are available, several methods have been developed to generate ontologies in a data-driven manner from high-throughput datasets [31,14,16]. These rely on data-driven generation of concepts (through clustering of networks) and using ontology mapping techniques [26] to align these clusters to ontology classes. Other methods rely, for example, on recurrent neural networks that can combine distributed representations [51,15]. However, the full "neuro-symbolic" cycle from learning symbolic representations from data, manipulating and combining the symbols into more complex symbol structures, and providing results back to classification of data is still unsolved.

4. Limits of Data Science

It is also important to identify fundamental limits for any statistical, data-driven approach with regard to the scientific knowledge it can possibly generate. Some important domain

concepts simply cannot be learned from data alone. For example, the set of Gödel numbers for halting Turing machines can, arguably, not be "learned" from data or derived statistically, although the set can be characterized symbolically. Furthermore, many empirical laws cannot simply be derived from data because they are idealizations that are never actually observed in nature; examples of such laws include Galileo's principle of inertia, Boyle's gas Law, zero-gravity, point mass, friction-less motion, etc. [39]. Although these concepts and laws cannot be observed, they form some of the most valuable and predictive components of scientific knowledge. To derive such laws as general principles from data, an addition creative step seems to be required that abstracts from observations to scientific laws. This step relates to our human cognitive ability of making idealizations, and has early been described as necessary for scientific research by philosophers such as Husserl [27] or Ingarden [1].

Inspired by progress in Data Science and statistical methods in AI, Kitano [30] proposed a new Grand Challenge for AI "to develop an AI system that can make major scientific discoveries in biomedical sciences and that is worthy of a Nobel Prize". A more tangible challenge may be to design an algorithm that can identify the principle of inertia, given unlimited data about moving objects and their trajectory over time and all the knowledge Galileo had about mathematics and physics in the 17th century. This is a task that Data Science should be able to solve, which relies on the analysis of large ("Big") datasets, and for which almost infinite data points can be generated.

One of Galileo's key contributions was to realize that laws of nature are inherently mathematical and expressed symbolically, and to identify symbols that stand for force, objects, mass, motion, and velocity, ground these symbols in perceptions of phenomena in the world. This task may be achievable through feature learning, or ontology learning, methods. Additionally, the symbols need to be combined in a way to express the principle of inertia. However, given sufficient data about moving objects on Earth, any algorithm will likely come up with Aristotle's theory of motion [45] instead, not Galileo's principle of inertia. On a high level, Aristotle's theory of motion states that all things come to a rest, heavy things on the ground and lighter things on the sky, and force is required to move objects. It was only when a more fundamental understanding of objects outside of Earth became available through the observations of Kepler and Galileo that this theory on motion no longer yielded useful results.

The challenges were (1) to identify that motion processes observed on Earth and the motion observed at stellar objects are essentially instances of the same concept (i.e., "motion"), (2) to identify the inconsistency between the established theory on motion and the data derived from repeated observations (of moving stellar objects), and (3) finding a theory that was more comprehensive and predictive of both phenomena as well as supported by experimental evidence (data) in both domains or areas of observation. Identifying the inconsistencies is a symbolic process in which deduction is applied to the observed data and a contradiction identified. Generating a new, more comprehensive, theory, i.e., the principle of inertia, is a creative process, with the additional difficulty that not a single instance of that theory could have been observed (because we know of no objects on which no force acts). Generating such a theory in the absence of a single supporting instance is the real grand challenge!

If we ever wish to build machines that can "discover" natural laws from data and observations, we will need a revolution similar to the scientific revolution in the 16th and 17th century that resulted in the creation of the scientific method and our modern

understanding of natural science. Data Science, due to its interdisciplinary nature and as the scientific discipline that has as its subject matter the question of how to turn data into knowledge will be the best candidate for a field from which such a revolution will originate.

References

[1] *Gesammelte Werk; Band 7: Zur Grundlegung Der Erkenntnistheorie, Teil 1*. Walter de Gruyter, 1996.

[2] Rakesh Agrawal, Tomasz Imieliński, and Arun Swami. Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, SIGMOD '93, pages 207–216, New York, NY, USA, 1993. ACM.

[3] Christof Angermueller, Tanel Pärnamaa, Leopold Parts, and Oliver Stegle. Deep learning for computational biology. *Molecular Systems Biology*, 12(7), 2016.

[4] Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter Patel-Schneider. *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press, 2003.

[5] Sebastian Bader and Pascal Hitzler. Dimensions of neural-symbolic integration - a structured survey. In Sergei N. Artmov, Howard Barringer, Artur S. d'Avila Garcez, Lus C. Lamb, and John Woods, editors, *We Will Show Them! (1)*, pages 167–194. College Publications, 2005.

[6] Dave Beckett. RDF/XML syntax specification (revised). W3C recommendation, World Wide Web Consortium (W3C), February, October 2004.

[7] T. Berners-Lee, J. Hendler, O. Lassila, et al. The Semantic Web. *Scientific American*, 284(5):28–37, 2001.

[8] David Brown. *Mesopotamian planetary astronomy-astrology*. Styx, Groningen, 2000.

[9] Lorenz Bhmann, Jens Lehmann, and Patrick Westphal. Dl-learnera framework for inductive learning on the semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 39:15 – 24, 2016.

[10] Alison Callahan, José Cruz-Toledo, Peter Ansell, and Michel Dumontier. *Bio2RDF Release 2: Improved Coverage, Interoperability and Provenance of Life Science Linked Data*, pages 200–212. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[11] Tom Heath Christian Bizer and Tim Berners-Lee. Linked data – the story so far. *International Journal on Semantic Web and Information Systems*, 2009. in press.

[12] Claudia d'Amato, Nicola Fanizzi, and Floriana Esposito. A semantic similarity measure for expressive description logics. *CoRR*, abs/0911.5043, 2009.

[13] Lucas Drumond, Steffen Rendle, and Lars Schmidt-Thieme. Predicting rdf triples in incomplete knowledge bases with tensor factorization. In *Proceedings of the 27th Annual ACM Symposium on Applied Computing*, SAC '12, pages 326–331, New York, NY, USA, 2012. ACM.

[14] Janusz Dutkowski, Michael Kramer, Michal A Surma, Rama Balakrishnan, J Michael Cherry, Nevan J Krogan, and Trey Ideker. A gene ontology inferred from molecular networks. *Nature Biotechnology*, 31:38–45, 2012.

[15] L. Ferrone and F. Massimo Zanzotto. Symbolic, Distributed and Distributional Representations for Natural Language Processing in the Era of Deep Learning: a Survey. *ArXiv e-prints*, February 2017.

[16] Vladimir Gligorijevi, Vuk Janji, and Nataa Prulj. Integration of molecular network data reconstructs gene ontology. *Bioinformatics*, 30(17):i594, 2014.

[17] B. Grau, I. Horrocks, B. Motik, B. Parsia, P. Patelschneider, and U. Sattler. OWL 2: The next step for OWL. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(4):309–322, November 2008.

[18] Thomas R. Gruber. Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, 43(5-6), 1995.

[19] Nicola Guarino. Formal ontology and information systems. In Nicola Guarino, editor, *Proceedings of the 1st International Conference on Formal Ontologies in Information Systems*, pages 3–15. IOS Press, 1998.

[20] Sbastien Harispe, Sylvie Ranwez, Stefan Janaqi, and Jacky Montmain. Semantic similarity from natural language and ontology analysis. *Synthesis Lectures on Human Language Technologies*, 8(1):1–254, 2015.

[21] Stevan Harnad. The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42:335–346, 1990.

[22] G. E. Hinton, J. L. McClelland, and D. E. Rumelhart. Parallel distributed processing: Explorations in the microstructure of cognition, vol. 1. chapter Distributed Representations, pages 77–109. MIT Press, Cambridge, MA, USA, 1986.

[23] Robert Hoehndorf, Anika Oellrich, Michel Dumontier, Janet Kelso, Dietrich Rebholz-Schuhmann, and Heinrich Herre. Relations as patterns: Bridging the gap between OBO and OWL. *BMC Bioinformatics*, 11(1):441+, 2010.

[24] Robert Hoehndorf, Paul N. Schofield, and Georgios V. Gkoutos. The role of ontologies in biological and biomedical research: a functional perspective. *Briefings in Bioinformatics*, March 2015.

[25] Ian Horrocks, Oliver Kutz, and Ulrike Sattler. The even more irresistible sroiq. In Patrick Doherty, John Mylopoulos, and Christopher A. Welty, editors, *KR*, pages 57–67. AAAI Press, 2006.

[26] L. Huang, G. Hu, and X. Yang. Review of ontology mapping. In *2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet)*, pages 537–540, April 2012.

[27] Edmund Husserl and W. Biemel. *Die Krisis der Europischen Wissenschaften und die Transzendentale Phnomenologie*. Springer Netherlands, 1 edition, 1976.

[28] Simon Jupp, James Malone, Jerven Bolleman, Marco Brandizi, Mark Davies, Leyla Garcia, Anna Gaulton, Sebastien Gehant, Camille Laibe, Nicole Redaschi, Sarala M. Wimalaratne, Maria Martin, Nicolas Le Novre, Helen Parkinson, Ewan Birney, and Andrew M. Jenkinson. The EBI RDF platform: linked open data for the life sciences. *Bioinformatics*, 30(9):1338–1339, 2014.

[29] Suleiman A. Khan, Eemeli Leppäaho, and Samuel Kaski. Bayesian multi-tensor factorization. *Machine Learning*, 105(2):233–253, 2016.

[30] Hiroaki Kitano. Artificial intelligence to win the nobel prize and beyond: Creating the engine for scientific discovery. *AI Magazine*, 37(1), 2016.

[31] Michael Kramer, Janusz Dutkowski, Michael Yu, Vineet Bafna, and Trey Ideker. Inferring gene ontologies from pairwise similarity data. *Bioinformatics*, 30(12):i34, 2014.

[32] Yann Lecun et al. Deep learning. *Nature*, 521(7553):436–444, 5 2015.

[33] H. Lu, R. Setiono, and H. Liu. NeuroRule: A Connectionist Approach to Data Mining. *ArXiv e-prints*, January 2017.

[34] Marvin Minsky. Logical versus analogical or symbolic versus connectionist or neat versus scruffy. *AI Mag.*, 12(2):34–51, April 1991.

[35] Allen Newell and Herbert A. Simon. Computer science as empirical inquiry: Symbols and search. *Commun. ACM*, 19(3):113–126, March 1976.

[36] M. Nickel, K. Murphy, V. Tresp, and E. Gabrilovich. A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, 104(1):11–33, Jan 2016.

[37] Maximilian Nickel, Lorenzo Rosasco, and Tomaso Poggio. Holographic embeddings of knowledge graphs. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI'16, pages 1955–1961. AAAI Press, 2016.

[38] Maximilian Nickel, Volker Tresp, and Hans peter Kriegel. A three-way model for collective learning on multi-relational data. In Lise Getoor and Tobias Scheffer, editors, *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 809–816, New York, NY, USA, 2011. ACM.

[39] Leszek Nowak. *Idealization VII: Structuralism, Idealization and Approximation*, chapter Remarks on the nature of Galileo's methodological revolution. Martti Kuokkanen, 1994.

[40] Bryan Perozzi et al. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, pages 701–710, New York, NY, USA, 2014. ACM.

[41] D. Rav, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G. Z. Yang. Deep learning for health informatics. *IEEE Journal of Biomedical and Health Informatics*, 21(1):4–21, Jan 2017.

[42] Petar Ristoski and Heiko Paulheim. Rdf2vec: Rdf graph embeddings for data mining. In *The Semantic Web - ISWC 2016 : 15th International Semantic Web Conference, Kobe, Japan, October 17-21, 2016, Proceedings, Part I*, volume 9981, pages 498–514, Cham, 2016. Springer International Publishing.

[43] Tim Rocktäschel and Sebastian Riedel. Learning knowledge base inference with neural theorem provers. In *Proceedings of the 5th Workshop on Automated Knowledge Base Construction, AKBC@NAACL-HLT 2016, San Diego, CA, USA, June 17, 2016*, pages 45–50, 2016.

[44] David E. Rumelhart, James L. McClelland, and CORPORATE PDP Research Group, editors. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations*. MIT Press, Cambridge, MA, USA, 1986.

[45] Joe Sachs. *Aristotle's Physics: A Guided Study*. Rutgers University Press, 1 edition, 1995.

[46] Subrata Saha and Sanguthevar Rajasekaran. Ergc: an efficient referential genome compression algorithm. *Bioinformatics*, 31(21):3468, 2015.

[47] T.A. Skolem. *Ueber einige Grundlagenfragen der Mathematik*. Skrifter utgitt av det Norske videnskaps-akademi i Oslo. 1, Matematisk-naturvidenskapelig klasse. Dybwad, 1929.

[48] B. Smith, W. Ceusters, B. Klagges, J. Köhler, A. Kumar, J. Lomax, C. Mungall, F. Neuhaus, A. L. Rector, and C. Rosse. Relations in biomedical ontologies. *Genome Biol*, 6(5):R46, 2005.

[49] Barry Smith, Michael Ashburner, Cornelius Rosse, Jonathan Bard, William Bug, Werner Ceusters, Louis J. Goldberg, Karen Eilbeck, Amelia Ireland, Christopher J. Mungall, Neocles Leontis, Philippe R. Serra, Alan Ruttenberg, Susanna A. Sansone, Richard H. Scheuermann, Nigam Shah, Patricia L. Whetzel, and Suzanna Lewis. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotech*, 25(11):1251–1255, 2007.

[50] Richard Socher, Danqi Chen, Christopher D. Manning, and Andrew Y. Ng. Reasoning With Neural Tensor Networks For Knowledge Base Completion. In *Advances in Neural Information Processing Systems 26*. 2013.

[51] Richard Socher, Brody Huval, Christopher D. Manning, and Andrew Y. Ng. Semantic compositionality through recursive matrix-vector spaces. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, EMNLP-CoNLL '12, pages 1201–1211, Stroudsburg, PA, USA, 2012. Association for Computational Linguistics.

[52] Daniel Whalen. Holophrasm: a neural automated theorem prover for higher-order logic. *CoRR*, abs/1608.02644, 2016.