Comprehensive Solution for Resolving Master Data Management Issues in Electronic Medical Records (EMR) Systems

Aditya Patel, PhD Scholar, Medical Technology, Weill Cornell Graduate School of Medical Sciences, New York

Abstract

Master Data Management (MDM) issues in Electronic Medical Records (EMR) systems, such as inconsistent drug information, pose significant challenges for healthcare providers. This paper outlines a comprehensive solution to these issues by implementing advanced data standardization, normalization, and matching techniques. The solution leverages machine learning algorithms, Natural Language Processing (NLP), and robust database management practices to ensure accurate, reliable, and consistent master drug lists.

Keywords*: Master Data Management, Electronic Medical Records, EMR systems, data standardization, data normalization, data matching, machine learning algorithms, Natural Language Processing, NLP, master drug list, healthcare providers, patient safety*

Introduction

Electronic Medical Records (EMR) systems are essential for modern healthcare, providing a digital version of patients' paper charts. However, inconsistencies in data, particularly with drug information, can lead to significant clinical and operational challenges. Common issues include misspellings, dosage information along with the drug, and the use of generic or brand names. These discrepancies complicate matching drug names to a master drug list, affecting patient safety and treatment efficacy.

Problem Statement

The primary challenge addressed in this paper is the inconsistency and inaccuracy of drug information in EMR data. Specific issues include:

1. **Misspelled Drug Names**: Variations in spelling can hinder the accurate matching of drug names to the master list.
2. **Inconsistent Dosage Information**: Differences in how dosage is recorded (e.g., mg, milligrams, micrograms) can create confusion and errors.

3. **Generic vs. Brand Names**: Using generic or brand names without standardization can lead to duplication and mismatches in the master list.

EMR Data Ingestion Process

- **Integration**: Patients, drugs, diseases, treatment, and other health information from the EMR for retina practices within the US and Canada are integrated into the warehouse from different EMRs.
- **ETL Process**: The process ingests raw EMR export files in various formats (CSV, TXT, XML, JSON, or direct database access).
- **Mapping**: Multiple different datasets are mapped to standard code lists, including:
  - Doctor's names and practice addresses
  - Drug names
  - Treatment names
  - Disease information and related information

Challenges

During the data mapping from EMR data to our standard code list, drug names are not a simple one-to-one mapping due to issues such as:

- Misspelled drug names
- Drug names with dosage information provided
- Drug names with treatment information
- Generic vs. brand-name drugs
- Vitamins and other categories mixed with drugs
- Plain text descriptions of treatments instead of drug names
- New drugs not in the master drug table or categories

Possible Solutions

1. **Fuzzy Text Matching**: Use fuzzy text matching to handle misspelled drug names and find the closest match in the master drug list.
2. **Classification Algorithms**: Logistic regression or other machine learning algorithms are used to classify data into relevant categories.
   - **Two-Step Classification**:

  - Step 1: Classify the entry as a drug or non-drug.
  - Step 2: Apply fuzzy text matching for exact or similar master drug match.

3. **NLP Processing**: Implement Natural Language Processing (NLP) techniques to process and standardize textual data.

Conclusion

By leveraging advanced data standardization, normalization, matching techniques, machine learning algorithms, and NLP, the comprehensive solution ensures accurate, reliable, and consistent master drug lists in EMR systems, improving patient safety and treatment efficacy.

References

1. Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, 2(1), 3**.**
2. Hossain E, Rana R, Higgins N, Soar J, Barua PD, Pisani AR, Turner K. Natural Language Processing in Electronic Health Records in relation to healthcare decision-making: A systematic review. Comput Biol Med. 2023 Mar;155:106649. doi: 10.1016/j.compbiomed.2023.106649. Epub 2023 Feb 10. PMID: 36805219.
3. Shah, N. H., & Tenenbaum, J. D. (2012). The coming age of data-driven medicine: translational bioinformatics' next frontier. *Journal of the American Medical Informatics Association*, 19(e1), e2-e4.
4. Dagliati A, Sacchi L, Zambelli A, Tibollo V, Pavesi L, Holmes JH, Bellazzi R. Temporal electronic phenotyping by mining careflows of breast cancer patients. J Biomed Inform. 2017 Feb;66:136-147. doi: 10.1016/j.jbi.2016.12.012. Epub 2017 Jan 3. PMID: 28057564.**.**
5. Hersh, W. R., Weiner, M. G., Embi, P. J., Logan, J. R., Payne, P. R., Bernstam, E. V., Lehmann, H. P., Hripcsak, G., Hartzog, T. H., Cimino, J. J., & Saltz, J. H. (2013). Caveats for the use of operational electronic health record data in comparative effectiveness research. Medical care, 51(8 Suppl 3), S30–S37. https://doi.org/10.1097/MLR.0b013e31829b1dbd
6. Saxena, Pushkarprabhat & Mayi, Krishna & Arun, R & Kumar, Mr & Biswo, Dr & Mishra, Ranjan & Praveen, K. (2023). Impact of Artificial Intelligence on Healthcare Informatics: Opportunities and Challenges. Journal of Informatics Education and Research. 3. 2309. 10.52783/jier.v3i2.384.
7. Singh, Pushpa & Singh, Dr & Bajaj, G. (2020). Role of Data Mining Techniques in Bioinformatics. International Journal of Applied Research in Bioinformatics. 11. 51-60. 10.4018/IJARB.2021010106.
8. Data-driven Precision Medicine: AI, ML, and Master Data Management in Healthcare (P. M. A. Khan , Trans.). (2024). International Journal of Creative Research In Computer Technology and Design, 6(6), 1-8. https://jrctd.in/index.php/IJRCTD/article/view/37
9. Murphy, S. N., & Chueh, H. C. (2002). A security architecture for query tools used to access large biomedical databases. Proceedings. AMIA Symposium, 552–556.