Text Summarization of Medical Records using Text Mining

Spoorthi D S[1], Varsha S R[2], Raksha M[3], Harshitha N S[4]

[1,2,3,4]Dept. of CSE Vidyavardhaka College of Engineering, Mysuru, Karnataka, India

---***---

Abstract - *The analysis of medical records is a major challenge, considering they are generally presented in plain text, have a very specific technical vocabulary, and are nearly always unstructured. It is an interdisciplinary work that requires knowledge from several fields. The analysis may have several goals, such as assistance on clinical decision, classification of medical procedures, and to support hospital management decisions. This work presents the concepts involved, the relevant existent related work and the main open problems for future analysis among the analysis of electronic medical records, using data and text mining techniques. It provides a comprehensive contextualization to all those who wish to perform an analytical work of medical records, enabling the identification of fruitful research fields. With the digitalization of medical records and the large amount of medical information offered, this is often a locality of wide research potential.*

***Key Words* - data mining, text mining, electronic medical records, ICD codes, machine learning, healthcare informatics.**

1.INTRODUCTION

Analysis of medical records victimization text mining may be a complicated field stern goodish time and energy (Brown &amp; Trimble, 2000). Many sources of information like symptoms, exams, patient history, procedures, treatments, and medications got to be taken into consideration for an accurate scrutiny. Additionally, this analysis needs information in many completely different fields, namely, the clinical specific space, data processing, text mining, medical records, hospital and clinical procedures. Developing a method to mine medical records is hard. Usually these records area unit in free text, have an unstructured format and a specific and complex domain. In fact, every medico has sometimes his own approach of describing events or symptoms, betting on his previous learning experiences and medical practices (Weiner, Swain, Wolf, & Gottlieb, 2001). Medical doctors use a specific language, that sometimes demands extra tools to interpret the selected terms and symptoms and to extract linguistics data from medical records (Fernandez et al., 2004).

Another challenge is the huge amount of data. Online medical data is increasing every day generating huge amounts of electronic knowledge. Medical Literature Analysis and Retrieval System on-line, or MEDLARS on-line (MEDLINE) databases contain over twelve.5 million records, growing at the speed of five hundred 000 new citations annually (Cohen & Hersh, 2004). The increase of electronic medical records provides a large amount of knowledge to method and, at an equivalent time, offers a chance for developing analysis so as to cut back the time and effort to find and classify a correct diagnosis for a particular patient.

There are many varieties of medical knowledge, e.g., patient demographics, anamnesis, and science laboratory tests which will be accustomed completely different ends, such as, ordering, managing, scheduling, and charge. This knowledge helps physician's diagnosis and treating patients and therefore the correct management of resources (Ludwick & Doucette, 2008).

Furthermore, physicians use customary codes to explain diagnoses, procedures or treatments, which can embrace options like symptoms, diseases or disorders. International Classification of Diseases (ICD) and systematic terminology of drugs Clinical Terms (SNOMED) are 2 samples of customary codes (Schulz, Rector, Rodrigues, & Spackman, 2012). This classification is essential for sharing info among clinicians and additionally for request. The classification method isn't easy, representing another challenge within the space. In the following sections, we have a tendency to introduce the relevant ideas of knowledge mining, text mining, medical records, and international customary classifications for diagnose, clinical procedures and coverings, indicating necessary techniques, procedures and tools used. We carry on by reviewing techniques, main challenges and tools in text mining applied to medical records, call support, health management and classification systems. Finally, relevant open problems are mentioned presenting promising future analysis trends.

2. LITERATURE SURVEY

There are differing kinds of medical data sources, namely, hospital data systems (HIS), electronic health records (EHR), or electronic medical records (EMR). HIS is a system that can manage medical, administrative, financial and legal aspects of a hospital (Tsumoto & Hirano, 2011). EHR area unit a group of medical records of individual patients or a population. They permit patient pursuit and supply call support mechanisms to access patient info across facilities of an establishment (Hoerbst &

Ammenwerth, 2010). Electronic medical records (EMR) grant the chance to store all relevant patient info in electronic format. This info could contain symptoms, notes, remarks created by one or a lot of physicians, and descriptions of relevant patient events. These records give medical support for physicians, organizations and for communication among them. They can be in numerous formats describing, for instance, symptoms, medical history, social problems, and lab test results. The documentation is often structured or unstructured, could have grammatical eleven errors, abbreviations, with specific and local vocabulary which can be difficult to classify and analyse. These systems might demand vital initial investments, may be difficult to implement or adapt, and can generate technical problems. However, it's potential to recover the investment in a very future with higher service quality, reducing medical errors, providing faster access to resources and better administrative management (Wang et al., 2003). In 1907, salad dressing was the primary clinic to develop centralized medical records, in which each patient had its own records (Melton, 1996). In 1969, Lawrence Weed created a regular that organized these records (Weed, 1969). Still, the records might solely be accessed by one person, required a lot of space to be stored, and should be organized to a faster access. Only in 1972, the Regenstrief Institute developed one in all the primary EMR systems (Luo, 2006). Several protocols were created to ensure interoperability and communication, such as Health Level 7 (HL7) to exchange information and ensure communication between different hospitals; Electronic Data Interchange (EDI), which can be utilized by yankee National Standards Institute X12 (ANSI X12)(Institute, 2013), to exchange structured information; and Digital Imaging and Communications in medication (DICOM) (Association, 2013), to save, print and exchange medical images (Luo, 2006; Mcdonald, 1997). There are several EMR applications in this field, like the Veterans Health Information Systems and Technology Architecture (VISTA) that allows the visualization of multimedia information in several fields like cardiology or radiology. Furthermore, it permits the verification, analysis, and update of patient data like, as an example, medication orders. EMR Experts, EMRRitus and Charting Plus are other examples of technological implementations of EMR. In the next section we will explorer the several standards used to classify clinical procedures, treatments and diagnoses

3. CONCLUSION

In this review we've given a comprehensive assortment of representative works on the sphere of text mining applied to Electronic Medical Records (EMR). As the analysis subject is roofed by a broad vary of various mining techniques and ideas, we've conjointly represented the background ideas and predominate techniques associated. The main areas embrace call support systems, classification of diagnosing, treatments and proceedings, and health management systems. Additionally, this text discusses key open problems in these numerous areas. It is possible to conclude that using text mining techniques to analyse medical information can reduce the effort and time to diagnose a patient or suggest treatments or proceedings to improve healthcare. Text mining will determine adverse events and develop algorithms to recommend proceedings or watching risky patients. Nevertheless, developing a method capable of suggesting or taking proceedings needs a posh understanding of many data fields, like clinical texts, medical proceedings, text mining, data processing, among others. Unstructured info, such as free texts provides many features that can be analysed, although the identification and extraction of information may offer a huge complexity in this field. Hence, we will conclude that text mining is already extraordinarily valuable once applied to EMR, but can, nevertheless, become crucial in future applications. Further analysis within the space of text mining applied to EMR can most actually follow white box approaches. Physicians square measure still apprehensive with this sort of technology, given its low accuracy, questions regarding security and access control that may exist. Therefore, the utilization of measures to get rid of or replace non-public patient info are pursued.

REFERENCES

[1] Ananiadou, S., Kell, D. B., & Tsujii, J.-i. (2006). Text mining and its potential applications in systems biology. Trends Biotechnol, 24(12).

[2] Armstrong, D. (2011). Diagnosis and nosology in primary care. Sociology of Diagnosis, 73(6), 801–807.

[3] Association, N. E. M. (2013). Digital Imaging and Communications in Medicine.

[4] Baesens, B. (2011). Guest Editorial: Special Section on White Box Nonlinear Prediction Models. IEEE Transactions On Neural Networks, 22(12).

[5] E. Berwick, D. M., & Hackbarth, A. D. (2012). Eliminating Waste in US Health Care. JAMA, 307(14), 1513-1516.

[6] Brown, Richard J., and Michael R. Trimble. "Dissociative psychopathology, non-epileptic

seizures, and neurology." Journal of Neurology, Neurosurgery & Psychiatry 69.3 (2000): 285- 289.

[7] Carrington, M. J., Kok, S., Jansen, K., & Stewart, S. (2013). The Green, Amber, Red Delineation of Risk and Need (GARDIAN) management system: a pragmatic approach to optimizing heart health from primary prevention to chronic disease management. European Journal of Cardiovascular Nursing, 12(4), 337-345. doi: 10.1177/1474515112451702

[8] Chaovalit, P., & Zhou, L. (2005). Movie Review Mining: a Comparison between Supervised and Unsupervised Classification Approaches. Paper presented at the Proceedings of the 38th Hawaii International Conference on System Sciences.

[9] Chapelle, O., Schölkopf, B., & Zien, A. (2006). Semi-Supervised Learning (Vol. 1). Cambridge, MA: MIT Press.

[10] Cheng, S., Azarian, M. H., & Pecht, M. G. (2010). Sensor systems for prognostics and health management. [Review]. Sensors (Basel), 10(6), 5774-5797. doi: 10.3390/s100605774

[11] Chowdhury, G. G. (2005). Natural language processing. Information Science and Technology, 37(1), 51-89

[12] Christen, P. (2012). A Survey of Indexing Techniques for Scalable Record Linkage and Deduplication. IEEE Transactions On Knowledge And Data Engineering, 24(5).

[13] Chuang, C. C., Su, S.-F., Jeng, J. T., & Hsiao, C. C. (2002). Robust support vector regression networks for function approximation with outliers. Neural Networks, 13(6), 1322 - 1330.

[14] Cohen, A. M., & Hersh, W. R. (2004). A survey of current work in biomedical text mining. Briefings in Informatics, 6(1), 57–71.

[15] Coonan, K. M. (2004). Medical informatics standards applicable to emergency department information systems: making sense of the jumble. Academic Emergency Medicine, 11(11), 1198-1205.