Springer Link

Chapter

Machine Learning, Optimization, and Big Data

Volume 9432 of the series *Lecture Notes in Computer Science* pp 306-317

Date: 06 January 2016

Conceptual Analysis of Big Data Using Ontologies and EER

- Kulsawasd Jitkajornwanich
- , Ramez Elmasri

Abstract

Large amounts of “big data” are generated every day, many in a “raw” format that is difficult to analyze and mine. This data contains potential hidden meaningful concepts, but much of the data is superfluous and not of interest to the domain experts. Thus, dealing with big raw data solely by applying a set of distributed computing technologies (e.g., MapReduce, BSP [Bulk Synchronous Parallel], and Spark) and/or distributed storage systems, namely NoSQL, is generally not sufficient. Extracting the full knowledge that is hidden in the raw data is necessary to efficiently enable analysis and mining. The data needs to be processed to remove the superfluous parts and generate the meaningful domain-specific concepts. In this paper, we propose a framework that incorporates conceptual modeling and EER principle to effectively extract conceptual knowledge from the raw data so that mining and analysis can be applied to the extracted conceptual data.

Keywords

Conceptual modeling Big data NoSQL Distributed computing

References

1. Embley, D.W., Liddle, S.W.: Big data—conceptual modeling to the rescue. In: 32nd International Conference on Conceptual Modeling (2013)

2. Dean, J., Ghemawat, S.: MapReduce: simplified data processing on large clusters. In: 6th Symposium on Operating Systems Design and Implementation (2004)

3. Valiant, L.G.: A bridging model for multi-core computing. In: 16th Annual European Symposium (2008)

4. Apache. Apache Spark™. *http://spark.apache.org* *(http://spark.apache.org)*

5. Zou, B., Ma, X., Kemme, B., Newton, G., Precup, D.: Data mining using relational database management systems. In: 10th Pacific-Asia Conference (2006)

6. Lam, C.: Hadoop in Action. Dreamtech Press, New Delhi (2011)

7. Edlich, S.: List of NOSQL Databases. *http://nosql-database.org* *(http://nosql-database.org)*

8. Amazon. Amazon DynamoDB. *http://aws.amazon.com/dynamodb* *(http://aws.amazon.com/dynamodb)*

9. MongoDB. *http://www.mongodb.org* *(http://www.mongodb.org)*

10. Jitkajornwanich, K., Elmasri, R., Li, C., McEnery, J.: Extracting storm-centric characteristics from raw rainfall data for storm analysis and mining. In: 1st ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data (2012)

11. Jitkajornwanich, K., Gupta, U., Elmasri, R., Fegaras, L., McEnery, J.: Using mapreduce to speed up storm identification from big raw rainfall data. In: 4th International Conference on Cloud Computing, GRIDs, and Virtualization (2013)

12. Jitkajornwanich, K., Gupta, U., Shanmuganathan, S.K., Elmasri, R., Fegaras, L., McEnery, J.: Complete storm identification algorithms from big raw rainfall data. In: 2013 IEEE International Conference on Big Data (2013)

13. Overeem, A., Buishand, A., Holleman, I.: Rainfall depth-duration-frequency curves and their uncertainties. J. Hydrol. **348**, 124–134 (2008)
CrossRef *(http://dx.doi.org/10.1016/j.jhydrol.2007.09.044)*

14. Elmasri, R., Navathe, S.: Fundamentals of Database Systems, 6th edn. Pearson Education, New Delhi (2010)

15. Asquith, W.H., Roussel, M.C., Cleveland, T.G., Fang, X., Thompson, D.B.: Statistical characteristics of storm interevent time, depth, and duration for eastern New Mexico, Oklahoma, and Texas. Professional Paper 1725, US Geological Survey (2006)

16. Lanning-Rush, J., Asquith, W.H., Slade, Jr., R.M.: Extreme precipitation depth for Texas, excluding the trans-pecos region. Water-Resources Investigations Report 98–4099, US Geological Survey (1998)

17. NOAA's national weather service. The XMRG File Format and Sample Codes to Read XMRG Files. *http://www.nws.noaa.gov/oh/hrl/dmip/2/xmrgformat.html* *(http://www.nws.noaa.gov/oh/hrl/dmip/2/xmrgformat.html)*

18. Consortium of universities for the advancement of hydrologic science, Inc. (CUAHSI). ODM Databases. *http://his.cuahsi.org/odmdatabases.html* *(http://his.cuahsi.org/odmdatabases.html)*

19. Asquith, W.H.: Depth-duration frequency of precipitation for Texas. Water-Resources Investigations Report 98–4044, US Geological Survey (1998)

20. Asquith, W.H.: Summary of dimensionless Texas hyetographs and distribution of storm depth developed for texas department of transportation research project 0–4194. Report 0–4194-4, US Geological Survey (2005)

21. National Oceanic and Atmospheric Administration (NOAA). National Weather Service River Forecast Center: West Gulf RFC (NWS-WGRFC). *http://www.srh.noaa.gov/wgrfc*

(http://www.srh.noaa.gov/wgrfc)

22. Unidata. What is the LDM? *https://www.unidata.ucar.edu/software/ldm/ldm-6.6.5/tutor-ial/whatis.html (https://www.unidata.ucar.edu/software/ldm/ldm-6.6.5/tutor-ial/whatis.html)*

23. Chang, F., Dean, J., Ghemawat, S., Hsieh, W.C., Wallach, D.A., Burrows, M., Chandra, T., Fikes, A., Gruber, R.E.: Bigtable: a distributed storage system for structured data. In: 7th USENIX Symposium on Operating Systems Design and Implementation (2006)

24. NOAA. MPE: Multisensor Precipitation Estimate. *http://www.erh.noaa.gov/marfc/Maps/xmrg/index_java.html (http://www.erh.noaa.gov/marfc/Maps/xmrg/index_java.html)*

25. Mishra, S.K., Singh, V.P.: Soil Conservation Service Curve Number (SCS-CN) Methodology. Kluwer Academic Publishers, Boston (2003) *CrossRef (http://dx.doi.org/10.1007/978-94-017-0147-1)*

26. Jitkajornwanich, K.: Analysis and modeling techniques for geo-spatial and spatio-temporal datasets. Doctoral Dissertation, The University of Texas at Arlington (2014)

27. Cheng, T., Haworth, J., Anbaroglu, B., Tanaksaranond, G., Wang, J.: Spatio-Temporal Data Mining. Handbook of Regional Science. Springer, Heidelberg (2013)

28. IBM Big Data and Analytics Hub. Understanding Big Data: e-book. *http://www.ibmbigdatahub.com/whitepaper/understanding-big-data-e-book (http://www.ibmbigdatahub.com/whitepaper/understanding-big-data-e-book)*

29. Jin, R. NoSQL and Big Data Processing: Hbase, Hive and Pig, etc. *http://www.cs.kent.edu/~jin/Cloud12Spring/HbaseHivePig.pptx (http://www.cs.kent.edu/%7ejin/Cloud12Spring/HbaseHivePig.pptx)*

30. Widom, J. NoSQL Systems: Overview. *http://openclassroom.stanford.edu/Main-Folder/courses/cs145/old-site/docs/slides/NoSQLOverview/annotated.pptx (http://openclassroom.stanford.edu/Main-Folder/courses/cs145/old-site/docs/slides/NoSQLOverview/annotated.pptx)*

31. World Wide Web Consortium (W3C). OWL Web Ontology Language Guide. *http://www.w3.org/TR/owl-guide/ (http://www.w3.org/TR/owl-guide/)*

About this Chapter

Title

Conceptual Analysis of Big Data Using Ontologies and EER

Book Title

Machine Learning, Optimization, and Big Data

Book Subtitle

First International Workshop, MOD 2015, Taormina, Sicily, Italy, July 21-23, 2015, Revised Selected Papers

Pages

pp 306-317

DOI

10.1007/978-3-319-27926-8_27

Print ISBN
978-3-319-27925-1
Online ISBN
978-3-319-27926-8
Series Title
Lecture Notes in Computer Science
Series Volume
9432
Series ISSN
0302-9743
Publisher
Springer International Publishing
Copyright Holder
Springer International Publishing Switzerland
Additional Links

- *About this Book*

Topics

- *Information Systems Applications (incl. Internet)*
- *Algorithm Analysis and Problem Complexity*
- *Artificial Intelligence (incl. Robotics)*
- *Computation by Abstract Devices*
- *Database Management*
- *Information Storage and Retrieval*

Keywords

- Conceptual modeling
- Big data
- NoSQL
- Distributed computing

Industry Sectors

- *Pharma*
- *Automotive*
- *Biotechnology*
- *Electronics*
- *IT & Software*
- *Telecommunications*
- *Consumer Packaged Goods*
- *Aerospace*
- *Oil, Gas & Geosciences*
- *Engineering*

Editors

- *Panos Pardalos* ✉ (13)
- *Mario Pavone* ✉ (14)
- *Giovanni Maria Farinella* ✉ (15)
- *Vincenzo Cutello* ✉ (16)

Editor Affiliations

- 13. University of Florida
- 14. University of Catania
- 15. University of Catania
- 16. University of Catania

Authors

- *Kulsawasd Jitkajornwanich* (17)
- *Ramez Elmasri* (18)

Author Affiliations

- 17. Geo-Informatics and Space Technology Development Agency (Public Organization), Ministry of Science and Technology of Thailand, 120 The Government Complex, Chaeng Wattana Road, Lak Si, Bangkok, 10210, Thailand
- 18. Department of Computer Science and Engineering, The University of Texas at Arlington, 701 S Nedderman Dr, Arlington, TX, 76019, USA