DDC-Outlier: Preventing medication errors using unsupervised learning

Henrique D.P. dos Santos[1], Ana Helena D.P.S. Ulbrich[2], Vinícius Woloszyn[3] and Renata Vieira[1]
henrique.santos.003@acad.pucrs.br, anaulbrich@ghc.com.br, vwoloszyn@inf.ufrgs.br, renata.vieira@pucrs.br

***Abstract*—Electronic Health Records (EHR) have brought valuable improvements to hospital practices by integrating patient information. In fact, the understanding of this data can prevent mistakes that may put patients' lives at risk. Nonetheless, to the best of our knowledge, there are no previous studies addressing the automatic detection of outlier prescriptions, regarding dosage and frequency. In this paper, we propose an unsupervised method, called Density-Distance-Centrality (DDC), to detect potential outlier prescriptions. A dataset with 563 thousand prescribed medications was used to assess our proposed approach against different state-of-the-art techniques for outlier detection. In the experiments, our approach achieves better results in the task of overdose and underdose detection in medical prescriptions, compared to other methods applied to this problem. Additionally, most of the false positive instances detected by our algorithm were potential prescriptions errors.**

***Index Terms*—Electronic Health Records, Prescription Errors, Unsupervised Learning**

I. Introduction

ELECTRONIC Health Records (EHR) have played an important role in hospital environments providing many benefits in terms of patient safety and health care quality [1]. These systems can produce a large and rich amount of information that reveals intrinsic relations between symptoms, diseases, drug interaction, and diagnoses, which can be used for different purposes. For instance, in a recent study, Rotmensch et al. performed an analysis of diseases and symptoms based on textual narratives in EHR, creating a knowledge graph from the co-occurrence of these two pieces of information [2].

Information extracted from EHR systems can also be employed to reveal hidden knowledge [3]. A study performed by Ashcroft [4] revealed that errors in prescribing are a common issue in the health care process, affecting around 9% of all medication orders. Although not all of these errors put patients' lives at risk, a harmless error could lead to undesirable side effects and affect patients' confidence in the medical treatment. Along these lines, a global campaign to prevent medication errors was launched by the *World Health Organization* to highlight the importance of this subject to health care quality [5]. This particular campaign aims to significantly reduce the indices of severe and harm medication errors in the next 5 years. However, to the best of our knowledge, there is no previous study addressing the automatic detection of wrong dosages and frequencies for medications in electronic prescriptions.

To address this issue, our work considers past data from electronic records to detect outlier prescriptions in order to avoid wrong prescriptions. To achieve this, we propose an unsupervised algorithm based on graph models to detect prescription outliers. Our approach uses previous prescriptions to automatically learn the threshold between normal and abnormal doses for each medication in electronic medication orders, highlighting the potential misuse. This context-aware characteristic (i.e., the DDC automatically learns the normal dose for each medication) is crucial, since there are many different prescribing practices in the world [6].

Several research domains may use graph structure to detect outliers, and the node centrality score is one of the features used to solve this task [7]. Centrality measures such as eigenvector [8] exist for symmetric, if *xRy* then *yRx*, and asymmetric relations between nodes, if *xRy* then not *yRx*. In this paper, we use graph centrality to detect outliers in prescription data.

The main contributions of this work are the following:

- an unsupervised method to detect overdoses and underdoses in electronic prescriptions, i.e., it does not depend on an annotated training set;
- our model relies only on two features, which does not require extensive pre-processing, thus reducing the computational cost;
- a context-aware graph that performs well in different medications, as it automatically learns the normal and abnormal doses for each medication;
- a sample dataset containing 150 thousand prescriptions for replicability purposes.

The rest of this paper is organized as follows: Section II presents previous works on detecting prescription errors and related graph centrality approaches. Section III gives details of our proposed algorithm and definitions used to identify outliers in prescriptions. In Subsection III-C we describe the tertiary hospital dataset and the pre-processing tasks. Section III-G describes the experiment performed with five state-of-the-art algorithms, followed by the results in Section IV. In addition, we perform a qualitative analysis of the results in Section V. Finally, in Section VI we summarize our conclusions and present further research directions.

*This work was partially supported by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) Foundation (Brazil), PUCRS (Pontifical Catholic University of Rio Grande do Sul), UFRGS (Federal University of Rio Grande do Sul) and GEP/GHC (Teaching and Research Group at Conceição Hospital Group).

[1]Pontifical Catholic University of Rio Grande do Sul
[2]Nossa Senhora da Conceição Hospital
[3]Federal University of Rio Grande do Sul

II. Background

Electronic and manual prescriptions have been compared by researchers, and it has been found that electronic records alone may not reduce prescription errors. More advanced systems with dose and frequency checking are needed to prevent potentially harmful errors [9]. However, to the best of our knowledge, electronic prescriptions have not been exploited to detect medication errors concerning dose and frequency. Usually, the focus is to avoid prescription errors by creating better systems that calculate the right unit for the medication [10] or that prevent medication errors by using patient histories [11], [12].

Park et al. [13] created a probabilistic graphical model to extract patterns from large prescription data and then showed divergent patterns in prescriptions for the same diseases. Nangle et al. [14] employed electronic prescription messages to extract drug dose amounts, units, and frequencies from freely-typed texts, with the aid of natural language processing techniques.

In another line, drug information from several sources is used to avoid drug side effects in prescriptions. Reps et al. developed a system that combines different methods and sources to provide side effect alerts when prescribing medicines [15].

In terms of detecting prescription errors, three previous works try to tackle this problem. Hauskrecht et al. [16] used historical EHR data (e.g. laboratory tests, medication orders, and procedures) to develop a system that is able to detect outlier actions for a given patient. One of the actions detected by the alert system is a possible mistake regarding the medication order. In addition, Hash et al. [17] used the patients' historical medical orders and diagnostic claims as data to detect look-alike/sound-alike medication errors. Both studies focus on possible medication mistakes considering only the medication name itself, not taking dosage and frequency errors into account.

The only work addressing dosage outliers used a commercial piece of software that tackles the problem to detect prescription errors [18]. This approach evaluated medication dosage outliers using a machine-learned dosage distribution of the medication in the population and/or the patient's own history. The paper, however, does not detail the techniques. In our work, we propose an unsupervised method to detect potential prescription errors regarding dose and frequency based on the prescription history data.

Our method uses a graph centrality approach for outlier detection. Graph models have been employed for outlier detection in other domains. For instance, Muller [19] employed centrality algorithms to rank nodes based on their centrality index to distinguish them between inliers and outliers. The same idea was successfully employed, for instance, to rank textual information of documents and automatically create textual summaries that contain the most common words based on their centrality on a graph [20], [21]. Another approach in the graph centrality field uses random walk on a graph to perform outlier detection [22]. It relies on computing the node similarity and the number of shared neighbors between nodes. Afterwards, they use a Markov chain model to compute the score for each node of the graph. Nodes with lower scores are considered outliers.

In the next section, we detail our Density Distance Centrality algorithm to detect outliers in prescription data.

III. Methods and Materials

The intuition behind the Density-Distance-Centrality (DDC) is that the detection of medication outliers can be regarded as the problem of finding groups of prescriptions with low density and low similarity among other prescriptions. Low density, in this sense, is an uncommon prescription, rarely prescribed in historical data; and low similarity is a prescription like any other regarding dose and frequency. To solve this problem, our approach relies on the concept of graph centrality to rank prescriptions according to their centrality index. Overdoses or underdoses are probably prescriptions whose centrality score lies below a mean centrality index for each medication.

The main step to compute prescription outliers is how to represent each prescription in the vector space. Here we fit each prescribed medication in a bi-dimensional vector with dose and daily frequency. Since each medication/presentation is used in its own way, regarding dosage and frequency, the graph is built considering each medication/presentation (e.g. Omeprazole 20 mg and Omeprazole 40 mg dispersible tablet belong to different graphs).

We represent the relationship between prescriptions as a graph, in which the vertices are the prescriptions (e.g. Omeprazole 20 mg: 40 mg twice a day) and the edges are defined in terms of the similarity between a pair of prescriptions. We define the similarity function as the pairwise similarity between the bi-dimensional vectors (dose and frequency). The pairwise similarity accepts any pairwise metric, discussed in Section III-B. Our hypothesis is that a normal prescription has a high centrality index since it is similar to many others prescriptions.

Let P be a set of prescriptions for a specific medication and $p \in P$ a tuple $\langle d, f \rangle$, where $p.d$ represents the dose of the medication and $p.f$ the daily frequency that the physician prescribed for their patient. First, the DDC builds a distribution list for this medication D counting the frequency of the tuple $\langle d, f \rangle$. Then, the DDC builds a graph representation $G = (V, E)$, where V is the unique dose/frequency and E is the set of edges that connects pairs $\langle u, v \rangle$ where $v, u \in V$. In the next step, it uses Weighted PageRank to calculate the centrality scores for each vertex. Finally, considering the mean centrality score as the outlier threshold, the DCC generates the outlier list, when each prescription above or equal to the threshold is assigned as inlier and all those below are considered outlier prescriptions.

The pseudo-code of the DDC is displayed in Algorithm 1, where G is represented as adjacency matrices W. In the remainder of this section, we detail the process to obtain the centrality index.

A. Weighted PageRank Centrality

To compute the centrality of each prescription, the DDC relies on PageRank [23], which considers each edge as a

Algorithm 1 - DDC-Outlier Algorithm (P, α): O

- Input: a set of prescribed medications P, α the frontier threshold.
- Output: list O containing the computed outlier value for each prescription $\in P$, 1 for inlier and -1 for outlier prescriptions.

$D \leftarrow 0$
for each $p \in P$ **do**
  $D[p.d, p.f] \leftarrow +1$
end for
for each $u, v \in D$ **do**
  $W[u,v] \leftarrow similarity(u,v)$
end for
$C \leftarrow WeightedPageRank(W, D)$
$\overline{E} \leftarrow mean(C)$
for each $p \in P$ **do**
  if $C[p.d, p.f] \geq \overline{E} * \alpha$ **then**
    $O[p] \leftarrow 1$
  else
    $O[p] \leftarrow -1$
  end if
end for
Return O

vote to determine the overall centrality score of each node in a graph. However, as in many types of networks, not all relationships are considered of equal importance. The premise underlying PageRank is that the importance of a node is measured in terms of both the number and the importance of the vertices it is related to.

In one extension of PageRank, the algorithm takes into account the importance of both the inlinks and the outlinks of the nodes [24]. In another extension, PageRank authors adopted a more realistic and less democratic stance by using a better (and more flexible) perturbation matrix, where the "personalization" vector $v^T > 0$ is a probability vector that allows non-uniform probabilities of teleporting to particular pages [25].

Our approach uses both extensions, weighted links, and weighted nodes, to compute the centrality score of prescriptions. The Weighted PageRank function is given by:

$$WPR(u) = \sum_{v \in B_u} W(v,u) \frac{WPR(v)}{N_v} \quad (1)$$

where, B_u is the set containing all neighborhoods of u, and N_v represents the number of neighborhoods of v. Besides, $W(v,u)$ is the weight of the outlink from v to u.

The intuition of using PageRank in the DCC is that the more a prescribed medication is connected to prescriptions that are highly similar to other prescriptions, the more representative it is in the prescription distribution.

B. Pairwise Metric

The DDC-Outlier algorithm could be used with any pairwise metric. The intuitive metric to gather similar prescribed medications is the cosine similarity. In our experiments, we also analyze other metrics to evaluate our algorithm to detect prescription outliers.

- DDC: cosine similarity between instances;
- DDC-C, Cosine: cosine distance between instances;
- DDC-J, Jaccard: a statistic used to compare the similarity and diversity of sample sets;

C. Prescription Data Source

The dataset was obtained from Hospital Nossa Senhora da Conceição (HNSC). The database contains 240 thousand Computerized Physician Order Entries (CPOE) entered between January and September 2017. All records are related to prescriptions, with 2 million medications prescribed to 16 thousand patients. The majority of the patients were born from the 1950s to the 1990s and were treated by the Brazilian public health care system.

The HNSC hospital belongs to the public healthcare system in Brazil, and, as a standard procedure, the hospitals at the public hospital environment always use generic names for medications (regardless of the brand acquired, the name registered in the system is the name of the active principle).

Each prescription record has the following information: the patient's register number, date of prescription, medication name and presentation, dose, route, frequency, and a free-text comment field. In our experiments, we only used the dose and frequency of each medication/presentation. In the next section, we detail the pre-processing tasks performed on these data.

Ethical approval to use the hospital dataset in this research was granted by the Research Ethics Committee of Grupo Hospitalar Conceição under the number 71571717.7.0000.5530.

D. Pre-Processing Prescriptions

We have noticed that there was noise in the raw data provided by the CPOE, requiring data cleansing to ensure good data quality in the experiments. We performed the four preprocessing tasks described below.

a) Unit Table: Some prescribed medications are presented in milligrams and others in grams. To avoid mistakes in the quantity, we set a standard unit for each medication. Then we listed all dose units used by the physicians in the CPOE and defined a factor to multiply the non-standard units.

b) Frequency Table: In the CPOE, each physician describes frequency in different ways. We standardized it to a daily frequency: 3 times a day (3), once a day (1), 6 times a day (6). This included terms such as "twice a day" (2), "6h/6h" (4) and "3x/day with a meal" (3).

c) Dosage Table: For validation purposes, we included the daily maximum and minimum doses for 345 prescribed drugs in the dataset, following two evidence-based references: Micromedex® [26] and UpToDate® [27]. Medications with no daily maximum and minimum doses were not considered in the experiments.

d) Medications Prune: Following the methodology proposed by Emmott et al. [28], medications with less than 1,000 records were discarded. In addition, when the number of candidate anomalous data points was small, we excluded the medication. We chose to discard medications with less than 10 outlier observations due to insufficient support.

E. Prescriptions Stats

After performing all the described pre-processing tasks, 51 medications remained in our dataset, with a total of 563 thousand records. Each medication has a particular prescription distribution: some of them are used only for disease treatments, others only to reduce symptoms, and others are intended for prophylactic use. These characteristics allow a wide evaluation of the proposed algorithm regarding a variety of prescription scenarios. In Table I we show the overall statistics concerning the prescriptions issued at HNSC.

TABLE I
PRESCRIPTION DATASET

Total Prescribed Medications	563,171
Overdosed Prescribed Medications	6,666
Underdosed Prescribed Medications	4,868

The large number of prescriptions (one thousand per day) hinders the screening process performed by the department of pharmacy services. Despite their efforts, there were more than 6 thousand overdose and almost 5 thousand underdose prescriptions in 2017. This scenario shows the importance of an automated system that is able to identify prescription outliers.

Figure 1 shows an example of the resulting data of the drug Acyclovir tablet. This shows a typical distribution, featuring some of the most common uses of this medication in the hospital, such as 200 mg three times a day and 400 mg three times a day. In addition, there are unusual prescriptions, such as 400 mg Acyclovir six times a day. The only points that are considered outliers for Acyclovir are the red ones, with prescriptions for 800 mg six times a day as overdose and 200mg once a day as underdose.

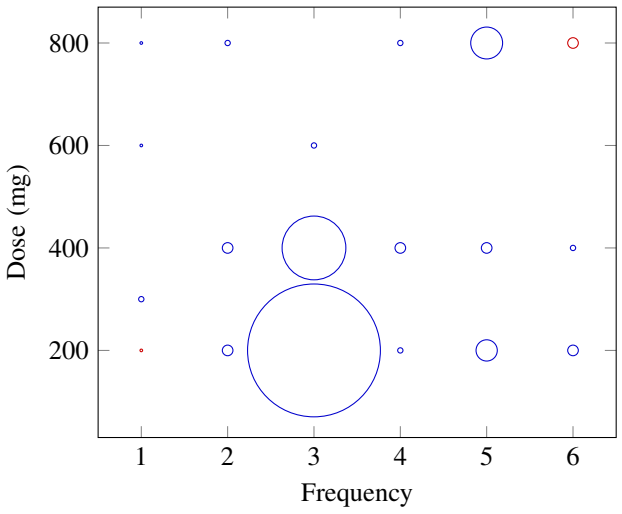

Fig. 1. Prescription of Acyclovir 200 mg tablet, where size means how many times this pair (dose, frequency) is prescribed. Blue points mean normal prescriptions and red points indicate outliers.

All low frequencies, represented by small circles, could be potential medication mistakes and could be highlighted by the department of pharmacy services for additional verification.

F. Data Sharing

A sample dataset with 150 thousand real prescriptions is provided for replicability purposes at the project's Github Page[1]. This dataset contains 21 different medications prescribed at Hospital Nossa Senhora da Conceição. In the next Section, we detail the experiments using the maximum and minimum dose information concerning the medications.

G. Experiments

In order to evaluate the performance of our proposed approach, we designed a task for each medication in the prescription dataset. The task consisted of identifying overdoses, that is, prescribed medication above the maximum daily dose specified in the literature, and underdoses, when the dose is below the lowest daily dose in the literature (regardless of indication).

Our experiments followed the methodology proposed by Emmott et al. work [28]:

- Top 3 rankings: it shows the number of medications in which each algorithm appeared in the top 3 algorithms when ranked by F-Measure;
- All analyzed algorithms require a parameter. We employed parameter search to find the best parameters to maximize the hits (as described below, in Section III-J). In all cases, we made a good faith effort to maximize the performance of all methods;

H. Baseline

To evaluate our approach, we selected as baseline several state-of-the-art unsupervised methods used to detect outliers [29], as explained below:

a) One Class SVM: it was introduced by Schölkopf et al. It requires the election of an SVM kernel and a scalar parameter to define the frontier of outlier instances. Here we chose the RBF kernel, which better fits our experiments [30].

b) Local Outlier Factor: it computes a score reflecting the degree of anomaly instances. It measures the local density deviation of a given data point with respect to its neighbors. The idea is to detect the samples that have substantially lower densities than their neighbors [31].

c) Gaussian Mixture: it is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters [32].

d) Robust Covariance: assuming that the inlier data are Gaussian distributed, it will estimate the inlier location and covariance in a robust way (i.e., without being influenced by outliers). The Mahalanobis distances obtained from this estimate are used to derive a measure of outlyingness [33].

e) Isolation Forest: it isolates the observations by randomly selecting a feature and then randomly selecting a split value between the maximum and minimum values of the selected feature. Random partitioning produces noticeably shorter paths for anomalies. Hence, when a forest of random trees collectively produces shorter path lengths for particular samples, they are highly likely to be anomalies [34].

[1]https://github.com/nlp-pucrs/prescription-outliers

I. Performance Metric

The task of detecting outliers is a classification with a class imbalance problem, where the main class of interest is rare. That is, the dataset distribution reflects a significant majority of the negative class (non-outlier prescriptions) and a minority of the positive class (overdose and underdose prescriptions). Therefore, choosing the right performance metric is essential to correctly evaluate all methods regarding this problem. In this sense, we briefly describe the most common metrics to evaluate binary classifiers [29]:

a) Accuracy: is the percentage of instances labeled as the correct class (positive or negative);

b) Recall: is a measure of completeness (i.e., what percentage of positive instances are labeled as such);

c) Precision: can be thought of as a measure of exactness (i.e., what percentage of instances labeled as positive are actually such);

d) F-Measure: corresponds to the harmonic mean between precision and recall.

An algorithm that predicts that the majority of instances belong to the negative class (not medication errors), in an imbalance problem, could have a high accuracy score, but it is useless to predict the aimed positive class (medication errors). On the other hand, an algorithm that predicts that all instances relate to the positive class will have a high recall, but it is unable to distinguish between positive and negative instances. For this reason, we selected F-Measure as the main metric to evaluate the performance of outlier algorithms. F-Measure gives equal weight to precision and recall, both important to evaluate the task of detecting overdose and underdose prescriptions.

J. Parameter Tuning

All outlier algorithms analyzed were sensitive to the parameter that defines the frontier between normal and abnormal observations [35]. Besides, each type of medication has its own distribution, making the frontier particular for each drug. To ensure the best F-Measure, we performed a simple parameter search, varying it for each algorithm and each medication. In all cases, we made a good faith effort to maximize the performance of all methods.

All algorithms allow only a specific range for the frontier parameter (some call it a contamination parameter). Below we list the interval searched for each set of algorithms:

- For Local Outlier Factor, Isolation Forest, and Robust Covariance, the search ranges from 0.01 to 0.5;
- For One Class SVM, Gaussian Mixture, and DDC, the search ranges from 0.01 to 1.0.
- For every algorithm, we applied a 0.01 step between intervals.

IV. Results

In this section, we discuss the evaluation of the DDC with regards to the adopted baselines in terms of detecting overdoses and underdoses in all 51 medications. We also address the results of the algorithm stability regarding its parameter search and its run-time performance.

In Table II we show the overall results of all algorithm-detecting overdose and underdose outliers in the 51 medications. The DDC-J, density-distance-centrality using the Jaccard similarity, achieved the best mean F-Measure and Top 3 rank. Isolation Forest and DDC using the cosine similarity also achieved good results, both remaining in the Top 3 rank for more than 20 medications.

TABLE II
Mean Performance of Outlier Detection.

Algorithm	Recall	Precision	F-Measure	Top 3 ↑
DDC-J	0.90	0.61	0.68	**31**
Iso. Forest	0.91	0.52	0.61	**26**
DDC	0.86	0.50	0.58	**21**
SVM	0.94	0.39	0.48	20
DDC-C	0.72	0.54	0.51	19
Covariance	0.60	0.37	0.39	16
Gau	0.95	0.29	0.37	13
LOF	0.87	0.38	0.44	12

When counting the best methods, the top 3 rankings discarded the F-Measure when it stood at less than 0.4. Besides, all methods are counted in the Top 3 when there is a tie in the third position.

Covariance, Gaussian Mixture, and Local Outlier Factor had the worst performance in F-Measure. In addition, these three algorithms achieved lower results in the Top 3 rank. Since each medication has its own outlier distribution, some algorithms work better for some medications, but the DDC-J had the best overall result.

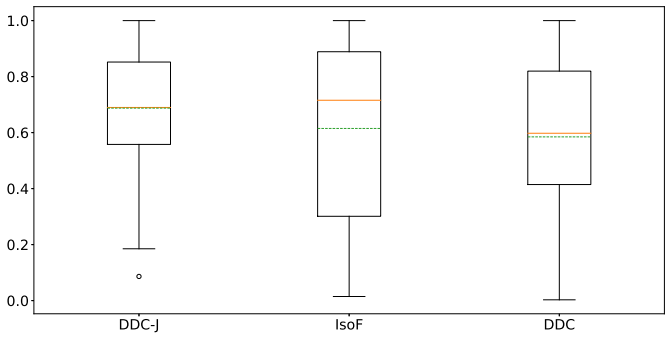

Fig. 2. Distribution of F-Measure obtained in Top 3 Algorithms. Solid lines are the medians and dash lines are the means.

In Figure 2 we show the F-Measure distribution for all medications in each of the Top 3 algorithms in Table II. The DDC-J also had lower variation in comparison with Isolation Forest. In the next section, we cover some insights about the algorithms' run time.

A. Run-Time Evaluation

Some anomaly detection algorithms are very time consuming when the dataset is greater than ten thousand instances. We developed an experiment to evaluate the algorithm's time performance. We selected two medications with 30 thousand instances, ran every algorithm from 3K to 30K instances with 3K steps, and computed the time spent on these medications.

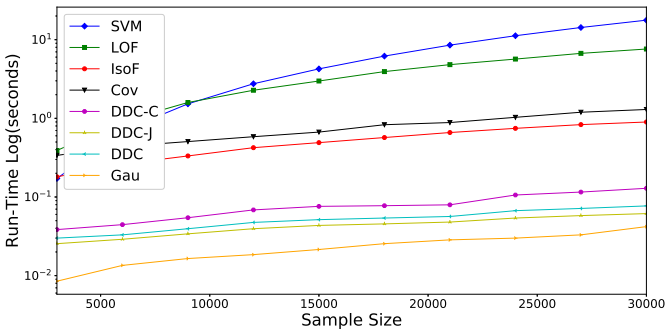

Fig. 3. Run-time comparison between the DDC and baselines for two medications with 30 thousand instances.

For the task of detecting outliers in a large historical dataset, the DDC algorithm has an important scalability property to perform this analysis. Figure 3 shows that One-class SVM and Local Outlier Factor have an exponential time consumption when the data size grows. The time in seconds is represented in log scale on the y-axis. This experiment shows that these algorithms are not fit for big data analysis.

B. Algorithm Stability

The frontier is a sensible parameter that needs to be set in outlier algorithms. With this aspect in mind, we performed an experiment to evaluate the stability of the frontier parameter for each algorithm.

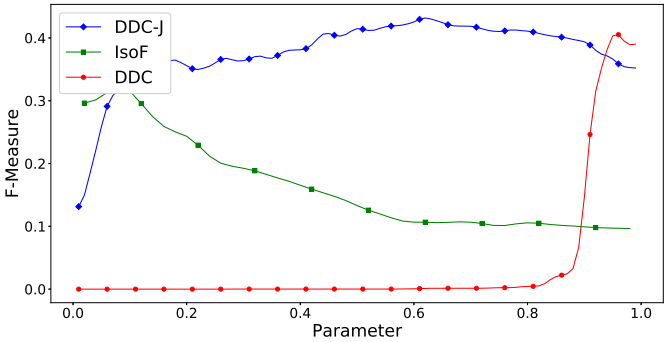

Fig. 4. Influence of parameters on F-Measure results for the Top 3 algorithms

In Figure 4 we show the mean F-Measure considering all parameters for each Top 3 algorithms in Table II. Regarding the DDC-J, it presents the lowest variation using all possibles values for the α parameter. Nevertheless, when $\alpha > 0.1$ it achieves the best results in comparison to the baselines. Regarding the DDC, it achieves better results when $\alpha > 0.9$.

C. Parameter Regression Estimation

All outlier algorithms are very sensitive to their parameters for each kind of dataset. To tackle this problem, we performed a regression analysis considering several statistical information on the medication distribution to estimate the best parameter.

The following statistics were used in the regression analysis: mean, standard deviation, median, and percentile at 20, 50, and 75. All statistics were computed across dose, frequency and both (dose and frequency). The best parameter found for each algorithm was used as the regression target.

Despite all regression algorithms evaluated and the efforts to combine specific sets of medications, the parameter obtained in the regression drastically decreased all algorithm performances regarding F-Measure in overdose/underdose experiments.

V. Discussion

Qualitative assessments were performed by a pharmacist, and the different reasons why a prescription might be outside the pattern were investigated. The F-Measure considered only overdoses/underdoses as true positives. Therefore, a more thorough analysis was required to assess the other cases of outliers. The algorithms could detect other prescriptions to be improved by the pharmacy service department. In the list below, we included a few examples of prescriptions detected by the algorithms, which were not overdoses/underdoses but whose singularity indicates other problems that should also be avoided and could be reviewed by double checking.

- Prescriptions with more suitable presentations for the prescribed dose. It was not necessary to split tablets or even dispose of medications unnecessarily (e.g.: Amlodipine 10 mg, 5 mg prescribed - Amlodipine 5 mg was available at the hospital);
- Prescriptions whose right dose is difficult to administer (e.g.: Levothyroxine 100mcg, a dose of 88 mcg was prescribed);
- Prescriptions with unusual frequency (e.g.: Meropenem 2g once a day)
- Unusual frequencies, making compliance difficult (e.g.: Hydralazine 50 mg 6x/day);
- Unusual dose (Allopurinol 600 mg once a day);
- Half-dose prescription of medications that should not be split (e.g.: Hydralazine 50 mg, with a 25 mg dose prescribed).

The performance of the algorithms, in general, regarding the outlier detection of drug prescriptions with homogeneous prescription distribution (e.g.: Doxazosin tablet, Enalapril 20 mg tablet - Figure 5) was better than for drugs with more sparse prescription distribution - dose x frequency (e.g.: Potassium chloride oral solution, Carbamazepine oral suspension - Figure 6).

In addition to being better in the outlier detection of prescriptions by assessing the F-Measure (with overdoses/underdoses as true positives), the DDC-J algorithm was also able to detect more cases as those described above. Therefore, the algorithm proved to be suitable to generate warnings for double checking, improving prescriptions and providing greater safety for patients. The DDC-J featured a good recall but poor precision (for cases of overdose/underdose); however, this characteristic allows it to select more prescriptions that are potentially incorrect or likely to be improved.

A. Limitations

Our experiments focused on the ability of the proposed algorithms to detect outlier instances among hospital prescriptions. However, any approach that seeks medication errors isolating

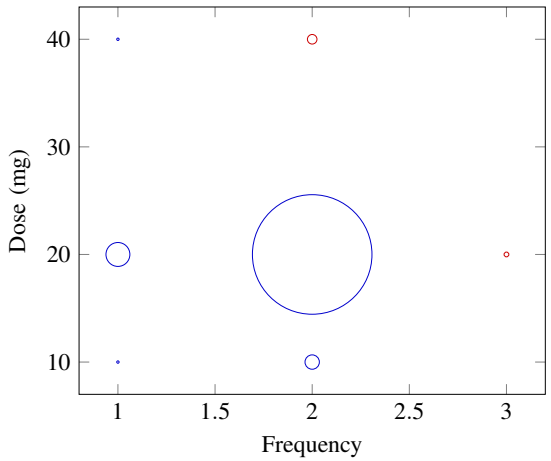

Fig. 5. Prescription of Enalapril 20mg cp, where blue points mean normal prescriptions and red points indicate outliers.

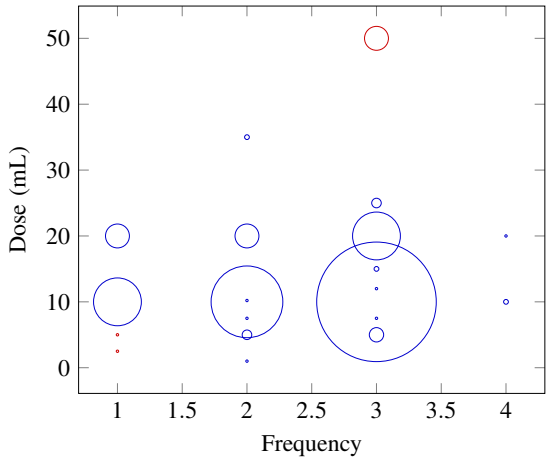

Fig. 6. Prescription of Carbamazepine 20mg/mL oral suspension, where blue points mean normal prescriptions and red points indicate outliers.

the drug from the prescription, as we do, inherently has some limitations. For example, drug-to-drug interactions and duplicated/redundant therapy problems must use all prescription data to handle this task. Therefore, our method should be construed as part of possible mistakes in prescriptions.

Some medications are usually prescribed taking into account patient weight. In that case, the outlier algorithm should use patient information to detect relative overdoses/underdoses considering each patient's characteristics. Nonetheless, because of the simplicity of the pre-processing pipeline and the lack of information on patients' weights in the HNSC dataset, we discarded any medicines that depend on weight.

Another limitation of our study is the outlier assumption in our experiments. All evaluation analyses set overdoses/underdoses as the unique type of outlier problem in medication errors. Nevertheless, in a future experiment, an annotated dataset could be used to evaluate other kinds of outliers as potential errors.

Last, the outlier detection is not intended to assess prescribing associated with disease or clinical conditions of the patient, but only to alert pharmacists of critical prescriptions that may contain an error. Non-standard prescriptions will be reassessed by pharmacists, who may then identify an error or only an unusual dose that is acceptable for the specific condition of that patient. It is a tool to help and to set priorities so that clinical pharmacists can focus on their activities.

VI. Conclusion

A novel algorithm was developed to identify prescription outliers in electronic medical records. This algorithm can rapidly create a distribution pattern in a graph structure to detect anomaly prescriptions. A comparative experiment was conducted against five state-of-the-art algorithms to detect overdose/underdose prescriptions. Moreover, a qualitative analysis indicated the proposed method with Jaccard similarity as the best approach to handle potential prescription errors.

There is a great advantage for the algorithm to learn the distribution from the institution's prescriptions. These characteristics allow the use of this method anywhere in the world with drug-specific standardization, self-adapting to the specifics of each hospital and being able to be automatically updated.

All the content of the work (algorithm, sample dataset, and experiments) is available at the project's GitHub Page[2] in order to be easily replicated. The sample dataset has no patient data; it contains only a subset of the medication dataset, with information on dose, frequency, overdose, and underdose.

This approach fits the prescription in a bi-dimensional space, using just the dose and frequency of each medication. Nonetheless, the same algorithm may be used to identify outliers with a multi-dimensional space, using the route and also the free-text comments that appear in the prescriptions. Based on this, an intelligent screening system could be deployed at the hospital to support the decisions of the pharmacy service.

Besides, the graph strategy could be applied to prescription data attached with laboratory results to track the shortest path between the disease diagnosis and its resolution. This approach can help physicians identify the best treatment prescribed for a given disease, leading to patient satisfaction and less length of hospital stay and increasing health care efficiency.

References

[1] M. B. Buntin, M. F. Burke, M. C. Hoaglin, and D. Blumenthal, "The benefits of health information technology: a review of the recent literature shows predominantly positive results," *Health affairs*, vol. 30, no. 3, pp. 464–471, 2011.

[2] M. Rotmensch, Y. Halpern, A. Tlimat, S. Horng, and D. Sontag, "Learning a health knowledge graph from electronic medical records," *Scientific Reports*, vol. 7, 2017.

[3] R. Hillestad, J. Bigelow, A. Bower, F. Girosi, R. Meili, R. Scoville, and R. Taylor, "Can electronic medical record systems transform health care? potential health benefits, savings, and costs," *Health affairs*, vol. 24, no. 5, pp. 1103–1117, 2005.

[4] D. M. Ashcroft, P. J. Lewis, M. P. Tully, T. M. Farragher, D. Taylor, V. Wass, S. D. Williams, and T. Dornan, "Prevalence, nature, severity and risk factors for prescribing errors in hospital inpatients: prospective study in 20 uk hospitals," *Drug safety*, vol. 38, no. 9, pp. 833–843, 2015.

[5] A. Sheikh, N. Dhingra-Kumar, E. Kelley, M. P. Kieny, and L. J. Donaldson, "The third global patient safety challenge: tackling medication-related harm," *Bulletin of the World Health Organization*, vol. 95, no. 8, p. 546, 2017.

[2]https://github.com/nlp-pucrs/prescription-outliers