Forecasting Informatics Competencies for Nurses in the Future of Connected Health
J. Murphy et al. (Eds.)

doi:10.3233/978-1-61499-738-2-165

Big Data and Nursing: Implications for the Future

Maxim TOPAZ[a] and Lisiane PRUINELLI[b]
[a] *Harvard Medical School & Brigham Women's Health Hospital, Boston, MA, USA*
[b] *School of Nursing, University of Minnesota, Minneapolis, MN, USA*

Abstract. Big data is becoming increasingly more prevalent and it affects the way nurses learn, practice, conduct research and develop policy. The discipline of nursing needs to maximize the benefits of big data to advance the vision of promoting human health and wellbeing. However, current practicing nurses, educators and nurse scientists often lack the required skills and competencies necessary for meaningful use of big data. Some of the key skills for further development include the ability to mine narrative and structured data for new care or outcome patterns, effective data visualization techniques, and further integration of nursing sensitive data into artificial intelligence systems for better clinical decision support. We provide growth-path vision recommendations for big data competencies for practicing nurses, nurse educators, researchers, and policy makers to help prepare the next generation of nurses and improve patient outcomes trough better quality connected health.

Keywords. Big data, Data science, Nursing, Natural language processing, Data mining, Data visualization

1. Introduction

The discipline of nursing is charged with several core missions, including: protecting, promoting and optimizing human health and ability; alleviating suffering by diagnosing and treating the human response; advocating for individuals, families, communities, and populations; and preventing injury and illness [1]. Nursing informatics (NI) incorporates the science of nursing with information, technological, communication and analytical sciences to support the integration of data, information, knowledge, and wisdom into the provision of evidence-based nursing care [1].

The concept of big data became widespread around 2010 [2] and is often defined by the "4Vs": *Volume, Velocity, Variety, and Veracity* [3]. *Volume* refers to the large amount of data, e,g., millions of patients records or detailed genomic data. *Velocity* refers to the rate of high-frequency, real-time generated data, such as data from smartphones and sensors devices. *Variety* refers to the heterogeneity of the data, e.g., structured, semi-structured, and unstructured data, such as electronic health records (EHRs), monitoring devices, genomics, sensors, imaging, claims data, social media, patient generated data, and others real time data. *Veracity* is the uncertainty of the data, either in terms of accuracy of the data for the original purpose for which the data were collected or appropriateness of the data for secondary use. Recently, an additional fifth "V" was suggested for *Value*, representing the value of information extracted from the data leading to knowledge discovery [4, 5].

The scholarship of working with data is known as data science and it draws on knowledge and theories from multiple disciplines, including nursing. Big data science uses a variety of methods for analyzing data from traditional statistics to visualization techniques, data mining, and natural language processing [3]. Use of big data is assuming a critical role in healthcare analytics as the complexity and variety of data available are increasing. In addition, advanced computational methods and tools to analyze big data are increasingly more available. However, nursing relevant big data applications and research are still in their infancy, often because they require a rare combination of multidisciplinary skills in computer science, statistics, mathematics, and health informatics. There is an urgent need in further understanding and recommendations on what nurses need to learn in order to be competent in data science and analytics. Once NI research is able to incorporate data that represent the "5Vs" into research, education and practice, nurses have the potential to achieve better outcomes for populations and individuals.

This chapter aims to identify competencies necessary for big data management and analytics, and provide recommendations concerning policy, educational, research, and practice needs related to big data. We understand that some of the identified competencies might present a challenge for practicing nurses and educators today; nevertheless, we strongly believe that our recommendations will provide a growth-path vision for nurses of the near future practicing in the realm of connected health.

2. Big data types and data processing steps

Health data is the raw and uninterpreted object that contains health attributes or characteristics [6]. In order to manage data effectively, nurses need to understand several basic concepts presented below:

a) *Types of data*: data can be divided in two generic types: qualitative (or categorical) and quantitative (or numeric). Commonly, qualitative data can be represented as nominal/ ordinal/ or free-text values whereas quantitative data can have interval/ratio/dichotomous or discrete values. Health data are often highly heterogeneous, such as electronic health records data that include lab values, narrative notes, diverse time points and large number of medical abbreviations. This data is also often quite messy with missing values, noise, outliers, inconsistencies, duplicate concepts, and might contain other types of human or system errors (e.g., misspellings or erroneously recorded time points). Thus, knowledge domain expertise needed to evaluate the quality of health data.

b) *Data processing steps*: an individual or an organization using health data need to have a clear understanding of the data use goals and the analytic methods they might want to apply. Then, the data needs to be pre-processed and prepared for the analysis. These steps may include: aggregation, sampling, dimensionality reduction, feature selection, feature creation, discretization, binarization, transformation and so on. Another import part is de-identification of data. With big data sets usually containing all health information from a person, additional security measures should be implemented to preserve patient's privacy and confidentiality. For instance, data about rare conditions may be discarded or grouped to prevent potential individual identification. A similar situation can occur if a specific condition is related to a postal code, leading to identification of an individual.

Nursing informaticians expertise, with the domain knowledge of health problems and implications of data for these problems, is crucial in driving more intelligent models where data can be used to improve health of populations through implementation of best evidence-based practices.

3. Common methods and technologies applied for big data with implications for nursing

3.1 Data mining

Data mining refers to a suite of statistical tools that enable knowledge discovery. These techniques help to identify novel patterns that might otherwise remain unknown in large data sets and diverse data types. Data mining techniques help to analyze large multidimensional datasets and identify relationships between different types of data. It is often used for data driven hypothesis generation, for example when novel relationships between diseases to abnormal lab findings are discovered. Data mining tasks are generally divided into two major categories: predictive modeling and descriptive tasks. In addition, data mining tasks can be supervised (e.g., when the outcome is known) or unsupervised (when there is no known outcome). An example of data mining in nursing can be found in Bowles et al. [7] where decision trees were used to identify factors associated with nurses post-acute care referrals. Another example is the Dey et al study [8] where discriminative pattern mining was used to discover patterns associated with improvement in homecare patient mobility.

Data mining is becoming increasingly more prevalent and those techniques are being applied in everyday applications, such as traffic detection, personalized advertisement, fraud detection, and many others. Several existing open access data mining applications, such as Weka [9] and KNIME [10], can be readily utilized by nurses to conduct data mining and knowledge discovery with big data.

3.2 Natural Language Processing

Natural language processing is concerned with creating automated approaches for understanding human language. It is an interdisciplinary field that combines linguistics, computer science and informatics. One of the main goals of health oriented natural language processing is to create computer algorithms capable of understanding different types of free text data. The different types of data include narrative, clinician generated texts from EHRs, social media data (e.g., Twitter, Facebook, online health forums), patient or family generated data, and research literature, among others. The increasing need in natural language processing is based on the fact that about 80% of the data in EHRs, and up to 100% in some other data sources, is stored as free text.

In the past several decades, a few health natural language processing systems were developed, mostly focused on working with texts from the medical domain. For example, an open source system called cTAKES can assist in extracting symptoms from physician narrative notes [11]. Another system called MTERMS can assist with capturing information about medication dosage, route, etc. in physician notes to enable improved medication reconciliation [12].

Although nurses' recorded clinical narratives are similar to data generates by other disciplines, e.g., medicine, there are several unique features that require specific

nursing-focused approaches. For example, when nursing narratives are analyzed, nursing interventions or nursing problems should be mapped to nursing standard terminologies (e.g., to the International Classification for Nursing Practice, ICNP). This mapping would allow further use of extracted data for scalable analytics or automated reasoning (e.g., clinical decision support). For example, when a computer algorithm is trained to identify patients with pressure ulcers (ICNP code 10015612) or self-management issues (ICNP code 10022155) in a large pool of clinical notes, the system can generate automated nursing specific treatment recommendations (e.g., facilitate adherence to regime, ICNP code 10036273) or reminders for health promotion (e.g., provide health promotion, ICNP code 10008776). For more information on the role of nursing terminologies please see this book's chapter addressing with semantics.

Unfortunately, there are only a very few natural language processing systems that focus explicitly on nursing data [13]. To analyze nursing-relevant narrative big data, nurses will need the support and close collaboration of other disciplines, such as computational linguistics and computer science and understand machine learning methods.

3.3 Artificial Intelligence

Artificial intelligence is often referred to as the theory and development of information systems able to perform tasks that usually require human intelligence, such as visual or speech recognition, decision-making, and language translation. In health related fields, artificial intelligence systems can be used to assist in data analytics, clinical decision making, and diagnostics, among others. Some examples of artificial intelligence systems in healthcare include esophageal image recognition to enable better intubation [14] or genetics based predictions [15].

In nursing, work on artificial intelligence based systems started in early 1980s when first rule based systems were proposed. For example, Ryan [16] described a development of an interesting system called COMMES (an acronym for Creighton On-line Multiple Medical Expert System). COMMES was an artificial intelligence based expert system simulating a professional consultant designed to assist with nursing decision making about a patient condition. The system had a robust knowledge base and an extensive hierarchy of nursing and medical diagnoses and treatment options. COMMES was envisioned to be used as an educational tool for nurses, capable of generating diagnosis oriented care protocols and testing nurses' knowledge. Similar work on nursing artificial intelligence systems was done by several other groups throughout the years [17-20].

However, most of the systems proposed so far were rule based systems where an expert would pre-define some logical order for the system to follow. This is in contrast to a more probability-based systems that can do reasoning on their own, as envisioned by artificial intelligence pioneers. Also, there is still a wide gap between the developed systems and nursing practice where such systems are not common. To create nursing-sensitive artificial intelligence systems of a new generation, there is a need in thoroughly developed and tested machine learning and natural language processing approaches, which are scarce. Other potential domains that need to be integrated to produce best artificial intelligence results are speech and image recognition. Seeing the trends of increasing use of artificial intelligence systems in healthcare, we expect significant advances in this field for nursing and nurses need to be ready to integrate nursing knowledge into the emerging systems.

3.4 Visualization

Information visualization is the study of visual representations of different data to assist in human cognition and decision making. Big data analytics and applications require appropriate visualizations to enable users to grasp the extent and sometimes the significance of big-data driven inferences. In healthcare, data visualization is often used to present patient specific (e.g., blood pressure levels over time) or population trends (geographical spread of zika virus over time). However, an increasing body of literature indicates that healthcare visualization approaches need better integration within the existing clinical systems. For example, a recent study evaluated the graphical displays of laboratory test results in eight EHRs in the U.S. None of the systems met all 11 quality of visualization criteria and the magnitude of deficiency was significant. One system even presented results in reverse chronological order [21]. Similarly, there were a few recent studies that indicated significant problems with EHR systems used by nurses in terms of visualizations and usability [22]. Traditionally, nurses are used to working with paper charts presenting trends in patient's condition, for example blood pressure, breathing, and heart rate. Big data approaches could potentially offer much insight and help interpret patient's vital signs compared to the population average, for example. To accomplish that, nurses working with big data will need to draw from other disciplines, such as human computer interaction and graphic design, to be able to come up with appropriate visualization approaches.

4. Big data related growth-path competency recommendations for nursing education, practice, research and policy

4.1 Education

Academic and inservice education programs in nursing should teach the concept of big data to students. Students need to be able to discuss big data characteristics and understand the increasing impact of using big data on clinical decision making. Ideally, students in academic nursing programs should understand the limitations and benefits of big data while practicing nurses should learn on how big data is used to shape everyday clinical practice. Nurse educators should also be knowledgeable about big data related concepts and their implications for students.

4.2 Practice

Practicing nurses should understand how to use big data to extract and apply clinically relevant evidence-based information in practice. Also, there is a need to be able to understand how to extract data from different formats and reuse it for quality improvement and better decision making.

4.3 Research

Nurse researchers need to be able to understand the big data characteristics and be able to collaborate effectively with interdisciplinary teams to use big data in their programs of research. They also need to learn about analytic tools and methods (e.g., data mining

and natural language processing) that assist in working with big data. More efforts should be spent on data quality validation and merging different data formats/types together for further use in clinical decision making.

4.4 Policy

Nursing policy makers should be able to use big data to visualize the effect of nursing on patient outcomes. Also, there is a need to start building comprehensive national and international data pools that are capable of storing different data types. The large data storages can help design nursing relevant policies that work (e.g., identify the ideal number of patients per nurse during the hospital shift to achieve best outcomes).

Table 1 provides an overall summary of growth-path vision recommendations for entry level and advanced level nurses.

Table 1: Big data growth-path vision competency recommendations for entry and advanced level nurses.

Entry level nursing competencies
Nurses should understand the types and sources of data captured by the information systems, including data quality, type (e.g., structured data, free text narratives, etc.)
Nurses should understand the importance and value of data for nursing care.
Nurses should develop and apply critical thinking when using the data in clinical decision making.
Nurses should be able to communicate with multidisciplinary teams to use data for enhanced clinical decision making.
Advanced level nursing competencies
Nurses should understand the big data characteristics and be able to collaborate effectively with interdisciplinary teams to use big data in their programs of research.
Nurses should learn about analytic tools and methods (e.g., data mining and natural language processing) that assist in working with big data.
Nurses should use data quality validation techniques to merge different data formats/types together for further use in clinical decision making.

5. Conclusion

Big data, with all its complexity, is becoming increasingly more prevalent and it affects the way nurses learn, practice, conduct research and develop policy. The discipline of nursing needs to maximize the benefits of big data to advance the vision of promoting human health and wellbeing. However, current practicing nurses and nurse scientists often lack the required skills and competencies necessary for meaningful use of big data. Some of those key skills for further development include the ability to mine narrative and structured data for new care or outcome patterns, effective data visualization approaches, and further integration of nursing sensitive data into artificial intelligence systems for better clinical decision support. We hope that our big data competencies recommendations' for practicing nurses, nurse educators, researchers, and policy makers will provide a vision for the growth-path and help prepare the next generation of nurses to improve patient outcomes trough better connected health.

References

[1] American Nurses Association. Nursing informatics: Scope and standards of practice. 2nd ed. Silver Spring: MD: Nursesbooks.org; 2014.

[2] Gandomi A, Haider M. Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management 35 (2015), 137-144.

[3] Bellazzi R. Big data and biomedical informatics: A challenging opportunity. Yearbook of Medical Informatics 9 (2014) 8-13. doi:10.15265/IY-2014-0024

[4] Kitchin R, McArdle G. What makes big data, big data? Exploring the ontological characteristics of 26 datasets. Big Data and Society 3 (2016),1-10.

[5] Brennan P, Bakken S. Nursing needs big data and big data needs nursing. J Nurs Scholarsh (2015), doi: 10.1111/jnu.12159

[6] McCormick K, & Saba V. Essentials of Nursing Informatics, 6th Edition McGraw-Hill Education, Philadelphia, PA, USA, 2015. Topaz M., Radhakrishnan K., John VL., Zhou L. Mining clinicians' electronic documentation to identify heart failure patients with ineffective self-management: A pilot text-mining study. Stud Health Technol Inform- NI2016 Proceedings (2016).

[7] Bowles KH, Holmes J, Ratcliffe S, Liberatore M, Nydick R, & Naylor MD. Factors identified by experts to support decision making for post acute referral. Nursing research, 58(2009), 115-128.

[8] Dey S, Cooner J, Delaney CW, et al. Mining Patterns Associated With Mobility Outcomes in Home Healthcare. Nurs Res 64(2105), 235-45.

[9] The University of Waikato. Weka 3: Data Mining Software in Java. http://www.cs.waikato.ac.nz/ml/weka/index.html (accessed 27 October 2016).

[10] KNIME.com AG. Open for Innovation. https://www.knime.org/ (accessed 27 October 2016).

[11] Lin C, Karlson E, Dligach D, et al. (2014). Automatic identification of methotrexate-induced liver toxicity in patients with rheumatoid arthritis from the electronic medical record. Journal of the American Medical Informatics Association 22(2105),e151-61.

[12] Zhou L, Plasek J, Mahoney LM, et al. Using Medical Text Extraction, Reasoning and Mapping System (MTERMS) to process medication information in outpatient clinical notes. AMIA Annu Symp Proc (2011), 1639-48.

[13] Topaz M, Lai K, Dowding D, Lei V, Zisberg A, Bowles K, Zhou L. Using natural language processing to automatically identify wound information in narrative clinical notes: application development and testing. Home Healthcare and Hospice Information Technology Conference (H3IT) Proceedings 2 (2015), 3-4.

[14] Carlson JN, Das S, De la Torre F. A Novel Artificial Intelligence System for Endotracheal Intubation. Prehosp Emerg Care 17 (2016), 1-5.

[15] Severin J, Beal K, Vilella K, Fitzgerald S, et al. eHive: an artificial intelligence workflow system for genomic analysis. BMC bioinformatics, 11(2010), 1-7.

[16] Ryan SA. An Expert System for Nursing Practice. Computers in Nursing 3(2)(1985), 77-84.

[17] Bloom KC, Leitner JE, Solano JL. Development of an Expert System Prototype to Generate Nursing Care Plans Based on Nursing Diagnoses. Computers in Nursing 5(2)(1987), 140-5.

[18] Ozbolt JG, Schultz S, Swain MA, Abraham IL. A Proposed Expert System for Nursing Practice: A Springboard to Nursing Science. Journal of Medical Systems 9(1-2)(1985), 57-68.

[19] Ozbolt JG. Developing Decision Support Systems for Nursing. Computers in Nursing 5(3)(1987), 105-11.

[20] Norris J, Cuddigan J, Foyt M, Leak G, Lazure L. Decision Support and Outcomes of Nurses' Care Planning. Computers in Nursing 8(5)(1990), 192-7.

[21] Sittig DF, Murphy DR, Smith MW. Graphical display of diagnostic test results in electronic health records: a comparison of 8 systems. Journal of the American Medical Informatics Association 22(2015), 900-904.

[22] Cho I, Kim E, Choi WH, Staggers N. Comparing usability testing outcomes and functions of six electronic nursing record systems. Int J Med Inform 88 (2016), 78-85.