POLITECNICO DI TORINO
Repository ISTITUZIONALE

Digging deep into weighted patient data through multiple-level patterns

Original
Digging deep into weighted patient data through multiple-level patterns / Baralis, ELENA MARIA; Cagliero, Luca; Cerquitelli, Tania; Chiusano, SILVIA ANNA; Garza, Paolo. - In: INFORMATION SCIENCES. - ISSN 0020-0255. - STAMPA. - 322:(2015), pp. 51-71. [10.1016/j.ins.2015.06.006]

Availability:
This version is available at: 11583/2615624 since: 2015-07-28T19:17:08Z

Publisher:
Elsevier

Published
DOI:10.1016/j.ins.2015.06.006

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

12 October 2024

Digging deep into weighted patient data through multiple-level patterns

Elena Baralis, Luca Cagliero*, Tania Cerquitelli, Silvia Chiusano, Paolo Garza

Dipartimento di Automatica e Informatica, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129, Torino, Italy

Abstract

Large data volumes have been collected by healthcare organizations at an unprecedented rate. Today both physicians and healthcare system managers are very interested in extracting value from such data. Nevertheless, the increasing data complexity and heterogeneity prompts the need for new efficient and effective data mining approaches to analyzing large patient datasets. Generalized association rule mining algorithms can be exploited to automatically extract hidden multiple-level associations among patient data items (e.g., examinations, drugs) from large datasets equipped with taxonomies. However, in current approaches all data items are assumed to be equally relevant within each transaction, even if this assumption is rarely true.

This paper presents a new data mining application targeted to patient data analysis. It tackles the issue of extracting generalized rules from weighted

*Corresponding author. Tel.: +39 011 090 7084. Fax: +39 011 090 7099.

Email addresses: `elena.baralis@polito.it` (Elena Baralis), `luca.cagliero@polito.it` (Luca Cagliero), `tania.cerquitelli@polito.it` (Tania Cerquitelli), `silvia.chiusano@polito.it` (Silvia Chiusano), `paolo.garza@polito.it` (Paolo Garza)

Preprint submitted to

patient data, where items may weight differently according to their importance within each transaction. To this aim, it proposes a novel type of association rule, namely the Weighted Generalized Association Rule (W-GAR). The usefulness of the proposed pattern has been evaluated on real patient datasets equipped with a taxonomy built over examinations and drugs. The achieved results demonstrate the effectiveness of the proposed approach in mining interesting and actionable knowledge in a real medical care scenario.
Keywords: Generalized Association Rule Mining, Weighted Data Mining, Medical Data

1. Introduction

In today's world large volumes of data have continuously been generated during patient care. However, from the analysis of medical datasets low profits can be made unless physicians and healthcare system managers become able to automatically gain actionable knowledge from potentially large data collections. Patient data analysis is attractive for both physicians, who can use new automatic tools for patient care and healthcare system management, and computer scientists, who can tackle the challenging issue of applying novel data mining techniques to real datasets characterized by an inherent sparseness.

Data mining techniques focus on studying algorithms to find implicit, previously unknown, and potentially useful information from data. In the context of medical care existing data mining approaches encompass different analyses, such as mining underlying associations among data items (e.g., [6, 20, 34, 38, 40]), clustering (e.g., [4, 5, 48]), and classification

(e.g.,[26, 33]). However, the increasing complexity and heterogeneity of medical data prompts the need for novel and effective approaches to automatically mining actionable knowledge. This knowledge can be exploited, for example, to improve the current patient care processes, to assess new medical guidelines, or to enrich existing ones.

In the last few years the use of exploratory techniques to discover hidden correlations among medical data items has received great attention by the research community. To discover valuable multiple-level correlations among data equipped with taxonomies, generalized association rule mining techniques [41] can be easily exploited. A taxonomy, i.e., a set of is-a hierarchies that aggregate data items into higher-level concepts, is used to analyze co-occurrences among data items at different abstraction levels. A generalized association rule is an implication in the form $A \rightarrow B$, where A and B are disjoint sets of generalized items belonging to the taxonomy. The rule is characterized by its frequency of occurrence in the dataset, which is called support, and by the strength of the implication, called confidence.

A major drawback of traditional association rule mining approaches is that all data items are assumed to be equally relevant within the analyzed data, even though in many application domains this assumption is not true. For example, in the medical context prescribed examinations and drugs have not all the same importance in patient care. To overcome this issue, weighted datasets can be analyzed. A weighted dataset is a dataset in which to each item a weight, denoting its relative importance in the corresponding transaction, is assigned. Some research efforts (e.g., [44, 47, 49]) have been made to consider also item weights during the association rule mining process. How-

ever, to the best of our knowledge, no generalized rule mining algorithm is currently able to successfully cope with weighted data. Therefore, there is a need for innovative solutions to discover interesting patterns at different abstraction levels from weighted data.

The main contributions of this paper can be summarized as follows:

- It presents a new data mining application, named Weighted Patient Data Analyzer (WeP-DatA), targeted to patient data analysis.
- It proposes a novel type of generalized association rule, namely the Weighted Generalized Association Rule (W-GAR), tailored to weighted data.
- It considers two different measures for weighting data items, which are targeted to different use cases.
- As a case study, the proposed approach has been applied to the medical care scenario to demonstrate the effectiveness of W-GARs in discovering interesting and actionable knowledge on real data.

W-GARs are a new type of generalized association rules, which also consider item weights during rule evaluation. W-GAR extraction entails the typical two-step process: (i) Frequent generalized weighted itemset mining from weighted datasets by enforcing a minimum weighted support threshold *minwsup*, and (ii) frequent W-GAR extraction, starting from the previously mined itemsets, by enforcing a minimum weighted confidence threshold *minwconf*. To filter out rules that contain irrelevant information during W-GAR extraction, itemset occurrences within each weighted transaction are

weighted by the least item weight. In such a way, we guarantee that all generalized items in a W-GAR have a minimal relevance score within each transaction.

The proposed approach has been applied to a real dataset of diabetic patients provided by the National Health Center of an Italian province. Two different weighting measures (i.e., simple frequency and tf-idf [32] of prescriptions) were tested. The experiments demonstrate that, starting from a large collection of raw patient weighted data, W-GARs represent interesting multiple-level associations among patient treatments, which are hardly inferable using traditional rules. The results were validated by clinical domain experts. The extracted rules appear to be consistent with the guidelines for diabetes disease [1, 23, 24].

This paper is organized as follows. Section 1.1 presents a motivating example, to exemplify the main advantages of the new kind of proposed rules in the specific context under analysis. Section 2 presents the Weighted Patient Data Analyzer environment. Section 3 assesses the effectiveness of the system in performing knowledge discovery from a real diabetic patient dataset. Section 4 compares our approach with most relevant related works, while Section 5 draws conclusions, presents future developments of this work and envisions different use case scenarios where the newly proposed rules may be profitability exploited to support advanced analyses.

1.1. Motivating example

Let us consider the dataset and taxonomy reported in Table 1 and Figure 1, respectively, which come from the medical domain. The dataset consists of 4 transactions. Each transaction corresponds to a different patient,

Table 1: Example of unweighted transactional dataset

Patient id	Examinations and drug prescriptions
1	$(Exam : Glucose)$ $(Drug : Pantoprazole)$ $(Drug : Acetylsalicylic\ Acid)$
2	$(Exam : Glucose)$ $(Drug : Pantoprazole)$ $(Drug : Moxifloxacin)$
3	$(Exam : Glucose)$ $(Drug : Pantoprazole)$ $(Drug : Acetylsalicylic\ Acid)$
4	$(Exam : Glucose)$ $(Drug : Pantoprazole)$ $(Drug : Moxifloxacin)$
5	$(Exam : Glucose)$ $(Drug : Omeprazole)$ $(Drug : Acetylsalicylic\ Acid)$

Table 2: Example of weighted transactional dataset. Measure: number of prescriptions

Patient id	Examinations and drug prescriptions
1	$\langle Exam : Glucose, 6\rangle$ $\langle Drug : Pantoprazole, 1\rangle$ $\langle Drug : Acetylsalicylic\ Acid, 10\rangle$
2	$\langle Exam : Glucose, 2\rangle$ $\langle Drug : Pantoprazole, 2\rangle$ $\langle Drug : Moxifloxacin, 2\rangle$
3	$\langle Exam : Glucose, 4\rangle$ $\langle Drug : Pantoprazole, 2\rangle$ $\langle Drug : Acetylsalicylic\ Acid, 5\rangle$
4	$\langle Exam : Glucose, 6\rangle$ $\langle Drug : Pantoprazole, 3\rangle$ $\langle Drug : Moxifloxacin, 2\rangle$
5	$\langle Exam : Glucose, 2\rangle$ $\langle Drug : Omeprazole, 5\rangle$ $\langle Drug : Acetylsalicylic\ Acid, 4\rangle$

which is identified by the respective patient id (Pid). Transactions contain the examinations undergone and the drugs prescribed to each patient during the last year. The taxonomy in Figure 1 generalizes examinations and drugs (e.g., *(Exam:Glucose)*, *(Drug:Acetylsalicylic Acid)*) as the corresponding higher-level categories (e.g., *(Exam:Routine)*, *(Drug:Analgesic)*). $\{$*(Drug:Analgesic)*$\} \rightarrow \{$*(Drug:Omeprazole)*$\}$ and $\{$*(Drug:Analgesic)*$\} \rightarrow \{$*(Drug:Pantoprazole)*$\}$ are examples of generalized rules. These rules can be used to decide which protective drug recommend to patients who have been treated with Analgesics for a long time. For example, since the latter rule has a higher confidence than the former one (i.e., $\frac{2}{3}$ versus $\frac{1}{3}$), drug Pantoprazole should be recommended first.

A weighted version of the dataset in Table 1 is reported in Table 2. In a weighted dataset each item has a weight, denoting its relative importance in the corresponding transaction. For example, the patient with Pid 5 has un-

Table 3: Example of weighted transactional dataset. Measure: tf-idf of prescriptions

Patient id	Examinations and drug prescriptions
1	$\langle Exam : Glucose, 0\rangle$ $\langle Drug : Pantoprazole, 0.032\rangle$ $\langle Drug : Acetylsalicylic\ Acid, 0.7395\rangle$
2	$\langle Exam : Glucose, 0\rangle$ $\langle Drug : Pantoprazole, 0.065\rangle$ $\langle Drug : Moxifloxacin, 0.1479\rangle$
3	$\langle Exam : Glucose, 0\rangle$ $\langle Drug : Pantoprazole, 0.065\rangle$ $\langle Drug : Acetylsalicylic\ Acid, 0.3698\rangle$
4	$\langle Exam : Glucose, 0\rangle$ $\langle Drug : Pantoprazole, 0.097\rangle$ $\langle Drug : Moxifloxacin, 0.1479\rangle$
5	$\langle Exam : Glucose, 0\rangle$ $\langle Drug : Omeprazole, 1.165\rangle$ $\langle Drug : Acetylsalicylic\ Acid, 0.2958\rangle$

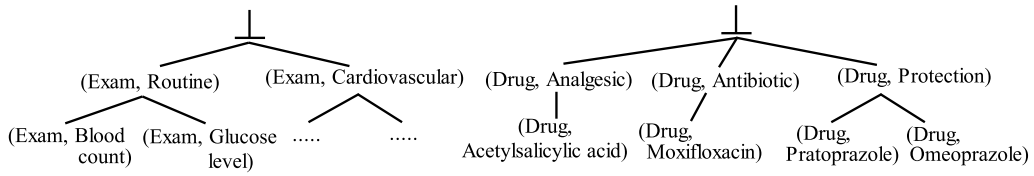

Figure 1: Example of taxonomy built over the transactional patient datasets

dergone Glucose test 2 times and has taken Omeoprazole Acetylsalicylic Acid 5 and 4 times, respectively.

W-GARs, the new type of generalized association rules proposed in this paper, consider item weights during rule evaluation. To filter out rules that contain irrelevant information during W-GAR extraction, all generalized items in a W-GAR have a minimal relevance score within each transaction. Specifically, itemset occurrences within each weighted transaction are weighted by the least item weight. For example, W-GAR {*(Drug:Analgesic)*} $\rightarrow$ {*(Drug:Omeprazole)*} has weighted support equal to 4, because the least weighted item in the transaction with Pid 5 has weight 4. Let us consider now rules {*(Drug:Analgesic)*} $\rightarrow$ {*(Drug:Omeprazole)*} and {*(Drug:Analgesic)*} $\rightarrow$ {*(Drug:Pantoprazole)*}. Both rules have the same antecedent but a different consequent and they can be used to decide which supplements could be recommended to patients who are treated with analgesics. It is worth notic-

ing that, by considering item weights, W-GAR ranking in order of confidence is reversed with respect to those of traditional rules ($\frac{4}{19}$ with Omeoprazole versus $\frac{3}{19}$ with Pantoprazole). Hence, recommending supplementary drugs according to the confidence value of traditional rules could be misleading.

2. Weighted patient data analyzer

Weighted Patient Data Analyzer (WeP-DatA) is a new data mining environment for the advanced analysis of medical data related to the history of examinations' and drugs' prescriptions to patients. WeP-DatA focuses on supporting physicians and healthcare system managers in discovering interesting and actionable knowledge from large patient data collections.

Currently, the main challenges in effectively coping with real patient data are: (i) the intrinsic sparseness of the analyzed datasets, which typically contain a very large number of different examinations and drugs and (ii) the inability of some data mining algorithms (e.g., the generalized association rule mining algorithms [9, 21, 41]) in coping with data equipped with item weights. To overcome these issues, WeP-DatA generates a new type of association rule [2], called Weighted Generalized Association Rule (W-GAR).

This section is organized as follows. Section 2.1 thoroughly describes the context under study, i.e., the analysis of real patient datasets. Section 2.2 discusses the problem of taxonomy generation on top of patient data. Taxonomies will be exploited to overcome the limitations of traditional association rule mining algorithms in coping with sparse datasets (the issue (i) mentioned above). Section 2.3 better clarifies the semantics used within this study for item weight assignment. Finally, Section 2.4 introduces the concept

of W-GAR and it describes in detail the process of W-GAR mining, which can be successfully applied to weighted patient data (issue (ii)).

2.1. Context of analysis

Healthcare systems usually collect heterogeneous medical information into potentially large datasets. To allow physicians to keep track of diagnosis and therapies, patient data are commonly stored into separate log files, which are then integrated into common repositories. For example, physicians commonly record examinations and drugs prescribed to each patient to analyze the temporal evolution of patients' state of health and to perform further analyses. Furthermore, the same information is also deemed worthy by healthcare system managers in charge of purchasing medical equipments or planning resource allocations.

Hereafter we will focus our analysis on two main characteristics of patient data:

- the undergone examinations and
- the prescribed drugs.

To perform data mining analyses, patient data are tailored to a weighted transactional data format. A weighted patient dataset is a set of weighted transactions, where each weighted transaction corresponds to a different patient and it consists of a set of pairs ⟨item, weight⟩ called *weighted items*. Items are related to examinations (e.g., *Glucose level*) or drugs (e.g., *Acetylsalicylic Acid*) and they are represented in the form (*feature*:*value*), where *feature* is *Examination* or *Drug* while *value* is the corresponding feature value.

Weights indicate the item relevance within the corresponding transaction. The semantics used for assigning item weights in our context of analysis will be thoroughly discussed in Section 2.3. A more formal definition of weighted patient dataset is given below.

Definition 2.1. Weighted patient dataset. *Let E be the set of all possible patient examinations, and M the set of all possible drugs. An item i_k is a pair (feature,v_q), where $v_q \in E$ if feature is equal to Examination or $v_q \in M$ if feature is equal to Drug. A weighted item is a pair $\langle i_k, w_k \rangle$, where w_k is a real number denoting item relevance. A weighted patient dataset $\mathcal{D}$ is a set of weighted transactions, where each weighted transaction $t_j \in \mathcal{D}$ is a set of weighted items $\{\langle i_k^j, w_k^j \rangle\}$.*

For instance, Table 2 reports an example of weighted patient dataset. For each patient examinations and drugs are weighted by the corresponding number of prescriptions (e.g., the patient with Pid 5 has taken Omeoprazole 5 times, while drug Acetylsalicylic Acid was prescribed 10 times to the patient with Pid 1). According to Definition 2.1, weighted item $\langle Exam : Glucose, 6 \rangle$, which occurs in the transaction with Pid 1, indicates that the corresponding patient has undergone the glucose level test 6 times.

2.2. Taxonomy generation

Real-world data are intrinsically sparse. For example, patient data often contain hundreds of examinations and thousands of drugs. Hence, the analysis of the raw data could be challenging because most relevant information can remain undisclosed at first glance.

Analyzing data at different abstraction levels allows experts to discover interesting and actionable knowledge which may remain hidden at the lowest granularity level. This approach is commonly used in data warehousing and data mining analyses. For example, OLAP analyses commonly exploit hierarchies built over the schema dimensions to perform aggregation and drill down operations over historical data [27].

We target the discovery of associations among patient data items at different arbitrary levels by means of generalized association rule discovery [41]. To our purpose, data are equipped with analyst-provided taxonomies. Taxonomies built over weighted patient data aggregate examinations and drugs into high-level concepts, i.e., examinations are generalized as examination categories and drugs as drug categories. A more formal definition follows.

Definition 2.2. Taxonomy. *Let $\mathcal{D}$ be a weighted patient dataset and $\mathcal{I}$ the set of items in $\mathcal{D}$ (disregarding the corresponding weights). A generalization hierarchy $GH_{\mathcal{I}_k}$ ($\mathcal{I}_k \subseteq \mathcal{I}$) built over $\mathcal{D}$ is a hierarchy of aggregations defined over a subset of items in $\mathcal{I}$, where hierarchy leaves are items in $\mathcal{I}$, while non-leaf nodes in $GH_{\mathcal{I}_k}$ are ancestors of their corresponding children. Each hierarchy has a root node (denoted as $\perp$) which aggregates all its items. A taxonomy $\mathcal{T}$ built over $\mathcal{D}$ consists of a set of generalization hierarchies $GH_{\mathcal{I}_k}$ for which $\cup_{GH_{\mathcal{I}_k} \in \mathcal{T}} \mathcal{I}_k = \mathcal{I}$.*

Although taxonomies can potentially contain many generalizations over the same item (e.g., many categories for the same examination), for the sake of simplicity hereafter we will consider only taxonomies containing at most one generalization per item.

Given a taxonomy $\mathcal{T}$ built over a weighted patient dataset $\mathcal{D}$, a weighted generalized item is a pair $\langle gi_k, W_k \rangle$, where gi is a non-leaf node in $\mathcal{T}$ (also called generalized item), while W_k is its corresponding weight. To mine rules at different abstraction levels in a single extraction, generalized rule mining algorithms handle both generalized and non-generalized items.

An example of taxonomy built over the running example dataset is reported in Figure 1. Examinations *Blood count* and *Glucose level* are classified as *Routine*, whereas drugs *Pantoprazole* and *Omeoprazole* are generalized as *Protection*. Finally, drugs *Acetylsalicylic Acid* and *Moxifloxacin* are classified as *Analgesic* and *Antibiotic*, respectively. (*Drug*:*Analgesic*) is an example of generalized item, while $\langle$*Drug*:*Analgesic*, 5$\rangle$ is an example of weighted generalized item, which indicates that the weight of item (*Drug*:*Analgesic*) is 5.

To define aggregations over examinations and drugs, analysts could exploit standard classification systems, e.g., for drugs the ATC classification system available at [7]. Even though our data model currently considers only examinations and drugs, it can be easily extended by inserting data related to additional patient features (e.g., social status, job, phenotype) and the corresponding high-level categories.

2.3. Weight semantics

Since real-life data analyses are commonly targeted to multiple data facets and measures, item weights in different datasets could potentially represent different information. We focus our empirical studies on two representative measures tailored to patient data analysis:

A) the number of the drug and examination prescriptions and

B) the tf-idf of the drug and examination prescriptions.

A more detailed description of each measure is given below.

Measure (A): number of prescriptions. Given a weighted transaction $t_j \in \mathcal{D}$, the weight w_k^j of weighted item $\langle i_k^j, w_k^j \rangle$ indicates the number of prescriptions of the examination/drug associated with i_k to the patient corresponding to transaction t_j.

Table 2 reports an example of weighted patient dataset in which weights are defined according to measure (A). For example, since patient with Pid 5 has taken Omeoprazole 5 times, item (*Drug:Omeprazole*) has weight 5. The use of this measure allows us to figure out interesting and hidden recurrences in the history of drug and examination prescriptions. For instance, such information is worth considering by healthcare system managers to efficiently set up and manage disease prevention protocols or to plan resource allocations.

Since generalized items represent either drug categories or examination categories, we are interested in analyzing the cumulative number of prescriptions per category and patient. Hence, given a weighted transaction $t_j \in \mathcal{D}$, the weight W_k^j of weighted generalized item $\langle gi_k^j, W_k^j \rangle$ is the number of prescriptions for the patient corresponding to t_j of the examinations/drugs belonging to the category represented by generalized item gi.

For example, according to the dataset in Table 2 and the taxonomy in Figure 1, the weight associated with generalized item (*Examination*:*Routine*) in the transaction with Pid 1 is 6, because the transaction contains just one of its descendant items, i.e., (*Examination*:*Glucose examination*), and its corresponding weight is 6.

Measure (B): tf-idf of prescriptions. Given a weighted transaction $t_j \in \mathcal{D}$, the weight w_k^j of weighted item $\langle i_k^j, w_k^j \rangle$ expresses the term frequency-inverse document frequency (tf-idf) statistics related to the examination/-drug associated with i_k and to the patient corresponding to transaction t_j.

The term frequency-inverse document frequency (tf-idf) index is an established statistic frequently used to analyze textual documents. In our context, each patient is equivalent to a document and each examination/drug is equivalent to a word.

Table 3 reports an example of weighted patient dataset in which weights are defined according to measure (B).

The use of this measure is aimed at discovering combinations of examinations/drugs that have frequently been prescribed together to few patients. Such patterns are worth considering to highlight peculiar prescriptions related to specific patient clusters, e.g., patients with specific diseases or profiles.

The tf-idf evaluator [32] is usually expressed in matrix form [45]. Let TI be the tf-idf matrix for weighted patient dataset $\mathcal{D}$, where each row represents a distinct patient (i.e., a distinct weighted transaction), while each column corresponds to a distinct examination or drug (i.e., an item). Each element ti_{jk} of the tf-idf matrix TI combines the frequency of the k-th item in the j-th transaction with the inverse of the logarithm of its transaction frequency in $\mathcal{D}$. The tf-idf matrix value ti_{jk} can be expressed as follows:

$$ti_{jk} = \frac{n_{jk}}{|t_j|} \cdot \log \frac{|\mathcal{D}|}{|\{t_j \in \mathcal{D} \ : \ \mathrm{i_k} \in \mathrm{t_j}\}|} \tag{1}$$

where n_{jk} is the number of prescriptions of the examination/drug corresponding to the k-th item i_k for the patient associated with transaction t_j, $\mathcal{D}$ is the weighted patient dataset, $|t_j|$ is the number of items that are contained in the j-th transaction t_j, and $\frac{|\mathcal{D}|}{|\{t_j \in \mathcal{D} \,:\, \mathrm{i_k} \in \mathrm{t_j}\}|}$ represents the logarithm of the inverse of the fraction of transactions in which item i_k occurs in the whole dataset, i.e., the inverse of the frequency of patients to whom the drug/examination corresponding to i_k has been prescribed at least once.

The logarithm is minimal when the inverse transaction frequency is equal to 1 (i.e., when a drug/examination has been prescribed to all patients in the datasets). In such a case, the corresponding td-idf value reduces to zero. For example, item (*Exam*:*Glucose*) has tf-idf weight equal to zero in Table 3 because the exam has been prescribed at least once to all patients and thus the idf component of the tf-idf statistics is reduced to zero. Conversely, a high tf-idf value indicates that the specific examination/drug has frequently been prescribed to few patients. For example, item (*Exam*:*Omeoprazole*) has the highest tf-idf value in Table 3 (1.165) because it has been prescribed many times (5) to only one out of five patients in the dataset.

2.4. Generalized weighted association rule mining

We focus on discovering a new type of association rule, i.e., the Weighted Generalized Association Rule (W-GAR), from weighted patient datasets equipped with taxonomies (see Definitions 2.1 and 2.2, respectively).

This section is organized as follows. Section 2.4.1 introduces preliminary concepts related to the traditional (unweighted) generalized association rule mining problem, while Section 2.4.2 formally introduces the concept of W-GAR and it thoroughly describes the W-GAR mining task addressed by this

paper.

2.4.1. Generalized association rules

Generalized association rule mining [41] is a widely exploratory data mining technique to discover hidden and multiple-level correlations among large datasets equipped with (analyst-provided) taxonomies. A generalized association rule is an implication $A \rightarrow B$, where A and B are disjoint sets of generalized or not generalized items, also called generalized k-itemsets. A (generalized) k-itemset is set of (generalized) items of size k. In the following A and B will be also denoted as antecedent and consequent, respectively.

In our context of analysis, generalized itemsets are arbitrary sets of examinations, drugs, examination categories, or drug categories. Generalized rules express implications between examinations or drugs, possibly at different abstraction levels. Generalized rules that contain only examinations or drugs, i.e., items at the lowest granularity level, will be denoted hereafter as *low-level* rules, rules containing only examination/drug categories will be denoted as *high-level* rule. Finally, rules that contain a mixture of generalized and not generalized items will be denoted as *cross-level* rules. For example, let us consider again the dataset in Table 1 and the taxonomy in Figure 1. {*(Exam:Cardiovascular)*} $\rightarrow$ {*(Exam:Glucose)*} is an example of cross-level generalized rule which indicates that cardiovascular examinations are frequently prescribed in conjunction with a specific routine examination, i.e., the glucose level test.

Since generating all the possible itemsets and rules is computationally intractable [2] and it would require experts to deal with a very large set of (potentially redundant) patterns, generalized association rule extraction

typically entails the following two steps [41]:

- Frequent generalized itemset mining, which addresses the extraction of all generalized itemsets that frequently occur in $\mathcal{D}$, i.e., generalized itemsets whose support value is above a given threshold $minsup$ and

- Strong generalized association rule extraction, which entails generating, from the subset of previously mined itemsets, all generalized rules that frequently occur and that hold in most cases in $\mathcal{D}$, i.e., generalized rules whose support value is above a given threshold $minsup$ and whose confidence value is above a given threshold $minconf$.

The support of a generalized itemset indicates its observed frequency of occurrence in the source dataset. If an item is generalized, its occurrence in a transaction is counted if and only if any of its descendant item (according to the input taxonomy) occurs.

For example, the generalized 2-itemset {*(Exam:Routine)*, *(Drug:Omeprazole)*} has (absolute) support value equal to 1 in Table 1, because it occurs only in the transaction with Pid 5 (the occurrence of exam category Routine is due to those of exam Glucose).

Hereafter we formally introduce the two main generalized association rule quality indexes, i.e., support and confidence [45].

Definition 2.3. Generalized rule support and confidence. *Let $\mathcal{D}$ be an (unweighted) patient dataset and $A \rightarrow B$ an arbitrary generalized association rule. Let $sup(A)$ and $sup(B)$ be support of A and B in $\mathcal{D}$, respectively.*

- *The support of $A \rightarrow B$, denoted as $sup(A \rightarrow B)$, is defined as the support of generalized itemset $A \cup B$ in $\mathcal{D}$.*

- *The confidence of $A \rightarrow B$, denoted as conf($A \rightarrow B$) is defined as the conditional probability of occurrence of B given the A in $\mathcal{D}$, i.e., conf($A \rightarrow B$)=$\frac{sup(A \cup B)}{sup(A)}$.*

Note that, based on the above definitions, item occurrences in each dataset transaction are treated equally, even if items are not equally relevant within each transaction.

For example, rule {*(Drug:Analgesic)*} $\rightarrow$ {*(Drug:Omeprazole)*} has support equal to 1 in Table 1, because it covers only the transaction with Pid 5. Its confidence value is $\frac{1}{3}$, because only one third of the patients in the dataset who took Analgesic drugs have also taken Omeoprazole.

To extract generalized association rules from transactional datasets many algorithms have already been proposed in literature (e.g., [9, 21, 41]). Unfortunately, to the best of our knowledge, none of them is able to cope with weighted data. In the next section, we overcome this issue by presenting a new type of generalized rule, i.e., the weighted generalized association rule.

2.4.2. Weighted generalized association rules

We present a new type of association rule that allows us to overcome both the following issues at the same time: (i) the sparseness of real-life data and (ii) the inability of state-of-the-art generalized rule mining to cope with weighted data. On the one hand, state-of-the-art association rule mining approaches tailored to weighted data (e.g., [49, 44]) are unable to effectively deal with sparse data, because they discover only associations among data items at the lowest granularity level. On the other hand, generalized association rule mining strategies (e.g., [9, 21, 41]) are currently unable to cope

with weighted data.

Weighted Generalized Association Rules (W-GARs) are generalized rules $A \rightarrow B$ extracted from transactional datasets equipped with item weights. In our context of analysis, we extract generalized rules from weighted patient datasets by considering not only the simple item occurrences in the source dataset but even the weights associated with data items within each transaction. Specifically, to evaluate generalized rule quality indexes, item occurrences within each transaction are weighted by the corresponding weights.

To our purpose, we preliminary extend the definition of generalized itemset support to the case of weighted data. We will denote such a measure as *w-support* [17]. The w-support of a generalized itemset is the sum of the weight of its least weighted item in the itemset for each transaction in which the itemset occurs.

Definition 2.4. Itemset w-support. *Let $\mathcal{D}$ be an weighted patient dataset and I an arbitrary generalized itemset. Let $\mathcal{W}(\mathcal{I}, t_j)$ be the matching weight of a generalized itemset I with respect to tr_j, which is defined as follows:*

$$\mathcal{W}(\mathcal{I}, t_j) = \begin{cases} \min_{k \mid i_k \in I} w_k^j & \textit{if all items } i_k \textit{ in } I \textit{ occur in } t_j, \\ 0 & \textit{otherwise} \end{cases}$$

The w-support of I in $\mathcal{D}$ is the summation of all matching weights of I for every transaction in $\mathcal{D}$:

$$\textit{w-sup}(I) = \sum_{t_j \in \mathcal{D}} \mathcal{W}(\mathcal{I}, t_j)$$

For example, the support of {*(Exam:Routine)*, *(Drug:Omeprazole)*} in Table 2 is 2 because the itemset occurs only in the transaction with Pid 5

and the least weighted item between *(Exam:Routine)* and *(Drug:Omeprazole)* has weight equal to 2.

The concepts of support and confidence of a traditional generalized rule (see Definition 2.3) are extended to W-GARs below. We will denote by w-support and w-confidence the respective W-GAR quality measures.

Definition 2.5. W-GAR w-support and w-confidence. *Let $\mathcal{D}$ be a weighted patient dataset and $A \rightarrow B$ a W-GAR. Let w-sup(A) and w-sup(B) be the w-support of A and B in $\mathcal{D}$, respectively.*

- *The w-support of $A \rightarrow B$, denoted as w-sup($A \rightarrow B$), is defined as the w-support of generalized itemset $A \cup B$ in $\mathcal{D}$.*
- *The w-confidence of $A \rightarrow B$, denoted as w-conf($A \rightarrow B$) is defined as the weighted conditional probability of occurrence of B given the A in $\mathcal{D}$ i.e. w-conf($A \rightarrow B$)=$\frac{w\text{-}sup(A \cup B)}{w\text{-}sup(A)}$*

For example, W-GAR {*(Exam:Routine)*} $\rightarrow$ {*(Drug:Omeprazole)*} has w-support equal to 2 and w-confidence equal to $\frac{2}{20}$=10% because the w-support of the rule antecedent ({*(Exam:Routine)*}) is 20. Note that disregarding item weights the same rule would have a confidence equal to 20%. The gap between the two values is due to the fact that the transaction in which the rule actually occurs (Pid 5) has a relatively low matching weight (2) compared to the others. Hence, the weighted conditional probability of occurrence of *(Drug:Omeprazole)* given *(Exam:Routine)* is lower than the unweighted one.

To discover interesting and actionable multiple-level associations among weighted patient data, we discover and select a worthwhile subset of W-GARs, denoted as *strong WARs*, from the analyzed data.

Definition 2.6. Strong W-GAR. *Strong W-GARs are W-GARs whose*

- *w-support in $\mathcal{D}$ is above a given threshold $minwsup$*
- *w-confidence in $\mathcal{D}$, is above a given threshold $minwconf$*

In Section 3 the strong W-GARs mined from a real weighted patient dataset were validated and some worthy examples of application of the discovered patterns are presented.

To extract strong W-GARs, we adopted the usual two-step process [3], i.e., frequent itemset mining followed by strong association rule extraction. To mine generalized frequent itemsets from weighted data, we adapted the FP-Growth-like [22] weighted itemset mining algorithm implementation, which was first proposed in [17], to generalized itemset mining. Since the algorithm proposed in [17] was designed to mine non-generalized itemsets, we followed the approach previously adopted in [10] to integrate taxonomy information. Specifically, we first extended each dataset transaction by appending the corresponding item generalizations. Then we mined frequent generalized itemsets while preventing the generation of invalid candidate itemsets, i.e., those generalized itemsets that contain both an item and any of its generalizations.

To perform W-GAR mining on top of frequent itemsets, we used our slightly modified implementation of the rule mining step of the Apriori algorithm [3].

2.4.3. Algorithm complexity

The complexity of the W-GAR algorithm is comparable to those of traditional Apriori-based [3] association rule mining algorithms. More specifically,

it is linear in the number of transactions, i.e., $O(n)$, where n is the number of transactions, while it is combinatorial in the number of average items per transaction. The complexity of the rule extraction process is mainly due to the generalized itemset mining step and it mainly depends on the analyzed data distribution. Enforcing a minimum support threshold reduces the number of generated item combinations thus making the mining problem tractable on real data. If no support threshold is enforced during the mining process, the complexity of the process of itemset generation is $O(d \cdot 2^{d-1})$, where d is the number of distinct items in the source data [46]. A thorough analysis of the impact of the support threshold on the characteristics of the mining result is given in Section 3.3.

3. Experimental results

To assess the effectiveness and efficiency of the proposed approach we performed a set of experiments on a real dataset gathered by an Italian Health Center.

This section is organized as follows. Section 3.1 describes the characteristics of the analyzed dataset and the semantics used for assigning item weights. Section 3.2 summarizes the most relevant results and it highlights the significance and usability of the rules discovered with different weighting measures. A comparison between traditional and weighted generalized rules is also reported. Section 3.4 compares the rules extracted by our approach with those mined by a different weighted association rule mining approach. Finally, Section 3.5 evaluates the efficiency of our approach in terms of execution time.

Table 4: Generalization hierarchy over examinations

Examination category	**Examination**
Routine examinations	*Checkup visit* *Glucose level* *Urine test* *Venous blood* *Complete blood count* *Hemoglobin*
Cardiovascular examinations	*Electrocardiogram* *Cholesterol* *HDL Cholesterol* *Triglycerides*
Eye examinations	*Fundus oculi* *Angioscopy* *Complete eye examination* *Retinal photocoagulation*
Liver examinations	*AST* *ALT* *Bilirubin* *Gamma GT*
Kidney examinations	*Urin acid* *Microscopic urine analysis* *Culture urine* *Creatinine clearance* *Creatinine* *Microalbuminuria*
Carotid examinations	*ECO Doppler carotid*
Limb examinations	*ECO Doppler limb*

All the experiments were performed on a quad-core 3.30 GHz Intel Xeon workstation with 16 GB of RAM, running Ubuntu Linux 12.04 LTS.

3.1. Dataset and taxonomy

To evaluate the effectiveness of the proposed approach we analyzed a real dataset collecting the drugs and examinations prescribed to the overt patients with diabetes of an Italian Health Center. The dataset, collected in year 2007, consists of 648,797 records, where each record corresponds to a set of daily examination/drug prescriptions to a given patient. Clearly, each examination/drug can be prescribed several times to the same patient or to different patients. The number of patients under analysis is 8,749. We used

two different weighting measures to evaluate examination/drug importance, i.e., simple frequency and tf-idf of the prescription. To this purpose, we generated one distinct dataset version per measure. Hereafter we will denote the two dataset versions as Diabetes-Freq and Diabetes-tf-idf, respectively. Each dataset contains a set of weighted transactions, one for each patient. Transactions consist of a set of examinations and/or drugs which were prescribed at least once to the corresponding patient. To each examination/drug a weight is assigned by using the appropriate measure, i.e., the number of prescriptions in Diabetes-Freq or the tf-idf measure in Diabetes-tf-idf.

To enable generalized rule mining we exploited two hierarchies built over examinations and drugs, respectively. Table 4 reports the generalization hierarchy defined on the set of examinations under analysis. It contains 26 examinations clustered into 7 examination categories. The examination categories were selected based on the expert-driven classification reported in [5].

To generate a hierarchy over drugs, we exploited the levels of the ATC classification system defined in [7]. Each level represents a different abstraction level of aggregation on the set of considered drugs. More specifically, the fifth level of the code indicates the chemical substance (i.e., the drug), while the chemical subgroup is encoded by the fourth level. The third level indicates the therapeutic/pharmacological subgroup and the second level represents the therapeutic main group of the drug. Finally, the first level of the code indicates the anatomical main group, also called main category in the following. In our experiments we used a generalization hierarchy over drugs whose leaves are the drugs encoded using the fifth level of the ATC classi-

fication system defined in [7]. Drugs are aggregated according to the ATC classification into the upper-level categories. For example, drug B01AC06 (i.e., acetylsalicylic Acid) is a leaf node of the generalization hierarchy and its is generalized as the B01AC group (i.e., platelet aggregation inhibitors excluding heparin). The platelet aggregation inhibitors excluding heparin is then generalized as the B01 group (i.e., Antithrombotic agents), which, in turn, is further aggregated into the main category B (i.e., Category Blood and blood forming organs). Table 5 reports some examples of first- and fifth-level elements of the ATC classification.

3.2. Evaluation of the mined rules

We performed several experiments on the analyzed data to identify most interesting rules. Section 3.2.1 analyzes the W-GARs mined from the Diabetes-Freq dataset, while Section 3.2.2 presents the most interesting W-GARs mined from Diabetes-tf-idf.

In our analyses we separately considered the rules containing examinations and drugs. Furthermore, since drug and examination weights are, in general, not directly comparable we disregarded the W-GARs containing a mixture of drugs and examinations. However, the proposed methodology is general and it can potentially handle item weights with different semantics provided that a compound weighting scheme is used.

To select a manageable subset of potentially interesting strong W-GARs we tested several values of w-support and w-confidence thresholds. Furthermore, to validate rule interestingness we also considered two entropy-based evaluators, i.e., AntInt and ConsInt [35]. Entropy-based evaluators have largely been used to pinpoint most reliable correlations from data items [46].

Table 5: Portion of the generalization hierarchy over drugs

Drug category (1st ATC level)	**Drug (5th ATC level)**
Category A: Alimentary tract and metabolism	*A01AA01: Sodium fluoride* *A05AX01: Piprozolin* ...
Category B: Blood and blood forming organs	*B01AC06: Acetylsalicylic Acid* *B03AA03: Ferrous gluconate* ...
Category C: Cardiovascular system	*C09AA05: Ramipril* *C10AA07: Rosuvastatin* ...
Category D: Dermatologicals	*D01AA02: Natamycin* *D01AA03: Hachimycin* ...
Category G: Genito-urinary system and sex hormones	*G04CB01: Finasteride* *G04CX03: Mepartricin* ...
Category H: Systemic hormonal preparations, excluding sex hormones and insulins	*H02AA02: Fludrocortisone* *H02AB07: Prednisone* ...
Category J: Antiinfectives for systemic use	*J01MA12: Levofloxacin* *J02AC04: Posaconazole* ...
Category L: Antineoplastic and immunomodulating agents	*L01AA07: Trofosfamide* *L01AB01: Busulfan* ...
Category M: Musculo-skeletal system	*M03AC10: Mivacurium chloride* *M03BA05: Febarbamate* ...
Category N: Nervous system	*N04AA02: Biperiden* *N04AB01 Etanautine* ...
Category P: Antiparasitic products, insecticides and repellents	*P01AA04: Chlorquinaldol* *P01AC01: Diloxanide* ...
Category R: Respiratory system	*R03AC02: Salbutamol* *R03BA02: Budesonide* ...
Category S: Sensory organs	*S02AA10: Acetic Acid* *S02BA03: Prednisolone* ...
Category V: Various	*V10XX01: Sodium phosphate* *V10XA01: Sodium iodide* ...

The achieved results, summarized in the following sections, show that the most appropriate minimum support threshold value to set depends on the analyzed data distribution. Furthermore, it appears that setting relatively high minimum confidence threshold values may result in pruning highly correlated and potentially actionable rules. Therefore, we identified the most appropriate support threshold value to set by comparing the results of multiple extractions on each dataset with different configuration settings and we deemed low-confidence rules to be worth considering during manual inspection as well as high-confidence ones.

3.2.1. Analysis of the number of prescriptions

We considered the frequency of drug/examination prescriptions because we deemed it as a significant indicator to support physicians in the following analyses: (i) select the most appropriate treatments, (ii) check the adherence of prescriptions to standard guidelines, and (iii) plan healthcare resource allocations.

To perform our analyses, we first differentiated between low-, cross-, and high-level W-GARs according to the definition reported in Section 2.4. This preliminary rule classification allows us to categorize the rule content based on its corresponding abstraction level in the input taxonomy.

The trend of variation of w-confidence values of low-, cross-, and high- W-GARs is pretty similar to those of traditional rule confidence values (i.e., the confidence of traditional generalized rules mined disregarding item weights). On average, W-GAR w-confidence appears to be lower than traditional confidence for cross- and high-level rules because generalized item occurrences are weighted by the actual descendant item weight within each transaction

and thus in w-support counting all descendant item occurrences are no longer weighted equally as in traditional support counting.

Notably, rule ranking in order of confidence often changes from traditional rules to W-GARs. More specifically, some of top-ranked traditional rules were downgraded because embedding item weight information the traditional confidence value decreases significantly. This result demonstrates that considering only examination/drug co-occurrences rather than their relevance weight is a suboptimal choice, which could yield rather different and potentially unreliable results. Consequently, W-GARs are more suitable than traditional generalized rules for effectively addressing analyses such as the ones mentioned above.

Weighted generalized rules related to examinations. Table 6 reports a representative subset of W-GARs mined from Diabetes-Freq and containing only examinations. The experiments were performed by setting *minsup* to 500[1] and *minconf* to 0. Table 7 reports the corresponding (unweighted) rules mined from the unweighted version of the diabetes dataset. Below we report a detailed comparison between a worthwhile subset of low-, cross-, and high-level W-GARs and the corresponding traditional rules. Although, for the sake of simplicity, in the following sections we mainly focus on 2-length rules (i.e., rules whose antecedent and consequent are singletons), our approach can extract rules of arbitrary length.

Analysis of cross-level rules. The cross-level W-GARs in Table 6 represent the association between the examinations in the Liver category and

[1]When not otherwise specified, we consider absolute minimum support thresholds throughout the paper.

Table 6: Examples of W-GARs related to examinations. Weighting measure: simple frequency of prescriptions

ID	Rule	W-sup	W-conf	AntInt	ConsInt
	Cross-level rules				
1	$\{(Exam{:}Liver)\} \rightarrow \{(Exam{:}Creatinine)\}$	8395	30%	0.83	0.66
2	$\{(Exam{:}Liver)\} \rightarrow \{(Exam{:}Uric\ Acid)\}$	7215	25%	0.73	0.66
3	$\{(Exam{:}Liver)\} \rightarrow \{(Exam{:}Microscopic\ urine\ analysis)\}$	7107	25%	0.84	0.68
4	$\{(Exam{:}Liver)\} \rightarrow \{(Exam{:}Culture\ urine)\}$	5426	19%	0.83	0.72
5	$\{(Exam{:}Liver)\} \rightarrow \{(Exam{:}Creatinine\ clearance)\}$	4078	14%	0.87	0.81
6	$\{(Exam{:}Liver)\} \rightarrow \{(Exam{:}Microalbuminuria)\}$	3633	13%	0.91	0.81
	High-level rules				
7	$\{(Exam{:}Liver)\} \rightarrow \{(Exam{:}Kidney)\}$	23953	84%	0.60	0.45
8	$\{(Exam{:}Kidney)\} \rightarrow \{(Exam{:}Cardiovascular)\}$	28777	75%	0.75	0.43
9	$\{(Exam{:}Eye)\} \rightarrow \{(Exam{:}Cardiovascular)\}$	6298	85%	0.96	0.43
10	$\{(Exam{:}Cardiovascular)\} \rightarrow \{(Exam{:}Eye)\}$	6298	17%	0.96	0.70
11	$\{(Exam{:}Liver), (Exam{:}Cardiovascular)\} \rightarrow \{(Exam{:}Kidney)\}$	20562	92%	0.67	0.45
	Low-level rules				
12	$\{(Exam{:}Creatinine\ clearance)\} \rightarrow \{(Exam{:}HDL\ Cholesterol)\}$	3760	89%	0.84	0.54
13	$\{(Exam{:}Microalbuminuria)\} \rightarrow \{(Exam{:}HDL\ Cholesterol)\}$	3425	88%	0.89	0.54
14	$\{(Exam{:}Urin\ acid)\} \rightarrow \{(Exam{:}HDL\ Cholesterol)\}$	6300	84%	0.79	0.54
15	$\{(Exam{:}Microscopic\ urine\ analysis)\} \rightarrow \{(Exam{:}HDL\ Cholesterol)\}$	5848	74%	0.85	0.54
16	$\{(Exam{:}Creatinine)\} \rightarrow \{(Exam{:}HDL\ Cholesterol)\}$	5937	65%	0.89	0.54

one specific examination belonging to the Kidney category. Among them, $R_1 : \{(Exam : Liver)\} \rightarrow \{(Exam : Creatinine)\}$ is the W-GAR with top confidence value (30%) and it is characterized by relatively high antecedent and consequent interest values (83% and 66%, respectively). Note that while entropy-based evaluators evaluate the correlation between data items regardless of their corresponding weights, the w-confidence value depends on the weights of the items in the rules. Rule R_1 indicates that the patients who performed 10 (100) examinations belonging to the liver category usually repeat the creatinine examination 3 (30) times. If we consider the corresponding traditional rule (see Table 7), R_1 appears to be the rule with lowest confidence value (39%). According to traditional rule definition, R_1 implies that most patients who performed at least one examination belonging

Table 7: Examples of traditional unweighted generalized rules related to examinations

ID	Rule	Sup	Conf	AntInt	ConsInt
Cross-level rules					
1	$\{(Exam{:}Liver)\} \rightarrow \{(Exam{:}Creatinine)\}$	1111	39%	0.83	0.66
2	$\{(Exam{:}Liver)\} \rightarrow \{(Exam{:}Uric\ Acid)\}$	2135	75%	0.73	0.66
3	$\{(Exam{:}Liver)\} \rightarrow \{(Exam{:}Microscopic\ urine\ analysis)\}$	1899	66%	0.84	0.68
4	$\{(Exam{:}Liver)\} \rightarrow \{(Exam{:}Culture\ urine)\}$	1890	66%	0.83	0.72
5	$\{(Exam{:}Liver)\} \rightarrow \{(Exam{:}Creatinine\ clearance)\}$	1444	51%	0.87	0.81
6	$\{(Exam{:}Liver)\} \rightarrow \{(Exam{:}Microalbuminuria)\}$	1146	40%	0.91	0.81
High-level rules					
7	$\{(Exam{:}Liver)\} \rightarrow \{(Exam{:}Kidney)\}$	2695	94%	0.60	0.45
8	$\{(Exam{:}Kidney)\} \rightarrow \{(Exam{:}Cardiovascular)\}$	2948	90%	0.75	0.43
9	$\{(Exam{:}Eye)\} \rightarrow \{(Exam{:}Cardiovascular)\}$	2275	72%	0.96	0.43
10	$\{(Exam{:}Cardiovascular)\} \rightarrow \{(Exam{:}Eye)\}$	2275	55%	0.96	0.70
11	$\{(Exam{:}Liver), (Exam{:}Cardiovascular)\} \rightarrow \{(Exam{:}Kidney)\}$	2563	95%	0.67	0.45
Low-level rules					
12	$\{(Exam{:}Creatinine\ clearance)\} \rightarrow \{(Exam{:}HDL\ Cholesterol)\}$	1400	93%	0.84	0.54
13	$\{(Exam{:}Microalbuminuria)\} \rightarrow \{(Exam{:}HDL\ Cholesterol)\}$	1137	90%	0.89	0.54
14	$\{(Exam{:}Microscopic\ urine\ analysis)\} \rightarrow \{(Exam{:}HDL\ Cholesterol)\}$	1925	87%	0.85	0.54
15	$\{(Exam{:}Creatinine)\} \rightarrow \{(Exam{:}HDL\ Cholesterol)\}$	1050	84%	0.89	0.54
16	$\{(Exam{:}Urin\ acid)\} \rightarrow \{(Exam{:}HDL\ Cholesterol)\}$	2100	93%	0.79	0.54

the Liver category have undergone the creatinine examination at least once. Such information is definitely less precise and thus potentially misleading for non-expert users. For example, physicians who have to allocate resources for medical examinations could make wrong decisions unless considering the actual number of examination prescriptions per patient. W-GAR w-confidence actually depends also on the expected frequency of exam repetitions. For example, according to the guidelines [25] the main Liver examinations are recommended to be prescribed three times a year to diabetics, whereas Creatinine level examination just once a year. Hence, W-GAR R_1 confirms the adherence of doctor's prescriptions to guidelines.

The first six W-GARs in Tables 6 summarize the most frequently prescribed kidney examinations prescribed to patients who have already under-

gone liver examinations. Such information is worth considering for verifying treatment adherence to standard guidelines as well as to support physicians in making future decisions. All these rules have relatively high antecedent and consequent interest values (AntInt $\geq$ 73%, ConsInt $\geq$ 66%). However, from the comparison between the W-GARs in Tables 6 and the rules in Table 7 a different confidence ranking appears. Therefore, weighting exam occurrences by the number of prescriptions really matters in deciding which guidelines are mostly disobeyed.

Analysis of high-level rules. Let us consider the following two traditional rules: $R_{10} : \{(Exam : Cardiovascular)\} \rightarrow \{(Exam : Eye)\}$ and $R_9 : \{(Exam : Eye)\} \rightarrow \{(Exam : Cardiovascular)\}$ (Table 7). When comparing them with the corresponding W-GARs (Table 6), it appears that the R_{10}'s confidence value significantly decreases (17% vs 55%) whereas the R_9's confidence value slightly increases (85% vs 72%). Both rules R_9 and R_{10} consist of strongly correlated items (AntInt=96%, ConsInt=96%). W-GAR R_{10} implies that patients who performed 10 (100) cardiovascular examinations on average perform 1.7 (17) examinations belonging the Eye category as well. Hence, the correlation between cardiovascular and eye examinations seems to be weak. On the other hand, W-GAR R_9 highlights a strong correlation between eye and cardiovascular examinations. Specifically, it indicates that, on average, patients who have undergone 10 (100) eye examinations have also performed 8.5 (85) cardiovascular examinations. Since guidelines for diabetes treatments [25] recommend to repeat cardiovascular and eye examinations at least once a year, it implies that a relatively large number of patients (15%) did not adhere to guidelines, i.e., they have repeated eye tests

but not cardiovascular tests.

Note that, in the analyzed scenario, if we enforce relatively high confidence thresholds (e.g., *minconf*=70%) many interesting low-confidence rules (e.g., R_{10} in Table 6) would be discarded thus potentially actionable information would be lost. Hence, to evaluate the adherence of examination prescriptions to guidelines low-confidence rules are deemed to be as interesting as high-confidence ones. Therefore experts are recommended to set averagely low confidence thresholds. In case the number of mined rules becomes too large for manual inspection, rules may be ranked by decreasing AntInt/ConsInt values and only the top ranked rules can be manually explored.

Analysis of low-level rules. Let us consider the high-level rule $R_8 : \{(Exam : Kidney)\} \rightarrow \{(Exam : Cardiovascular)\}$ first (see Table 6). It indicates that patients who have undergone any examination belonging the Kidney category are very likely to undergo cardiovascular examinations as well (75% of likelihood). Low-level rules allow us to deepen into the analysis of such a pattern. For example, we can consider the rules that contain the HDL cholesterol examination in the rule consequent to figure out the Kidney examinations that are likely to be prescribed in conjunction with the HDL cholesterol test. Traditional rule $R_{16} : \{(Exam : Urin\ acid)\} \rightarrow \{(Exam : HDL\ cholesterol)\}$ has the top confidence value (93%), but it appears to be misleading, because the corresponding W-GAR has a relatively low confidence compared to other similar W-GARs (63%). Conversely, rule R_{12}: $\{(Exam : Creatinine\ clearance)\} \rightarrow \{(Exam : HDL\ colesterol)\}$ and its corresponding W-GAR have both high (w-)confidence (respectively 93% and 89%) and thus they can be used to analyze the correlation between

Table 8: Examples of W-GARs related to drugs. Weighting measure: simple frequency of prescriptions.

ID	Rule	W-sup	W-conf	AntInt	ConsInt
Cross-level rules					
1	$\{(Drug{:}Category\ B)\} \rightarrow \{(Drug{:}B01AC06)\}$	8373	59%	0.51	0.82
2	$\{(Drug{:}Category\ B)\} \rightarrow \{(Drug{:}B01AC05)\}$	2058	15%	0.92	0.97
3	$\{(Drug{:}Category\ B)\} \rightarrow \{(Drug{:}B01AB06)\}$	1384	10%	0.88	0.95
4	$\{(Drug{:}Category\ B)\} \rightarrow \{(Drug{:}B01AA03)\}$	741	5%	0.96	0.98
High-level rules					
5	$\{(Drug{:}Category\ B)\} \rightarrow \{(Drug{:}Category\ C)\}$ B = Blood and blood forming organs C = Cardiovascular system	12922	91%	0.96	0.42
6	$\{(Drug{:}Category\ C)\} \rightarrow \{(Drug{:}Category\ B)\}$	12922	17%	0.96	0.72

Table 9: Examples of traditional unweighted generalized rules related to drugs

ID	Rule	Sup	Conf	AntInt	ConsInt
Cross-level rules					
1	$\{(Drug{:}Category\ B)\} \rightarrow \{(Drug{:}B01AC06)\}$	1670	71%	0.51	0.82
2	$\{(Drug{:}Category\ B)\} \rightarrow \{(Drug{:}B01AC05)\}$	323	14%	0.92	0.97
3	$\{(Drug{:}Category\ B)\} \rightarrow \{(Drug{:}B01AB06)\}$	485	21%	0.88	0.95
4	$\{(Drug{:}Category\ B)\} \rightarrow \{(Drug{:}B01AA03)\}$	54	3%	0.96	0.98
High-level rules					
5	$\{(Drug{:}Category\ B)\} \rightarrow \{(Drug{:}Category\ C)\}$ B = Blood and blood forming organs C = Cardiovascular system	2101	90%	0.96	0.42
6	$\{(Drug{:}Category\ C)\} \rightarrow \{(Drug{:}Category\ B)\}$	2101	50%	0.96	0.72

the HDL cholesterol test and other examinations. More in detail, based on the W-GAR rule, patients who repeated 10 (100) Creatinine Clearance examination are likely to repeat HDL cholesterol examination approximately 9 (89) times.

Summarizing, we observed that, in many cases, estimates made using W-GARs appear to be more consistent and realistic than those made using traditional generalized rules.

Weighted generalized rules related to drugs. We performed a second round of experiments by considering drugs instead of examinations. For this experimental session, we used again the simple frequency of prescription as weighting measure and we set *minsup* to 500 and *minconf* to 0. Even in this case we selected a subset of potentially interesting patterns, which are reported in Table 8. These patterns can be exploited by healthcare system managers to profile drug prescriptions and thus to verify prescription adherence to guidelines or to plan drug provision. We compared again traditional generalized rules (see Table 9) with W-GARs (see Table 8).

The obtained results are similar to those achieved with examinations. Specifically, W-GAR and traditional rule ranking in order of confidence relevantly change. For example, let us consider the four cross-level W-GARs in the form $\{(\textit{Drug:Category B})\} \rightarrow \{(\textit{Drug:Specific drug of category B})\}$ reported in Table 8. They can be used to plan Category B drug provision. For example, drug B01AC06 (acetylsalicylic Acid) appears to be the most prescribed drug among those belonging to Category B. More specifically, 59% of the prescriptions of Category B drugs are B01AC06 prescriptions, whereas only 15%, 10%, and 5% of the Category B's prescriptions are related to drugs B01AC05, B01AB06, and B01AA03, respectively. These results can be deemed worthy by the National Health Center to plan drug provision. Let us now consider the four corresponding traditional rules in Table 9. The confidence value of a rule in the form $\{(\textit{Drug:Category B})\} \rightarrow \{(\textit{Drug:Specific drug of category B})\}$ represents the percentage of patients to whom a specific Category B drug has been prescribed at least once with respect to the total number of patients who have taken a Category B drug at least once. Based

on traditional rule confidence, we could make only rough estimates and thus a wrong drug provision could be planned. For example, the ranking of Category B drugs is misleading, because B01AB06 appears to be more frequently prescribed than B01AC05, even if this is not case. Note that most of the rules in Table 8 related to Category B drugs are characterized by relatively high interest values, independently of their confidence values.

Let us now consider the high-level W-GARs, which represents the association between Category B and Category C drugs (rules R_5 and R_6 in Table 8). The W-GARs reported in Table 8 indicate that the strength of the "implication" $\{(\textit{Drug:Category B})\} \rightarrow \{(\textit{Drug:Category C})\}$ is significantly higher than those of the opposite rule. In other words, patients with blood complications are frequently treated with cardiovascular drugs as well, whereas the opposite implication is unlikely. A contrasting result is achieved if we consider traditional rules rather than W-GARs (see in Table 9). For example, the confidence of traditional rule $\{(\textit{Drug:Category C})\} \rightarrow \{(\textit{Drug:Category B})\}$ is 50% where the w-confidence of the corresponding W-GAR is only 17%. The high-confidence W-GAR R_5 :$\{(\textit{Drug:Category B})\} \rightarrow \{(\textit{Drug:Category C})\}$ represents a strong recurrence among data items (i.e., it holds in 91% of the cases). Patients who do not adhere to this pattern should be analyzed separately, because they could represent either anomalous behaviors.

Finally, rules showing the association between a therapeutic group and the corresponding chemical subgroup are also extracted. An example follows: $\{(\textit{Drug:ATC2_A11})\} \rightarrow \{(\textit{Drug:ATC4_A11CC})\}$, w-conf=100%. This rule indicates that 100% of the prescribed vitamins (ATC2_A11 = Vitamins) are Vitamin D or analogues (ATC4_A11CC = Vitamin D and analogues).

Vitamin D is particularly recommended to diabetics. Unlike other vitamins, it is found in very few foods and thus physicians commonly prescribe it as a supplement.

3.2.2. Analysis of the tf-idf of the prescriptions

Since we are analyzing patients with the same illness (i.e., diabetes), we are very interested in identifying segments of patients treated with common treatments because they suffer from similar disease variations or complications. With this goal in mind, we deemed the simple frequency of drug/examination prescriptions as not appropriate for this new analysis, because it targets examinations/drugs prescribed to all patients indifferently. In contrast, we aim at identifying peculiar treatment features. Hence, we used the tf-idf statistics. Strong W-GARs mined from patient data enriched with tf-idf weights allow us to identify the examinations/drugs that are frequently prescribed only to a small subset of patients (see Section 2.3).

From the analysis of the extracted high-level rules, it appears that, as expected, routine examinations and drugs are characterized by relatively low support values. Hence, very common examinations/drugs are early pruned or ranked in last place. On the other hand, some low-level rules highlight specific examinations/drugs that have been prescribed to a small subset of patients, as thoroughly discussed below.

Weighted generalized rules related to examinations. Table 10 reports a subset of weighted generalized rules representing worthwhile correlations between examinations. The patterns were extracted by setting *minsup* to 50 and *minconf* to 0.

Table 10: Examples of weighted generalized rules related to examinations. Weighting measure: tf-idf

ID	Rule	W-Sup	W-conf	AntInt	ConsInt
High-level rules					
1	$\{(Exam{:}Liver)\} \rightarrow \{(Exam{:}Kidney)\}$	130.3	73%	0.60	0.45
2	$\{(Exam{:}Kidney)\} \rightarrow \{(Exam{:}Cardiovascular)\}$	125.3	69%	0.60	0.52
Low-level rules					
3	$\{(Exam{:}ALT)\} \rightarrow \{(Exam{:}AST)\}$	85.8	98%	0.21	0.58
4	$\{(Exam{:}AST)\} \rightarrow \{(Exam{:}ALT)\}$	85.8	96%	0.21	0.57
5	$\{(Exam{:}Triglycerides)\} \rightarrow \{(Exam{:}HDL\ cholesterol)\}$	68.4	95%	0.31	0.54
6	$\{(Exam{:}HDL\ cholesterol)\} \rightarrow \{(Exam{:}Triglycerides)\}$	68.4	94%	0.31	0.53

Rules R_3-R_6 in Table 10 represent segments of patients with specific treatments. More specifically, rules R_3 and R_4 cover a segment of patients for which ALT and AST examinations are frequently prescribed. Both examinations belong to the Liver category. Hence, these W-GARs are likely to represent diabetics with liver complications. On the other hand, rules R_5 and R_6 are related to cardiovascular examinations. Since the support of the aforementioned rules is relatively high (compared to those of the other mined patterns), we can conclude that both examinations have a high tf-idf and they are both frequently prescribed to a specific subset of patients. More specifically, the selected rules highlight patients with diabetes and cardiovascular complications. Note also that the confidence of rules R_3-R_6 is always higher than 94% and the same property holds for the opposite implication. Hence, we can state that, for instance, the ALT and AST prescriptions are related to a specific subset of patients and both drugs are prescribed, approximately, in the same quantity.

As expected, rule ranking in order of w-support using tf-idf weight produces a different ranking with respect to the simple frequency of prescrip-

tions. For example, let us consider the high-level rules reported in Table 10 (i.e., rules R_1 and R_2) and the corresponding rules reported in Table 6 (i.e., rules R_7 and R_8). Note that rule $\{(Exam{:}Liver)\} \rightarrow \{(Exam{:}Kidney)\}$ ranked higher than $\{(Exam{:}Kidney)\} \rightarrow \{(Exam{:}Cardiovascular)\}$ with tf-idf weights, whereas the opposite ranking is achieved using the number of prescriptions as weighting measure (see Table 6). Therefore, the two measures produce rather different results and they can be used by domain experts to address complementary issues.

Weighted generalized rules related to drugs. We also analyzed the associations among drugs using the tf-idf weights and by setting *minsup* to 500 and *minconf* to 0. In Table 11 a worthwhile subset of selected rules is reported.

Table 11: Examples of weighted generalized rules related to drugs. Weighting measure: tf-idf

ID	Rule	W-sup	W-conf	AntInt	ConsInt
Low-level rules					
1	$\{(Drug{:}ATC4_L04AD)\} \rightarrow \{(Drug{:}ATC4_L04AA)\}$ ATC4_L04AD = Calcineurin inhibitors ATC4_L04AA = Selective immunosuppressants	798.6	100%	1.00	1.00
High-level rules					
2	$\{(Drug{:}Category\ A)\} \rightarrow \{(Drug{:}Category\ C)\}$ A = Alimentary tract and metabolism C = Cardiovascular system	105262.4	87%	0.99	0.42

The mined rules are representative of diabetics with specific complications. For example, the first rule reported in Table 11 identifies a subset of patients who are treated with drugs related to transplants, e.g., patients who have undergone pancreas transplants. Similarly, high-level rule R_2 :$\{(Drug{:}Category\ A)\} \rightarrow \{(Drug{:}Category\ C)\}$ highlights a segment of

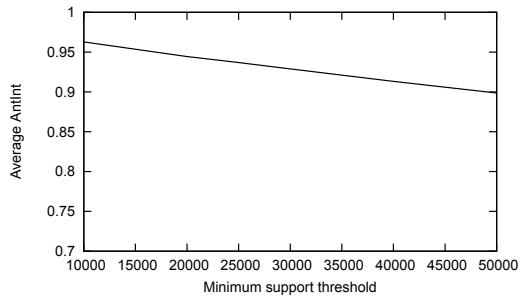

Figure 2: Impact of the *minsup* threshold on the average AntInt of the mined rules related to drugs. Weighting measure: tf-idf. $minconf$=0

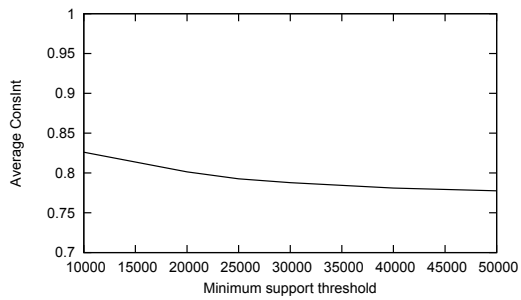

Figure 3: Impact of the *minsup* threshold on the average ConsInt of the mined rules related to drugs. Weighting measure: tf-idf. $minconf$=0

patients with alimentary and metabolism diseases associated with cardiovascular complications.

3.3. Analysis of the algorithm parameters

The W-GAR mining algorithm is driven by two (user-defined) parameters: the minimum w-support threshold $minwsup$ and the minimum w-confidence threshold $minwconf$. However, setting these parameters properly could be a challenging task. We performed a set of experiments to analyze the impact of $minwsup$ and $minwconf$ on the quality of the mined rules in terms of average AntInt and ConsInt [35]. Due to the lack of space, we reported only the results achieved on drugs by using tf-idf weights. Similar

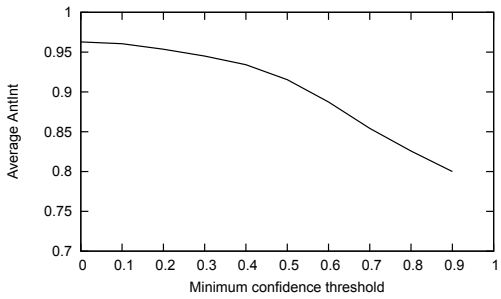

Figure 4: Impact of the *minwconf* threshold on the average AntInt of the mined rules related to drugs. Weighting measure: tf-idf. *minwsup*=1000

results were obtained on the other datasets. Figures 2 and 3 respectively show the average AntInt and ConsInt values achieved by varying the value of the *minwsup* threshold and by setting *minwconf* to 0 (i.e., no confidence threshold).

Low-support rules are, on average, characterized by higher AntInt and ConsInt values with respect to medium- and high-support ones. However, as shown in Figures 2 and 3, the rule quality decrease achieved by enforcing medium *minwsup* values appears to be limited. For example, by enforcing a support threshold equal to 35000 the average AntInt value is 0.93 and the average ConsInt value is 0.8, whereas by enforcing a minimum support threshold equal to 10000 the average AntInt is 0.96 and the average ConsInt is 0.83. Since the algorithm execution time scales more than linearly with the number of the enforced support threshold, to limit the computational complexity of the rule mining process we recommend experts to set medium *minwsup* values to perform their analyses. A more detailed analysis of the execution time taken by the proposed algorithm is reported in Section 3.5.

Figures 4 and 5 respectively show the average AntInt and ConsInt values

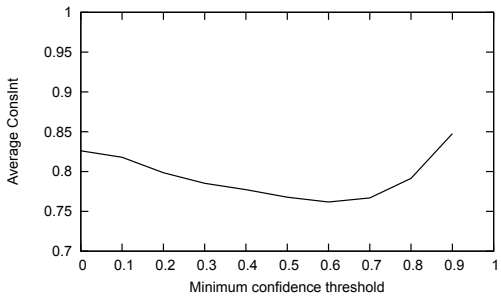

Figure 5: Impact of the *minwconf* threshold on the average ConsInt of the mined rules related to drugs. Weighting measure: tf-idf. *minwsup*=1000

achieved by varying the value of the *minwconf* threshold and by setting a fixed *minwsup* value (1000). The AntInt measure appears to be inversely correlated with the *minwconf* value, while the correlation between ConsInt and *minwconf* appears to be weaker. However, for both measures the rule quality is maximal when low *minwconf* values are enforced. Hence, to avoid discarding low-confidence yet interesting rules we recommend experts to set no *minconf* threshold (i.e., *minwconf*=0). To ease the manual exploration of top interesting rules the mined rules can be ranked by decreasing AntInt and ConsInt values. Examples of low-confidence rules that are particularly interesting in the analyzed context are given in Section 3.2.1.

3.4. Comparison with different weighted association rule mining approaches

We compared the number and quality of the rules mined by our approach with those of the rules generated by a previous approach, namely WARM [47]. Note that in [47] each transaction is first weighted by the average of all its item weights. Then, the rule support is counted by averaging the weights of all the covered transactions. Hence, the set of rules mined by WARM is potentially different from those extracted by our approach.

Figures 6-9 plot the number of rules mined by the W-GAR and WARM algorithms on the examination and drug datasets with different weighting measures (i.e., number of prescriptions and tf-idf). Since in WARM the weight of a transaction depends on the weight of all of its items, all the rules covering the same transactions have the same support value, even if they cover a different subset of items. Hence, the support of the rules generated by WARM is on average higher than those extracted by W-GAR. Indeed, the number of frequent rules mined by WARM is higher on all datasets and for all configuration settings. The support count made by WARM appears to be particularly unreliable when coping with datasets consisting of relatively large transactions. As an example, in Tables 10 and 12 we reported a subset of representative rules in common between W-GAR and WARM. They report the w-support and w-confidence values counted by W-GAR and WARM, respectively. As expected, the support and confidence values counted by WARM are higher than those achieved by W-GAR, because all the items in the transactions are considered. Furthermore, since low- and high-level items are not differentiated the high-level item weights could bias the support count of the itemsets consisting of lower-level items.

We also compared the rules mined by the two approaches according to two established rule quality measures, i.e., AntInt and ConsInt [35]. To perform a fair comparison, first we set the minimum support threshold values so that the two algorithms extracted approximately the same number of rules. Then in Figures 10 and 11 we plotted the AntInt values of the the top 10000 rules in order of decreasing AntInt and ConsInt values, respectively. The results show that top ranked W-GARs are on average more interesting than those mined

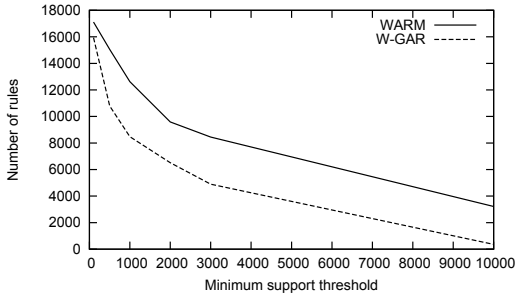

Figure 6: Number of mined rules related to examinations. Weighting measure: frequency of prescriptions

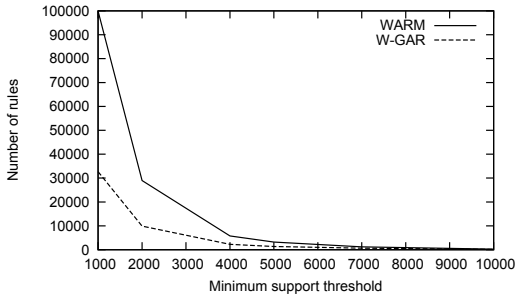

Figure 7: Number of mined rules related to drugs. Weighting measure: frequency of prescriptions

by WARMs according to the considered quality measures. Similar results, omitted due to the lack of space, were achieved on examinations by using tf-idf weights. Conversely, slightly different results were achieved using the simple frequency weights. More specifically, with this configuration setting the quality measures of the top ranked rules extracted by the two approaches appear to be pretty similar with each other, because since the distribution of the frequency count values is denser than those of tf-idf values, the differences in terms of rule quality indices become negligible.

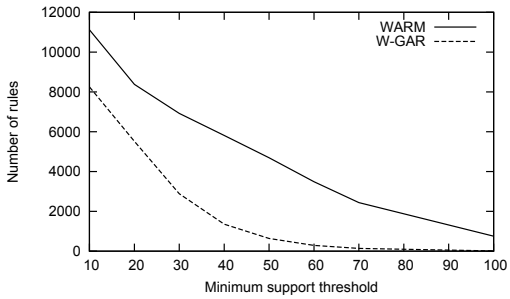

Figure 8: Number of mined rules related to examinations. Weighting measure: tf-idf

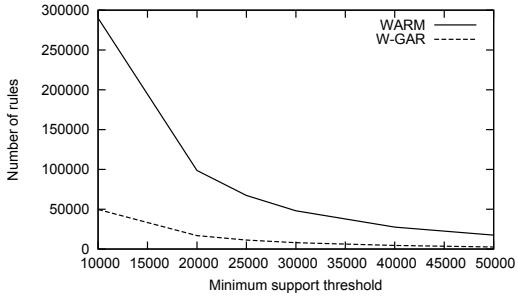

Figure 9: Number of mined rules related to drugs. Weighting measure: tf-idf

3.5. Execution time

We also analyzed the performance of the proposed approach in terms of execution time by varying the minimum support threshold. Figure 12 compares the execution times taken by the newly proposed W-GAR algorithm and the WARM algorithm [47] by varying the $minwsup$ value and by setting no $minwconf$ threshold (i.e., $minwconf$=0).

As representative example, we considered the results achieved on the drugs dataset with the tf-idf weighting measure, because for both algorithms rule extraction on this dataset takes maximal time compared to the other datasets and configurations.

The execution times of the two algorithms are roughly comparable when

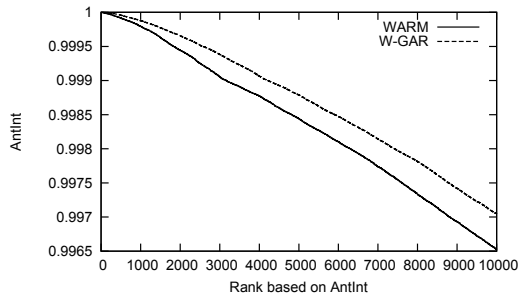

Figure 10: Rule ranking. Rules order by decreasing AntInt value. Dataset related to drugs. Weighting measure: tf-idf

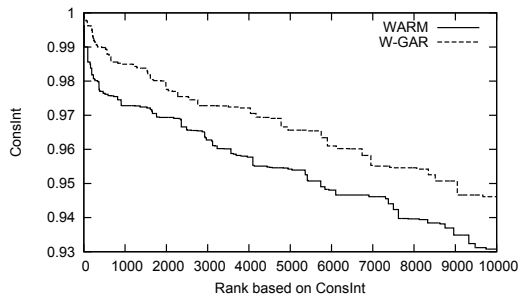

Figure 11: Rule ranking. Rules order by decreasing ConsInt value. Dataset related to drugs. Weighting measure: tf-idf

high support threshold values are enforced (higher than 40000), while W-GAR is about 5 times faster than WARM when low support thresholds are considered (e.g., 10000) . The execution time of W-GAR ranges from few seconds by enforcing relatively high *minwsup* values (e.g., 40000) to approximately 15 seconds by enforcing very low support thresholds (e.g., 10000). The non-linear increase in the execution time is due to the combinatorial increase of the number of generated item combinations. The execution times of the W-GAR and WARM algorithms are strongly related with the number of mined rules (see Figure 9).

As already discussed in Section 3.3, to achieve the best trade-off be-

Table 12: WARM. Examples of weighted generalized rules related to examinations mined by using WARM. Weighting measure: tf-idf

ID	Rule	W-sup WARM	W-conf WARM	AntInt	ConsInt
High-level rules					
1	$\{(Exam{:}Liver)\} \rightarrow \{(Exam{:}Kidney)\}$	122.1	98%	0.60	0.45
2	$\{(Exam{:}Kidney)\} \rightarrow \{(Exam{:}Cardiovascular)\}$	124.6	89%	0.75	0.43
Low-level rules					
3	$\{(Exam{:}ALT)\} \rightarrow \{(Exam{:}AST)\}$	113.5	98%	0.21	0.58
4	$\{(Exam{:}AST)\} \rightarrow \{(Exam{:}ALT)\}$	113.5	99%	0.21	0.57
5	$\{(Exam{:}Triglycerides)\} \rightarrow \{(Exam{:}HDL\ cholesterol)\}$	114.7	97%	0.31	0.54
6	$\{(Exam{:}HDL\ cholesterol)\} \rightarrow \{(Exam{:}Triglycerides)\}$	114.7	99%	0.31	0.53

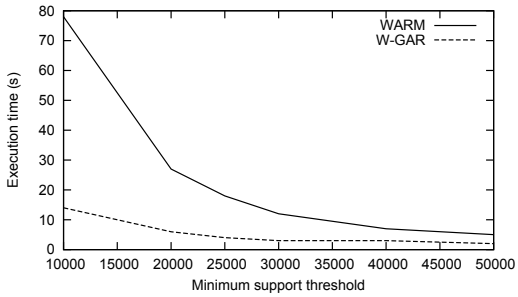

Figure 12: Algorithm execution time related to drugs. Weighting measure: tf-idf. $minwconf$=0

tween rule quality and computational complexity experts are recommended to set medium-value support thresholds (e.g., $minwsup$=35000). Nevertheless, based on the achieved results, all the tested configuration settings seem to be suitable for performing offline data analyses.

Due to the lack of space we do not report the results of scalability tests with respect to the number of records and features. However, similar to traditional generalized rule mining algorithms (e.g., [16, 9, 21, 41]), our algorithm scales linearly with the dataset cardinality and more than linearly with the average transaction length, because varying the number of transactions

the data item distribution remains approximately the same, whereas increasing the number of data items the number of generated item combinations combinatorially increases.

4. Related works

This section compares the proposed approach with the state-of-the-art related works. Based on the covered topic, the related works are discussed in separate subsections: (i) pattern mining from medical data, (ii) generalized (unweighted) association rule mining, and (iii) weighted (non-generalized) association rule mining.

Pattern mining from medical data. Data mining algorithms have largely been exploited to discover interesting patterns among medical data, such as frequent and interesting patterns among patient treatments (e.g., [20, 40]), temporal relationships in temporal clinical data [11, 18, 50], groups of correlated patients [4, 5, 48], patterns relevant for patient classification (e.g., [26, 33, 38]). Among the aforementioned approaches, association rules are worth considering in the analysis of healthcare data to transform huge amounts of raw data into actionable knowledge. Various kinds of patterns at different abstraction levels have been considered for analyzing medical data. Traditional association rules have been exploited in the heart disease scenario [34] to study sick and healthy factors. Three association rule extraction algorithms (i.e., Apriori [3], Predictive Apriori [39], and Tertius [20]) have been investigated. In [40], instead, association rules have been exploited to determine two important diseases in patients diagnosed with essential hypertension, i.e., non-insulin dependent diabetes mellitus and cerebral infarction.

Unlike [20, 34, 39] in this study we proposed a new kind of patterns, named weighted generalized association rules (W-GARs), to represent interesting multiple-level associations among patient treatments.

Preliminary attempts to discover generalized patterns from medical data have been made in [6, 12, 29]. Specifically, in [29] the authors analyzed multiple-level co-occurrences among diseases in a public health dataset, while in [12] generalized rules are used to represent biomedical relationships between concepts occurring in Medline. The authors in [6] proposed to analyze multiple-level associations among medical treatments and patient profiles. A level-wise rule categorization has also been proposed to ease manual result exploration. Similarly, to ease the exploration of the mined patterns different approaches, graph-based strategies (e.g., [36]) have also been proposed. However, none of the previous approaches is able to cope with data equipped with item weights. Hence, to the best of our knowledge, this paper is the first attempt to tackle generalized association rule mining from weighted data. To address this issue, it proposes a novel type of generalized association rule, namely the W-GAR. Furthermore, two different weighting measures have been exploited for weighting data items. Each weighting measure is targeted to a different use case (e.g., to identify peculiar treatment features, to select the most appropriate treatments, to check the adherence of prescriptions to standard guidelines, or to plan healthcare resource allocations).

A parallel research activity has been devoted to taking temporal information into account during pattern mining from healthcare data. For example, the authors in [18] proposed a rule-based approach to discovering complex temporal relationships in interval-based temporal clinical data, while in [19] a

sequential pattern mining algorithm has been customized to manage multi-dimensional healthcare data. Unlike [19, 18] this work does not consider sequential or temporal patterns.

Generalized association rule mining. A notable research effort has been devoted to efficiently extracting generalized frequent itemsets and association rules from (unweighted) transactional datasets. The first generalized frequent itemset mining algorithm has been proposed in [41] in the context of market basket analysis. It generates itemsets by considering for each item all its parents in the hierarchy. To avoid generating all the possible candidates in the taxonomy, the authors in [16, 42, 43] proposed to push (analyst-provided) constraints into the mining process. Many algorithm optimizations have also been proposed based on: a top-down hierarchy traversal [21], closed and maximal generalized itemsets [37] support-driven approaches [9, 13, 14]. Unlike all the aforementioned approaches, this paper addresses the problem of mining generalized rules from weighted datasets. Specifically, the goal is to differentiate between relevant items and not within each transaction by taking item weights into account during generalized rule extraction.

Weighted association rule mining. To consider the relative importance of items during the mining process, some attempts to mine association rules from weighted data have already been made [17, 44, 47, 49]. The approach presented in [49] is, to the best of our knowledge, the first attempt to consider item weights during association rule mining. Data items are enriched with weights denoting item relevance/intensity within each transaction. The goal of [49] is to segment of the domain of the item weights in the dataset and to generate reliable rules containing both items and weight

intervals. The mined rules have a different expression and semantics with respect to those mined by our approach. For example, rule $drugA[4,6] \rightarrow drugB[3,5]$ means that if drug A is prescribed in the quantity between 4 and 6 pills, then drug B is likely to be prescribed in the quantity between 3 to 5 pills. Conversely, our approach does not embeds item weights into the rule expression, but rather it considers them to compute the main rule quality indices, e.g., rule $drugA \rightarrow drugB$ is extracted if drugs A and B are characterized by high ratings (e.g., high number of prescriptions, high price, high customer satisfaction).

In [47] the authors defined the weighted support of a rule $A \rightarrow B$ as the fraction of the weight of the transactions that contains both A and B relative to the weight of all transactions, where the transaction weight is computed as the average of all its item weights. However, since rules are unlikely to contain all the transaction items the presence of a highly relevant item in a transaction could bias the support value of the rules that cover that transaction but do not contain any highly relevant item. Therefore, in [47] the analysis of the traditional rule quality measures (e.g., support, confidence [46]) is potentially misleading. Conversely, the approach adopted in this paper considers only the weights of the items in the rule because they really matter in rule support counting. Therefore, the extracted rules and their corresponding quality indices are deemed as more reliable for advanced analyses.

A parallel effort has been to mining weighted association rules without preassigned weights. To address this issue, in [44] the analyzed transactional dataset is represented as a bipartite hub-authority graph and evaluated by

means of a well-known indexing strategy [28] in order to automate item weight assignment. The proposed approach is significantly different from those presented in [28], because in our context of analysis (i.e., patient data analysis) item weights are given and they represent the number of drug/exam prescriptions. Therefore, there is no need for inferring item weights with indexing algorithms. More recently, in [17] the problem of mining infrequent itemsets from weighted data has also been addressed. Unlike [17, 44, 47, 49] the paper addresses the problem of mining weighted rules from datasets equipped with taxonomies. To this aim, it proposes a new type of rules (i.e., W-GARs), which represent associations among weighted data items at different abstraction levels.

5. Conclusions and future work

This paper presents a new type of generalized association rule, which considers item weights during the rule evaluation process. Item weights measure the relative item importance within each transaction.

The experiments performed on a real diabetic patient dataset highlight interesting and actionable correlations among patient treatments. The extracted knowledge is consistent with the guidelines for diabetes disease [1, 23] and it is particularly useful for performing advanced data analyses.

As future work, we aim at investigating the applicability of weighted generalized association rules in other application contexts, including financial data analysis [30], sensor data analysis [31], genetic data analysis [8], and social network data analysis [15]. For example, financial data can be straightforwardly modeled as weighted data, because for each company/financial in-

strument several key performance indicators are given and their values vary over time. For instance, stock prices continuously vary when stock markets are open. To gain insights into the analyzed data stocks can be generalized as the corresponding financial sectors or as the corresponding stock markets. Hence, discovering significant correlations between financial data items at different abstraction levels can be an appealing research issue. Similarly, sensor readings can be easily integrated into centralized data repositories and modeled as weighted data, where a reading collects the measurements acquired by all the sensors in a network at a given timestamp. Sensor data can be analyzed at different temporal and spatial granularities. On the one hand, experts may would like to analyze the underlying correlations among sensor data acquired at different time frequencies (e.g., one reading per second, one reading per day). On the other hand, sensors can be clustered according to their spatial position in the network topology. To reduce the maintenance cost of sensor networks the correlations between nearby sensor data can be analyzed and one representative sensor per group can be maintained. Therefore, in this context of analysis weighted generalized association rules can be exploited to study the spatial and temporal correlation between sensor measurements.

6. Acknowledgments

This work was partially supported by the GenData2020 project grant, which is funded by the Italian Ministry of Research (MIUR).

The authors wish to thank Prof. Dario Antonelli and Dr. Giulia Bruno for their advices and fruitful discussions.

References

[1] ADA, American Diabetes Association Standards of Medical Care in Diabetes 2013, Diabetes Care 36 (2013) S11–S66.

[2] R. Agrawal, T. Imieliński, A. Swami, Mining association rules between sets of items in large databases, in: ACM SIGMOD Record, ACM, New York, 1993, pp. 207–216.

[3] R. Agrawal, R. Srikant, Fast algorithms for mining association rules in large databases, in: Proceedings of the 20th International Conference on Very Large Data Bases, volume 1215, Morgan Kaufmann, San Fransisco, 1994, pp. 487–499.

[4] M.U. Ahmed, P. Funk, Mining rare cases in post-operative pain by means of outlier detection, in: IEEE Symposium on Signal Processing and Information Technology (ISSPIT), pp. 35–41.

[5] D. Antonelli, E. Baralis, G. Bruno, T. Cerquitelli, S. Chiusano, N. Mahoto, Analysis of diabetic patients through their examination history, Expert Systems with Applications 40 (2013) 4672 – 4678.

[6] D. Antonelli, E.M. Baralis, G. Bruno, L. Cagliero, T. Cerquitelli, S.A. Chiusano, P. Garza, N.A. Mahoto, Meta: Characterization of medical treatments at different abstraction levels, ACM Transactions on Intelligent Systems and Technology (2015).

[7] ATC, Norwegian-Institute-of-Public-Health: ATC/DDD Index 2013. Available: http://www.whocc.no/atc_ddd_index/ . Last access on November 2013, 2013.

[8] E. Baralis, L. Cagliero, T. Cerquitelli, S. Chiusano, P. Garza, Frequent weighted itemset mining from gene expression data, in: 13th IEEE International Conference on BioInformatics and BioEngineering, IEEE, 2013, pp. 1–4.

[9] E. Baralis, L. Cagliero, T. Cerquitelli, V. D'Elia, P. Garza, Support driven opportunistic aggregation for generalized itemset extraction, in: 5th IEEE International Conference on Intelligent Systems, IEEE, 2010, pp. 102–107.

[10] E. Baralis, L. Cagliero, T. Cerquitelli, P. Garza, M. Marchetti, Casmine: providing personalized services in context-aware applications by means of generalized rules, Knowl. Inf. Syst. 28 (2011) 283–310.

[11] I. Batal, H. Valizadegan, G.F. Cooper, M. Hauskrecht, A temporal pattern mining approach for classifying electronic health record data, ACM Trans. Intell. Syst. Technol. 4 (2013) 63:1–63:22.

[12] M. Berardi, M. Lapi, P. Leo, C. Loglisci, Mining generalized association rules on biomedical literature, in: Innovations in Applied Artificial Intelligence, volume 3533 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2005, pp. 500–509.

[13] L. Cagliero, Discovering temporal change patterns in the presence of taxonomies, IEEE Trans. Knowl. Data Eng. 25 (2013) 541–555.

[14] L. Cagliero, T. Cerquitelli, P. Garza, L. Grimaudo, Misleading generalized itemset discovery, Expert Syst. Appl. 41 (2014) 1400–1410.

[15] L. Cagliero, T. Cerquitelli, P. Garza, L. Grimaudo, Twitter data analysis by means of strong flipping generalized itemsets, Journal of Systems and Software 94 (2014) 16–29.

[16] L. Cagliero, P. Garza, Itemset generalization with cardinality-based constraints, Inf. Sci. 244 (2013) 161–174.

[17] L. Cagliero, P. Garza, Infrequent weighted itemset mining using frequent pattern growth, Knowledge and Data Engineering, IEEE Transactions on 26 (2014) 903–915.

[18] C. Combi, A. Sabaini, Extraction, analysis, and visualization of temporal association rules from interval-based clinical data, in: Artificial Intelligence in Medicine, volume 7885 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2013, pp. 238–247.

[19] E. Egho, C. Rassi, D. Ienco, N. Jay, A. Napoli, P. Poncelet, C. Quantin, M. Teisseire, Healthcare trajectory mining by combining multidimensional component and itemsets, in: New Frontiers in Mining Complex Patterns, volume 7765 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2013, pp. 109–123.

[20] P. Flach, V. Maraldi, F. Riguzzi, Algorithms for efficiently and effectively using background knowledge in tertius, 2006. Department of computer science, University of Bristol, Udine, Italy.

[21] J. Han, Y. Fu, Mining multiple-level association rules in large databases, IEEE Transactions on Knowledge and Data Engineering 11 (2002) 798–805.

[22] J. Han, J. Pei, Y. Yin, R. Mao, Mining frequent patterns without candidate generation: A frequent-pattern tree approach, Data Min. Knowl. Discov. 8 (2004) 53–87.

[23] I. ICD-9-CM, International Classification of Diseases, 9th revision, Clinical Modification. Available: http://icd9cm.chrisendres.com. Last access on March 2011, 2011.

[24] IDF, International Diabetes Federation. Available: http://www.idf.org/. Last access on November 2013, 2013.

[25] IDF, Italian Society of Diabetology. Available:http://www.siditalia.it/pubblicazioni.html. Last access on March 2014, 2014.

[26] A. Karegowda, M. Jayaram, A. Manjunath, Cascading k-means clustering and k-nearest neighbor classifier for categorization of diabetic patients, International Journal of Engineering and Advanced Technology (IJEAT) (2012) 147 – 151.

[27] R. Kimball, M. Ross, The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling, John Wiley & Sons, Inc., New York, NY, USA, 2nd edition, 2002.

[28] J.M. Kleinberg, Authoritative sources in a hyperlinked environment, J. ACM 46 (1999) 604–632.

[29] R. Kost, B. Littenberg, E. Chen, Exploring generalized association rule mining for disease co-occurrences, in: Proceedings of the AMIA Symposium, pp. 1284–1293.

[30] G. Kou, Y. Peng, G. Wang, Evaluation of clustering algorithms for financial risk analysis using mcdm methods, Information Sciences 275 (2014) 1–12.

[31] W.K. Lai, C.S. Fan, L.Y. Lin, Arranging cluster sizes and transmission ranges for wireless sensor networks, Information Sciences 183 (2012) 117–131.

[32] C.Y. Lin, E. Hovy, Automatic evaluation of summaries using n-gram co-occurrence statistics, in: Proceedings of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1, pp. 71–78.

[33] X.H. Meng, Y.X. Huang, D.P. Rao, Q. Zhang, Q. Liu, Comparison of three data mining models for predicting diabetes or prediabetes by risk factors, The Kaohsiung Journal of Medical Sciences 29 (2013) 93 – 99.

[34] J. Nahar, T. Imam, K. Tickle, Y. Chen, Association rule mining to detect factors which contribute to heart disease in males and females, Expert Systems with Applications (2012).

[35] E. Noda, A. Freitas, H. Lopes, Discovering interesting prediction rules with a genetic algorithm, in: Evolutionary Computation, 1999. CEC 99. Proceedings of the 1999 Congress on, volume 2, pp. –1329 Vol. 2.

[36] H. Pan, X. Tan, Q. Han, X. Feng, G. Yin, Gma: an approach for association rules mining on medical images, in: Proceedings of the 8th international conference on Intelligent Computing Theories and Applications, ICIC'12, pp. 425–432.

[37] N. Pasquier, Y. Bastide, R. Taouil, L. Lakhal, Discovering frequent closed itemsets for association rules, in: Proceedings of the 7th International Conference on Database Theory, ICDT '99, Springer-Verlag, London, UK, UK, 1999, pp. 398–416.

[38] B.M. Patil, R.C. Joshi, D. Toshniwal, Classification of type-2 diabetic patients by using apriori and predictive apriori, Int. J. Comput. Vision Robot. 2 (2011) 254–265.

[39] T. Scheffer, Finding association rules that trade support optimally against confidence, Principles of Data Mining and Knowledge Discovery (2001) 424–435.

[40] A. Shin, I. Lee, G. Lee, H. Park, H. Park, K. Yoon, J. Lee, Y. Kim, Diagnostic analysis of patients with essential hypertension using association rule mining, Healthcare informatics research 16 (2010) 77–81.

[41] R. Srikant, R. Agrawal, Mining generalized association rules, in: International Conference on Very Large Data Bases, Morgan Kaufmann, San Fransisco, 1995, pp. 407–419.

[42] R. Srikant, Q. Vu, R. Agrawal, Mining association rules with item constraints, in: Conference on Knowledge Discovery and Data Mining, volume 97, AAAI Press, USA, 1997, pp. 67–73.

[43] K. Sriphaew, T. Theeramunkong, A new method for finding generalized frequent itemsets in generalized association rule mining, in: Seventh International Symposium on Computers and Communications, IEEE, Washington, D.C., 2002, pp. 1040–1045.

[44] K. Sun, F. Bai, Mining weighted association rules without preassigned weights, IEEE Transactions on Knowledge and Data Engineering 20 (2008) 489 –495.

[45] P. Tan, V. Kumar, J. Srivastava, Selecting the right interestingness measure for association patterns, in: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, New York, 2002, p. 41.

[46] P. Tan, M. Steinbach, V. Kumar, et al., Introduction to data mining, Pearson Addison Wesley, Boston, 2006.

[47] F. Tao, F. Murtagh, M. Farid, Weighted association rule mining using weighted support and significance framework, in: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, KDD'03, pp. 661–666.

[48] S.M. Van Rooden, W.J. Heiser, J.N. Kok, D. Verbaan, J.J. van Hilten, J. Marinus, The identification of parkinson's disease subtypes using cluster analysis: A systematic review, Movement Disorders 25 (2010) 969–978.

[49] W. Wang, J. Yang, P.S. Yu, Efficient mining of weighted association rules (WAR), in: ACM SIGKDD international conference on Knowledge discovery and data mining, KDD'00, pp. 270–274.

[50] J.Y. Yeh, T.H. Wu, C.W. Tsao, Using data mining techniques to predict hospitalization of hemodialysis patients, Decis. Support Syst. 50 (2011) 439–448.