Sentiment Analysis for Performance Evaluation of Maintenance in Healthcare

Sentiment Analysis for Performance Evaluation of Maintenance in Healthcare

Lorenzo Mascii[1], Alessio Luschi[2], and Ernesto Iadanza[2][0000-0002-7291-4990]

[1] Master's Degree in Biomedical Engineering, University of Florence, Italy
[2] Department of Information Engineering, University of Florence, Italy

lorenzo.mascii@stud.unifi.it
alessio.luschi@unifi.it
ernesto.iadanza@unifi.it

Abstract. This paper presents a framework which makes use of Sentiment Analysis techniques for retrieving Real World Data (RWD) starting from scheduled and corrective maintenance data. The scope of the analysis is to automatically extract features from maintenance work orders, in order to calculate Key Performance Indicators of maintenance operations on medical devices, for Health Technologies Assessment purposes. Data are extracted from Computerized Maintenance Management System reports of healthcare facilities.

Keywords: clinical engineering, HTM, KPI, sentiment analysis, HTA, medical equipment.

1 Introduction

World Health Organization (WHO) defines Health Technology Assessment (HTA) as "the systematic evaluation of properties, effects, and/or impacts of health technology. It is a multidisciplinary process to evaluate the social, economic, organizational and ethical issues of a health intervention or health technology. The main purpose of conducting an assessment is to inform a policy decision making." [1]

In recent years, the analysis of Real-World Data (RWD) has become more and more important in HTA, allowing to generate reliable Real-World Evidence (RWE) which is very useful in terms of HTA analysis and reporting. The main challenge about RWD is their high degree of heterogeneity. They can originate from Randomized Clinical Trials (RCT), epidemiological studies, administrative databases, regulatory or institutional registers and secondary sources. In recent years, social media have proved to be one of the most popular and wide secondary sources of RWD, surely increasing the level of diversity and complexity [2]. Data from social media are public and can be analyzed relatively easily, overcoming many logistical obstacles associated with traditional approaches, while enabling fast and cost-effective data collection, providing a reliable insight into the patient experience. Therefore, the design of effective methods for the automated extraction of useful and aggregated Smart Data (i.e. data enabling accurate and precise decision-making processes) by using data mining processes on various heterogeneous sources is currently needed [3].

Sentiment Analysis is a fairly new field of study, grown in parallel with the spread of social media, for analyzing opinions, evaluations, attitudes and sentiments of people in relation to products, services, organizations, topics and personalities [4].

Maintenance is a crucial subject for medical equipment life cycle management. Evidence-based maintenance consists of the continuous performance monitoring of equipment, starting from the evidence and improvement of its effectiveness by making the required changes. Maintenance service work orders are stored in Computerized Maintenance Management Systems (CMMS) [5,6,7]. These are software applications which supports technicians and managers to track resources, manage work orders and costs, schedule maintenance plans and check the compliance to quality standards [8,9,10].

Other studies have developed failure classification models and KPIs for evaluating medical equipment maintenance performance [11]. These methods rely on data extracted and then classified directly by users, with a subsequential high cost in terms of computational evaluation time and resources.

This study aims to apply Sentiment Analysis and Natural Language Processing (NLP) techniques to extract features from maintenance's work orders reports stored in hospital's CMMS (both scheduled and corrective). The extracted features are then used as input to a supervised machine learning (ML) classifier to obtain fault codes and finally to evaluate KPIs, thus improving evidence-based maintenance.

2 Methods

2.1 Sentiment Analysis

According to literature, there are two main approaches to Sentiment Analysis: Lexicon-based approach and Artificial Intelligence (AI) approach.

- the former uses common dictionaries (e.g. SentiWordNet [12]) to compare nouns and adjectives and score them with a positive, neutral or negative weight [13,14].
- the latter uses AI algorithms to solve NLP problems, and to evaluate the final sentiment score by using classification models [15,16].

Both techniques require pre-processing normalization and transformation to transform the sentences recovered via text mining [17,18] into comprehensive and usable input text. Pre-processing task includes:

- sequence-splitting (extraction of single sequences from analyzed texts)
- spell-checking
- case-folding (making each word lowercased or uppercased)
- word-stopping (automatically excluding given words from further steps basing upon dictionaries)
- lemming/Stemming (converting all verb forms and tenses to present infinitive and all plurals nouns to singulars)
- tokenization (converting text into numerical representation)

When using the ML approach, feature extraction and selection are also needed. Typical NLP algorithms for feature extraction tasks are *BoW* (Bag of Words) and *TF-IDF* (Terms Frequency – Inverse Document Frequency) [15]. The former simply counts how many times each word occurs inside the document, while the latter links each word with a statistic measurement estimating the importance of the word itself inside the document. Specifically, the weight w_{ij} of the word t_i inside the document j is evaluated:

$$w_{ij} = tf_{ij} \times \log \frac{N}{df_i}$$

Where:

- tf_{ij} is how many times the word t_i occurs inside the document j
- N is the total number of documents
- df_i is the number of documents which includes the word t_i

Feature selection is the process of selecting the most significant and relevant features from a vast set of features in the given dataset. Feature selection helps in finding the smallest set of features which results in:

- faster training
- reducing the complexity of a model and making it easier to interpret
- building a sensible model with better predicting power
- reducing over-fitting by selecting the right set of features.

This task can be performed with filtered methods, wrappers or embedded methods. With a filtered method the selection is machine-learning-independent and thus can be used with both approaches (supervised and unsupervised). In wrapper methods, the feature selection process is based on a specific ML algorithm that we are trying to fit on a given dataset. With embedded methods, feature selection is done by observing each iteration of model training phase (Figure 1).

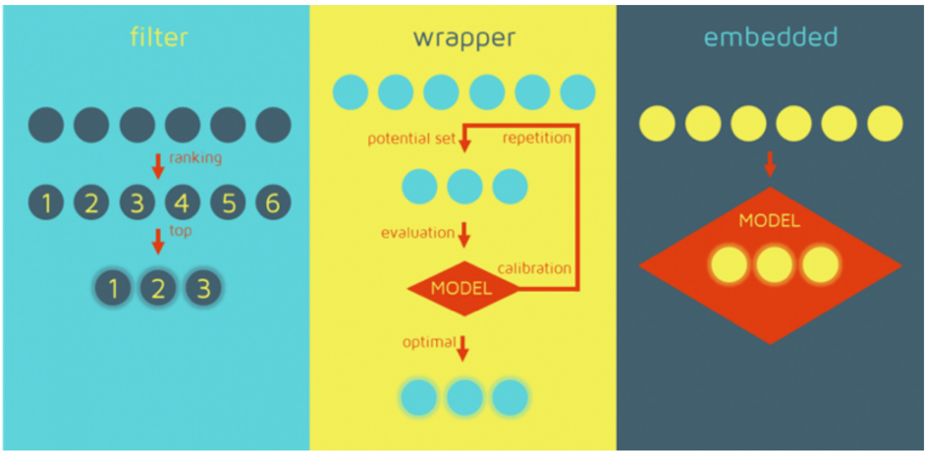

Figure 1. Schematic representation of Feature Selection methods.

Sentiment Analysis may use ML approaches to obtain a sentiment score for the initial data. In case of supervised ML, a specific algorithm is used to find a statistical model which better represents already classified elements, such as Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision-tree, Naïve Bayes (NB), Maximum Entropy or Neural Networks (NN).

In unsupervised ML, users do not need to provide the algorithm with output labels. Instead, the algorithm is able to autonomously discover patterns and information about the selected features.

A Recurrent Neural Network (RNN) is a class of artificial neural networks for unsupervised ML techniques where connections between nodes form a directed graph along a temporal sequence [19]. Convolutional Neural Network (CNN) is another class of deep neural networks, most commonly applied to analyzing visual imagery but which can also be implemented for NLP problems [20].

Figure 2 summarizes the whole described process.

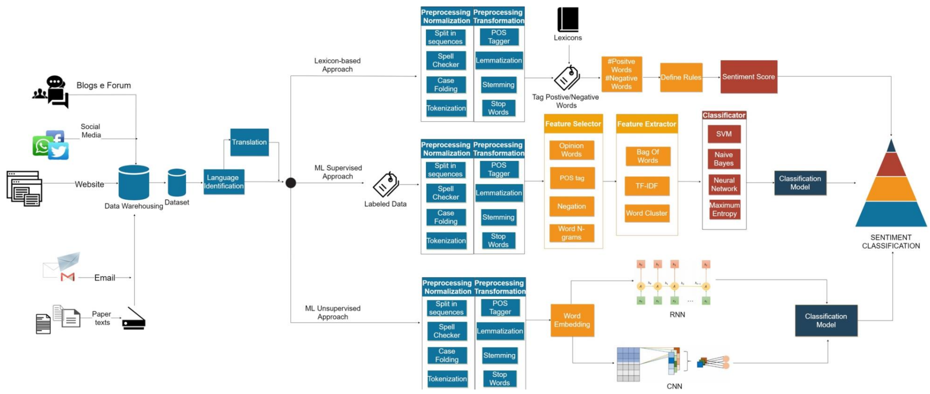

Figure 2. Different Sentiment Analysis approaches.

2.2 Failure classification and KPIs

Tecnicians from the Clinical Engineering service assign a fault code whenever a maintenance activity is performed. The code is chosen among a set of available codes [11]. This method allows evaluating and managing of the efficacy and efficiency of each maintenance operation.

Source	Activity	Code	Code definition	Severity
Equipment	Corrective maintenance (CM)	UPF	Unpreventable failure, evident to user, typically caused by normal wear and tear but is unpredictable	1
		PPF	Preventable and predictable failure, evident to user, typically caused by wear and tear that can be predicted or detected	3
		USE	Failures induced by use, eg, abuse, abnormal wear and tear, accident, or environment issues	1
		SIF	Service-induced failure, ie, failure induced by corrective or scheduled maintenance that was not properly completed or a part that was replaced and had premature failure ("infant mortality")	2
	Scheduled maintenance (SM)	PF	Potential failure, ie, failure is either about to occur or in the process of occurring but has not yet caused equipment to stop working or problems to patients or users	4
		EF	Evident failure, ie, a problem that can be detected but was not reported by the user without running any special tests or using specialized test/measurement equipment	1
		HF	Hidden failure, ie, a problem that could not be detected by the user unless running a special test or using specialized test/measurement equipment	5
	CM and SM	NPF	No problem found, including alleged failures that could not be duplicated ("cannot duplicate" [CND])	0
Accessories	CM or SM	BATT	Battery failure, ie, battery(ies) failed before the scheduled replacement time	1
		ACC	Other accessory failures, excluding batteries, evident to user, typically caused by normal wear and tear	1
Accessory (including rechargeable battery) failures are segregated from equipment failures to help determine the cause of problems and identify solutions.				

Figure 3. Set of available fault codes for both scheduled and corrective maintenance.

The UNI EN 15341:2007 standard [21] describes a system for managing KPIs to measure maintenance performance as influenced by key maintenance factors and to assess and improve efficiency and effectiveness. The standard suggests that the KPIs be structured into three groups to measure every aspect of the maintenance process: financial, technological and organizational. Previous studies have taken into account the criteria suggested by the above standard and designed a set of 20 KPIs described in Table 1 [11].

Table 1. List of KPIs.

Index	Name	Typology
KPI 1	Downtime (%) (non-availability time)	Technological
KPI 2	Uptime (%) (availability time)	Technological
KPI 3	MTTR (mean time to restoration)	Technological
KPI 4	MTBF (mean time between failures)	Technological
KPI 5	Class failure ratio (fails per class)	Technological
KPI 6	Global failure rate (defectiveness)	Technological
KPI 7	AFR: age failure rate	Technological
KPI 8	"Negligent" actions (%)	Organizational
KPI 9	"1 day" actions	Organizational
KPI 10	SM with failure (%)	Organizational
KPI 11	SM coverage rate (scheduled maintenance)	Organizational
KPI 12	No problem found (fake faults) (%)	Organizational

KPI 13	No. devices per technician (internal)	Organizational
KPI 14	Time cost of the workforce	Organizational
KPI 15	COSR (cost of service ratio = global maintenance to acquisition cost) (%)	Financial
KPI 16	External maintenance Cost (% with respect to total maintenance cost)	Financial
KPI 17	Internal maintenance cost (% with respect to total maintenance cost)	Financial
KPI 18	Corrective maintenance cost (CM) (% with respect to total maintenance cost)	Financial
KPI 19	Scheduled maintenance cost (SM) (% with respect to total maintenance cost)	Financial
KPI 20	Cost of spare parts (+ consumables) (% with respect to total maintenance cost)	Financial

3 Results

The input of the designed framework is a collection of work orders extracted from the CMMS. The work orders alone cannot be directly used to input the system. In fact, data are unstructured (i.e. they are written in several different ways), resulting in unstructured information. Hence, the first step is to implements a TimeML compliant time-expression tagger [22]. HeidelTime has been chosen to be implemented as temporal tagger to extract time information and normalize them according to TimeML's TimeX3 tag. Once data have been tagged and normalized, common Sentiment Analysis algorithms can be applied.

Preprocessing is initially needed to obtain semantically meaningful words and usable input text to subsequently perform feature extraction and selection on. This is achieved by using Sentiment Analysis preprocessing methods described in chapter 2.

Then TF-IDF algorithm is used for feature extraction. This has been chosen over BoW because more accurate and reliable. However, frequent words may not always have semantic significance and so they might be excluded from the obtained features. Moreover, this technique may easily lead to a huge amount of features which would cause the machine learning process to be spatial and temporal impracticable. *Drop-feature* predictive method is indeed used to give each feature a score (feature importance) to define which features have relevance in computational terms and which not, and so perform a feature selection.

Initial data contained inside the work orders of the CMMS can often be unbalanced, i.e. some classes may occur more frequently than others. This kind of data significantly reduces the performance of machine learning models: classifiers trained with unbalanced data tends to completely ignore the minority classes despite being useful for the final result. Many algorithms actually exist which prevent unbalanced learning by oversampling or undersampling the starting dataset. *Synthetic Minority Oversampling Technique* (SMOTE) has been successfully applied in many cases where unbalanced learning needed to be avoided, and it allows to add new data to the original dataset creating artificial datapoints based upon likeness between existent minorities [23].

Thereafter, machine-learning models are implemented to classify the extracted and filtered features. Support Vector Machine (SVM) supervised classifier has been chosen because it performs better in high-dimensions spaces and has a good cost-to-performance ratio [23,24].

The desirable output is a classification of the number of maintenance operations for type of fault code, both for scheduled and corrective maintenance. Finally, a subset of the previously described KPIs (especially the technological ones) can be automatically evaluated by using some fault codes or a combination of them (Figure 4).

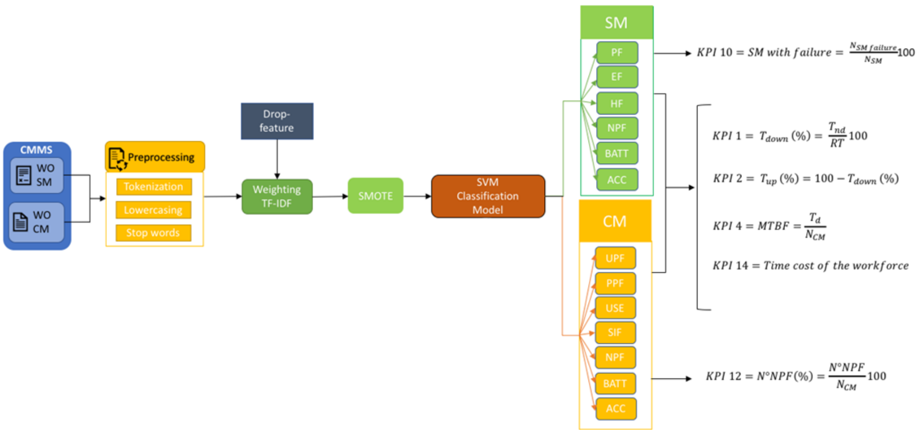

Figure 4. Designed framework for automatic classification of maintenance's work orders.

4 Conclusion

The proposed HTA framework implements Sentiment Analysis methods to analyze RWD extracted from maintenance's work orders reports manually produced inside CMMS.

The analysis itself is a restricted NLP problem. In fact, the system does not need to comprehend the semantic of each phrase, but just few features correlated to the maintenance classification according to a set of available fault codes. This type of RWD are extremely unstructured and unbalanced, so external algorithms and models (time-expression tagging, SMOTE) are needed to transform them to be suitable for the analysis. Sentiment Analysis' common preprocessing is then performed (lemming, stemming, case-folding, tokenization), followed by feature extraction via TD-IDF algorithm, feature selection via drop-feature algorithm and supervised machine learning classification with SVM.

The output is a multi-class classification of work orders with fault codes, allowing an evidence-based maintenance approach, which monitors the efficacy and efficiency of maintenance operations on medical devices. KPIs are (partly) automatically calculated by the system, allowing top-management to objectively better understand the weak points and improve the process.

Eventually, the designed framework leads to an optimization of costs and resources, and also to improvements in terms of patients' safety and risk management by predicting faults of critical assets.

References

1. WHO – Health technology assessment: https://www.who.int/medical_devices/assessment/en
2. Yingjie, L., Yang, W., Jingfang, L., Li, J., Pengzhu, Z.: Understanding Health Care Social Media Use from Different Stakeholder Perspectives: A Content Analysis of an Online Health Community. Journal of Medical Internet Research (2017).
3. Akay, A., Dragomir, A., Erlandsson, B.E.: Network-Based Modeling and Intelligent Data Mining of Social Media for Improving Care, Journal of Biomedical and Health Informatics (2013).
4. Liu, B.: Sentiment analysis and opinion mining. Synthesis lectures on human language technologies (2012).
5. Medenou, D., Fagbemi, L.A., Houessouvo, R.C., Jossou, T.R. et al.: Medical devices in Sub-Saharan Africa: optimal assistance via a computerized maintenance management system (CMMS) in Benin. Health and Technology, vol. 9, pp. 219-232 (2019).
6. Luschi, A., Marzi, L., Miniati, R., Iadanza, E.: A Custom Decision-Support Information System for Structural and Technological Analysis in Healthcare. IFMBE Proceedings, vol. 41, pp. 1350-1353 (2013).
7. Iadanza, E., Marzi, L., Dori, F., Biffi Gentili, G., Torricelli, M.C.: Hospital healthcare offer. A monitoring multidisciplinar approach. IFMBE Proceedings, vol. 14, pp. 3685-3688 (2006)
8. Luschi, A., Miniati, R., Iadanza, E.: A Web Based Integrated Healthcare Facility Management System. IFMBE Proceedings, vol. 45, pp. 633-636 (2015).
9. Iadanza, E., Luschi, A., Ancora, A.: Bed management in hospital systems. IFMBE Proceedings, vol. 68/3, pp. 313-316 (2019).
10. Iadanza, E., Luschi, A., Gusino, R., Terzaghi, F.: Designing a healthcare computer aided facility management system: a new approach. In: International Conference on Medical and Biological Engineering. IFMBE Proceedings, vol. 73, pp. 407-411 (2019).
11. Iadanza, E., Gonnelli, V., Satta, F. et al.: Evidence-based medical equipment management: a convenient implementation. Medical & Biological Engineering & Computing, vol. 57, pp. 2215–2230 (2019).
12. Miller, G., Beckwith, T., Fellbaum, C. et al: WordNet: an on-line lexical database. International Journal of Lexicography, vol. 3 (4), pp. 235-244 (1990).

13. Whitelaw, C., Garg, N, Argamon, S.: Using appraisal groups for sentiment analysis. Proceedings of the ACM SIGIR Conference on Information and Knowledge Management (CIKM), p. 625–31 (2005).
14. Vu, L., Le, T.: A lexicon-based method for Sentiment Analysis using social network data. International Conference on Information and Knowledge Engineering (IKE 17), Las Vegas, USA (2017).
15. Du, J., Xu, J., Song, H., Xiangyu, L., Cui, T.: Optimization on machine learning based approaches for sentiment analysis on HPV vaccines related tweets. Journal of Biomedical Semantics, vol. 8 (9) (2017).
16. Duyu, T., Bing, Q., Ting, L.: Deep learning for sentiment analysis: successful approaches and future challenges. WIREs Data Mining and Knowledge Discovery, vol. 5 (6), pp. 292-303 (2015).
17. Mine, G., Elder, J., Fast, A. et al..: Practical text mining and statistical analysis for non-structured text data applications. Academic Press (2012).
18. Heerschop, B, Goossen, F., Hogenboom, A. et al.: Polarity Analysis of Texts using Discourse Structure. 20th ACM Conference on Information and Knowledge Management (2011).
19. Socher, R., Perelygin, A., Wu, J., Chuang, J. et al.: Recursive deep models for semantic compositionality over a sentiment treebank. Proceedings of the 2013 conference on empirical methods in natural language processing, pp. 1631-1642 (2013).
20. Porumb, M., Iadanza, E., Massaro, S., Pecchia, L.: A convolutional neural network approach to detect congestive heart failure. Biomedical Signal Processing and Control, vol. 55 (2020).
21. UNI EN 15341:2007 Maintenance - Maintenance Key Performance Indicators
22. TimeML – Markup Language for Temporal and Event Expressions: http://www.timeml.org/
23. Zhang, T., Bhatia, A., Pandya, D. et al.: Industrial text analytics for reliability with derivative-free optimization. Computer & Chemical Engineering, vol. 135 (2020).
24. Hartmann, J., Huppertz, J., Schamp, C., Heitmann, M.: Comparing automated text classification methods. International Journal of Research in Marketing, vol. 36 (1) (2020).