MARAS: Signaling Multi-Drug Adverse Reactions *

Xiao Qin
xiao.qin@hhs.fda.gov
Worcester Polytechnic Institute

Tabassum Kakar
tabassum.kakar@hhs.fda.gov
Worcester Polytechnic Institute

Susmitha Wunnava
swunnava@wpi.edu
Worcester Polytechnic Institute

Elke A. Rundensteiner
rundenst@cs.wpi.edu
Worcester Polytechnic Institute

Lei Cao
lcao@csail.mit.edu
Massachusetts Institute of Technology

ABSTRACT

There is a growing need for computing-supported methods that facilitate the automated signaling of Adverse Drug Reactions (ADRs) otherwise left undiscovered from the exploding amount of ADR reports filed by patients, medical professionals and drug manufacturers. In this research, we design a Multi-Drug Adverse Reaction Analytics Strategy, called **MARAS**, to signal severe unknown ADRs triggered by the usage of a combination of drugs, also known as Multi-Drug Adverse Reactions (MDAR). First, **MARAS** features an efficient signal generation algorithm based on association rule learning that extracts *non-spurious* MDAR associations. Second, **MARAS** incorporates contextual information to detect drug combinations that are strongly associated with a set of ADRs. It groups related associations into Contextual Association Clusters (CACs) that then avail contextual information to evaluate the significance of the discovered MDAR Associations. Lastly, we use this contextual significance to rank discoveries by their notion of interestingness to signal the most compelling MDARs. To demonstrate the utility of **MARAS**, it is compared with state-of-the-art techniques and evaluated via case studies on datasets collected by U.S. Food and Drug Administration Adverse Event Reporting System (FAERS).

KEYWORDS

Public Health Surveillance; Adverse Drug Reaction; Association Rule Learning; Interestingness of Association

ACM Reference format:
Xiao Qin, Tabassum Kakar, Susmitha Wunnava, Elke A. Rundensteiner, and Lei Cao. 2017. MARAS: Signaling Multi-Drug Adverse Reactions [1]. In *Proceedings of KDD'17, August 13–17, 2017, Halifax, NS, Canada.,* , 9 pages.
DOI: http://dx.doi.org/10.1145/3097983.3097986

[1]Xiao Qin and Tabassum Kakar are ORISE Research Participants at U.S. Food and Drug Administration, Center for Drug Evaluation and Research, Office of Surveillance and Epidemiology. This project was supported in part by the National Science Foundation IIS-1560229 and CRI-1305258.

KDD'17, August 13–17, 2017, Halifax, NS, Canada.
 ISBN 978-1-4503-4887-4/17/08…$15.00
DOI: http://dx.doi.org/10.1145/3097983.3097986

1 INTRODUCTION

1.1 Background

An Adverse Drug Reaction (ADR) corresponds to an unwanted and often dangerous effect caused by the administration of a drug. ADRs are a major cause of morbidity and mortality worldwide. According to the U.S. Food and Drug Administration (FDA), every year hundreds of thousands of people die because of these ADRs while over two million serious ADRs are reported annually. ADRs can be caused by the administration of a single or multiple drugs either upon immediate or prolonged use or even overdose. ADRs caused by multiple drugs are known as Multi-Drug Adverse Reactions (MDAR). For example, *Aspirin* taken together with *Warfarin*, a blood-thinning drug, may lead to excessive bleeding [9]. It is critical that MDARs are detected early with minimum patient exposure to avoid further harmful incidents. Unlike the ADRs caused by one single drug, identifying MDARs in clinical trails is difficult, since it is impossible to try every possible drug combination.

For early detection of novel ADRs which are not captured during the clinical trials, Spontaneous Reporting Systems (SRS) are designed to collect information on adverse events related to drugs reported by patients, health care professionals and drug manufacturers filed via mail, telephone and Internet. FDA Adverse Event Reporting System (FAERS) is one such system [4].

Data collected from the surveillance programs is a useful resource to tap into MDARs. As thousands of new reports are added on a daily basis, discovering MDARs by aimlessly screening and analyzing all these reports is extremely difficult if not impossible. Therefore, computational methods, especially data mining techniques promise to be critical for identifying the most emerging MDAR signals from massive reports. These signals which can be seen as MDAR hypothesis along with the reports that derive these signals are then recommended to the drug safety evaluator for further investigation and validation.

1.2 Limitations of State-of-the-Art

As a well established data mining method for discovering interesting relationships among variables in the data, Association Rule Learning (ARL) is considered as a natural fit for MDAR signaling. For example Wei *et al.* [22] used ARL to signal vaccine MDARs in the US Vaccine Adverse Event Reporting System. Harpaz *et al.* [14] applied ARL to signal MDARs in FAERS. However it is shown that directly applying traditional rule learning algorithms to detect MDARs tend to produce a large amount of irrelevant, redundant, and even misleading rules. Worst yet, to date no mechanism is proposed that is able to effectively rank the rules based on their

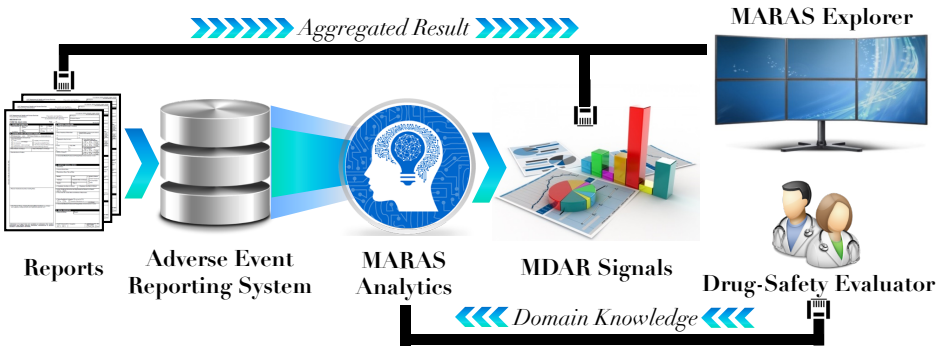

Figure 1: The MARAS approach.

possibilities of being real MDARs. Therefore finding real MDARs still requires manual validation of each and every one of the large number of rules produced by ARL. This is still a painful and difficult task for the drug-safety evaluator.

1.3 Research Challenges

To develop a Multi-Drug Adverse Reaction Analytics Strategy (MARAS) using ARL, the following research challenges must be addressed:

Amount of generated associations. Association learning applied to a set of thousands of drugs and ADRs generates an extremely huge number of associations that are impossible for a drug-safety evaluator to sift through. The volume of these association rules may at times be higher than the number of original reports they are derived from. Although the number of rules can be reduced by setting a high *support* parameter, this would risk missing damaging MDARs. This is so because low support does not necessarily indicate small possibility of being a dangerous MDAR for a given rule. Therefore, we encounter the dilemma of having to reduce the number of rules, while still guaranteeing no useful rule is overlooked.

Measure w.r.t MDAR. To effectively rank the produced rules and therefore help the drug-safety evaluator concentrate on the rules most likely to be real MDARs, measures that effectively reflect the significance of the association between a set of drugs and a set of ADRs have to be provided. However, the off-the-shelf common used association measures such as *support*, *confidence* and *lift* (RR) [7] focus only on a single association rule based on the number of its occurrences, while the correlations among different rules have to be considered when measuring the significance of a rule to be a MDAR. For examples, if two rules contain the same ADRs and overlaps on the medicines, their significance might be influenced by each other. Therefore, we are in need of a customized measure to quantify the significance of an association in terms of its signaled MDARs.

1.4 The MARAS Methodology

To tackle the above challenges, we design the Multi-Drug Adverse Reaction Analytics Strategy (MARAS) depicted in Figure 1 that successfully discovers all potential MDARs and recommends them to the analyst with an explicit significance score.

First, MARAS is based on our critical observation that a small set of associations – namely the closed Drug-ADR associations out of the huge amount of associations generated by directly applying the association learning techniques is sufficient to assist the analyst to discover all potential MDARs. Leveraging this observation, MARAS is able to accurately identify this small set of closed Drug-ADR associations without having to generate and preserve any irrelevant, redundant and misleading associations.

Second, by carefully analyzing the factors that influence the interestingness of the discovered associations, we observe that the common understanding in association learning literature, namely a long rule is more preferred than its shorter sub-rules, does not hold in our MDAR scenario. More specially, for example, if the chance of *Lorazepam* and *Abilify* taken together triggering a set of ADRs such as dizziness, drowsiness, confusion, and difficulty concentrating is high, while the possibility of *Lorazepam* taken individually triggering these ADRs is also high, then the combination of *Lorazepam* and *Abilify* together as a potential cause for these adverse effects is in fact not significant [1]. This is so because there is a high possibility the ADRs are triggered by Lorazepam alone. Based on this observation, we propose a *Contextual Association Cluster* (CAC) model to evaluate the interestingness of the discovered MDARs. By taking into account their contextual associations, our CAC model is shown to be able to better signal the MDARs as compared to the traditional rule interestingness measurements such as *support*, *confidence*, and *lift* (RR) [7].

1.5 Contributions

Our contributions include:

- We design and implement an end-to-end MDAR signaling solution that facilitates the drug safety evaluator to identify emerging severe unknown MDARs.
- We adopt and adapt association rule learning for MDAR signaling by proposing a pruning strategy to remove *spurious* associations while keeping the most relevant ones.
- We propose the *Contextual Association Cluster* model and the *contrast* measure to evaluate the interestingness of the associations in terms of the degree of being a true MDAR.
- We evaluate the utility of MARAS using adverse event reports extracted from FAERS both by conducting a case study as well as a comparative analysis with existing methods.

2 PRELIMINARIES

Association rule learning [5] is a popular technique used to detect relationships among the items in large databases. Let $\mathcal{I} = \{i_1, i_2, ..., i_n\}$ represent a set of **items**. $\mathcal{T} = \{t_1, t_2, ..., t_m\}$ is a collection of subsets of I called the **transaction database**. Each **transaction** t_i in $\mathcal{T}$ corresponds to a set of items such that $t_i \subseteq \mathcal{I}$. Let $\mathcal{Z} \subseteq \mathcal{I}$ be a nonempty set of items, called **itemset**. If $\mathcal{Z} \subseteq t_i$, transaction t_i *contains* $\mathcal{Z}$. $|\mathcal{Z}|$ denotes the number of transactions in $\mathcal{T}$ that contain $\mathcal{Z}$. If the cardinality of $\mathcal{Z}$ is k, $\mathcal{Z}$ is called a **k-itemset**.

Definition 2.1. An **association rule** is an expression of the form $\mathcal{R} \equiv \mathcal{X} \Rightarrow \mathcal{Y}$, where $\mathcal{X}$ and $\mathcal{Y}$ are itemsets and $\mathcal{X} \subseteq \mathcal{I}, \mathcal{Y} \subseteq \mathcal{I} \setminus \mathcal{X}$.

Number of Associations. ARL is a two-step process. First, the frequent itemsets are generated, then associations are induced from them in the second step. The total number of possible itemsets that can be generated based on $\mathcal{I}$ is:

$$\sum_{k=1}^{n} \binom{n}{k} = \binom{n}{1} + \binom{n}{2} + ... + \binom{n}{n} = 2^n - 1, \quad (1)$$

where k is the cardinality of the itemsets and n the number of unique items. The total number of possible associations derived from these itemsets is:

$$\sum_{k=1}^{n} \binom{n}{k} (2^{n-k} - 1) = \sum_{k=1}^{n} \binom{n}{k} 2^{n-k} - \sum_{k=1}^{n} \binom{n}{k} = 3^n - 2^{n+1} + 1, \quad (2)$$

where k is the cardinality of the antecedent and n the number of unique items.

Measures. Many measures [19] have been proposed to evaluate the interestingness of associations. The most commonly used ones *support*, *confidence* and *lift* (RR) for an association $\mathcal{R}$ are defined as follows:

$$support(\mathcal{R}) = P(\mathcal{X} \cup \mathcal{Y}) = |\mathcal{X} \cup \mathcal{Y}|, \quad (3)$$

$$confidence(\mathcal{R}) = P(\mathcal{Y}|\mathcal{X}) = \frac{|\mathcal{X} \cup \mathcal{Y}|}{|\mathcal{X}|}, \quad (4)$$

$$lift(\mathcal{R}) = \frac{P(\mathcal{Y}|\mathcal{X})}{P(\mathcal{Y})} = \frac{P(\mathcal{X}|\mathcal{Y})}{P(\mathcal{X})} = \frac{|\mathcal{X} \cup \mathcal{Y}| \times |\mathcal{T}|}{|\mathcal{X}| \times |\mathcal{Y}|}. \quad (5)$$

The *support* (Formula 3) describes the proportion of the transactions that *contain* all items in the association. *confidence* (Formula 4) describes the probability of finding the *consequent* $\mathcal{Y}$ of the association under the condition that these transactions also *contain* the *antecedent* $\mathcal{X}$. It is a maximum likelihood estimate of the conditional probability $P(\mathcal{Y}|\mathcal{X})$. *Lift* (Formula 5) measures how many times more often $\mathcal{X}$ and $\mathcal{Y}$ occur together than expected if they are statistically independent.

3 ASSOCIATION RULE MODEL FOR MDAR SIGNALING

3.1 Drug-ADR Association

Let $\mathcal{I}_{Drug} = \{d_1, d_2, ..., d_o\}$ and $\mathcal{I}_{ADR} = \{a_1, a_2, ..., a_u\}$ represent a set of drugs and a set of ADRs where $\mathcal{I}_{Drug} \cap \mathcal{I}_{ADR} \equiv \emptyset$. $\mathcal{T} = \{t_1, t_2, ..., t_m\}$ is a collection of ADR reports. Each report $t_i \equiv \mathcal{D}_i \cup \mathcal{A}_i$ contains a drug set $\mathcal{D}_i$ where $\mathcal{D}_i \subseteq \mathcal{I}_{Drug}$ and an ADR set $\mathcal{A}_i$ where $\mathcal{A}_i \subseteq \mathcal{I}_{ADR}$. Since we are only interested in modeling the associations from a set of drugs to a set of ADRs in a collection of ADR reports, we define the Drug-ADR association as below.

Definition 3.1. A **Drug-ADR association** is an expression of the form $\mathcal{R} \equiv \mathcal{D} \Rightarrow \mathcal{A}$ where $\mathcal{D} \subseteq \mathcal{I}_{Drug}$, $\mathcal{A} \subseteq \mathcal{I}_{ADR}$ and $\mathcal{I}_{Drug} \cap \mathcal{I}_{ADR} \equiv \emptyset$.

Irrelevant Association. If the traditional association rule model were to be directly applied on the ADR reports $\mathcal{T}$, the ARL algorithm can possibly generate $3^{o+u} - 2^{o+u} + 1$ ($O(3^n)$ where $n = o + u$) associations according to Formula 2 where o and u denote the total number of unique drugs and ADRs respectively. However, based on Definition 3.1, the number of possible Drug-ADR associations instead corresponds to:

$$|2^{\mathcal{I}_{Drug}} \times 2^{\mathcal{I}_{ADR}}| = \sum_{k=1}^{o} \binom{o}{k} \times \sum_{k=1}^{u} \binom{u}{k} = (2^o - 1) \times (2^u - 1). \quad (6)$$

According to Formula 6, the number of possible Drug-ADR associations ($O(2^n)$ where $n = o + u$) is much smaller than $O(3^n)$. The associations that do not confirm the defined Drug-ADR expression are *irrelevant*, therefore need to be pruned in the learning process. Also, since we study MDARs in this work, we focus on the Drug-ADR associations which contain at least two drugs in the antecedent.

3.2 Non-spurious Drug-ADR Association

Without pre-established dependency constraints among items, existing ARL algorithms [23] consider every possible combination of items that appears in a transaction as an *itemset* (Formula 1). This results in a huge amount of *redundant* [6, 18, 24] even *misleading* associations in the context of signaling ADRs from ADR reports as we show below.

3.2.1 Types of Drug-ADR Associations.

Explicitly Supported Drug-ADR Association. Let us consider an ADR report $t_i \equiv \mathcal{D}_i \cup \mathcal{A}_i$ with a set of drugs $\mathcal{D}_i \equiv \{d_1, d_2, d_3\}$ and a set of ADRs $\mathcal{A}_i \equiv \{a_1, a_2\}$. This particular ADR report explicitly establishes the association between $\mathcal{D}_i$ and $\mathcal{A}_i$, expressed by the association $\mathcal{R}_1 \equiv (d_1 \wedge d_2 \wedge d_3) \Rightarrow (a_1 \wedge a_2)$. However, based upon this single report, traditional ARL would generate 24 variants of Drug-ADR associations ($(3^2 - 1) \times (2^2 - 1)$), such as $(d_1 \wedge d_2) \Rightarrow (a_1)$, $(d_1 \wedge d_3) \Rightarrow (a_2)$ etc. including $\mathcal{R}_1$. All of these associations, except $\mathcal{R}_1$, are **partial interpretations** of the report, randomly leaving out certain item(s), e.g., some drugs or some ADRs mentioned in the report. In many scenarios, these associations could be misleading unless there is additional evidence to support them. For example, $\mathcal{R}_2 \equiv d_1 \Rightarrow a_2$ tells us that taking d_1 might lead to a_2. This may however not be true in our context since this report does not *explicitly indicate* that drug d_1 by itself will lead to ADR a_2 therefore cannot be confirmed by this ADR report.

Definition 3.2. A Drug-ADR association $\mathcal{R} \equiv \mathcal{D} \Rightarrow \mathcal{A}$ is **explicitly supported** by a collection of ADR reports $\mathcal{T}$ if there exists at least one report $t_i \in \mathcal{T}$ where $t_i \equiv \mathcal{D}_i \cup \mathcal{A}_i$ such that $t_i \equiv \mathcal{D} \cup \mathcal{A}$.

If a Drug-ADR association is *explicitly* supported, according to definition 3.2, at least one report must exist that refers exactly to drugs and ADRs in the association and no additional ones. Other reports that contain these drugs and ADRs can be used as evidence to measure the significance of this association.

Implicitly Supported Drug-ADR Association. In addition to t_i in the last example, let us consider adding another ADR report

$t_j \equiv \mathcal{D}_j \cup \mathcal{A}_j$ with a set of drugs $\mathcal{D}_j \equiv \{d_1,d_2,d_4\}$ and a set of ADRs $\mathcal{A}_i \equiv \{a_1,a_2\}$. According to Definition 3.2, $\mathcal{R}_3 \equiv (d_1 \wedge d_2 \wedge d_4) \Rightarrow (a_1 \wedge a_2)$ is *explicitly* supported by $\mathcal{T}$. Although the Drug-ADR association $\mathcal{R}_4 \equiv (d_1 \wedge d_2) \Rightarrow (a_1 \wedge a_2)$ is a **partial interpretation** of t_i or t_j, it may be of interest to the drug safety evaluator since it involves the **intersection** of two reports which can be interpreted as a commonly prescribed drug combination or a commonly caused ADRs. The Drug-ADR associations formed by the intersection of multiple reports such as $\mathcal{R}_4$ are defined as *implicitly supported* Drug-ADR associations:

Definition 3.3. A Drug-ADR association $\mathcal{R} \equiv \mathcal{D} \Rightarrow \mathcal{A}$ is **implicitly supported** by a collection of ADR reports $\mathcal{T}$ if there exist at least two ADR reports $t_i, t_j \in \mathcal{T}$ where $i \neq j, t_i \not\equiv t_j, t_i \equiv \mathcal{D}_i \cup \mathcal{A}_i$ and $t_j \equiv \mathcal{D}_j \cup \mathcal{A}_j$ such that $t_i, t_j \not\equiv \mathcal{D} \cup \mathcal{A}$, $\mathcal{D} \equiv \mathcal{D}_i \cap \mathcal{D}_j$ and $\mathcal{A} \equiv \mathcal{A}_i \cap \mathcal{A}_j$.

According to Definition 3.3, if a Drug-ADR association is *implicitly* supported, it models an association between a commonly prescribed drug combination and commonly caused ADRs suggested by at least two reports and it is not *explicitly* supported. If a Drug-ADR association is neither *explicitly* nor *implicitly* supported, it is a **spurious association** which must be treated with caution as it may convey misleading information. Next, we will discuss how our system identifies these associations.

3.2.2 Learning Non-spurious Drug-ADR Association.

$\mathcal{S}_{exp}$ and $\mathcal{S}_{imp}$ denote complete sets of *explicitly* and *implicitly* supported Drug-ADR associations learned from a collection of ADR reports $\mathcal{T}$. Below we show that identifying $\mathcal{S}_{exp} \cup \mathcal{S}_{imp}$ is equivalent to identifying *closed* associations [18] from all possible Drug-ADR associations in $\mathcal{T}$. *Closed* associations [6] compactly represent the same information as the full set of all possible associations and can be used to recover the full set. The notion of a *closed* association is defined as below:

Definition 3.4. An association $\mathcal{R}_i \equiv \mathcal{X}_i \Rightarrow \mathcal{Y}_i$ is called **closed** in a set of transactions $\mathcal{T}$ if there does not exist an association $\mathcal{R}_j \equiv \mathcal{X}_j \Rightarrow \mathcal{Y}_j$ where $i \neq j$ such that $\mathcal{X}_i \cup \mathcal{Y}_i \subset \mathcal{X}_j \cup \mathcal{Y}_j$ and $|\mathcal{X}_i \cup \mathcal{Y}_i| = |\mathcal{X}_j \cup \mathcal{Y}_j|$.

According to Definition 3.4, if an association $\mathcal{R}_i$ is not *closed* in a dataset, there exists another association $\mathcal{R}_j$ with additional items (richer information) which is also contained by the same set of transactions. For example, for associations $\mathcal{R}_1 \equiv (i_1 \wedge i_2) \Rightarrow (i_3 \wedge i_4)$ and $\mathcal{R}_2 \equiv (i_1) \Rightarrow (i_3 \wedge i_4)$ where i represents an item, if $|\{i_1, i_2, i_3, i_4\}| = |\{i_1, i_3, i_4\}|$, this means that $\mathcal{R}_1$ and $\mathcal{R}_2$ are contained by the same set of transactions. Regardless whether or not $\mathcal{R}_1$ is *closed*, $\mathcal{R}_2$ is not closed since it only presents partial information of $\mathcal{R}_1$.

Let $\mathcal{S}_{Drug-ADR}$ denote a complete set of Drug-ADR associations learned from a collection of ADR reports $\mathcal{T}$ and $\mathcal{S}^*_{Drug-ADR}$ be the complete set of *closed* Drug-ADR associations in $\mathcal{S}_{Drug-ADR}$. We have the following claim.

Lemma 3.5. *The closed Drug-ADR association set $\mathcal{S}^*_{Drug-ADR} \equiv \mathcal{S}_{exp} \cup \mathcal{S}_{imp}$ where $\mathcal{S}^*_{Drug-ADR}, \mathcal{S}_{exp}$ and $\mathcal{S}_{imp}$ are learned from the same collection of ADR reports $\mathcal{T}$.*

Proof. The proof is bi-directional. First, if a Drug-ADR association is *closed*, it is either *explicitly* or *implicitly* supported. Second, if a Drug-ADR association is either *explicitly* or *implicitly* supported, it must be *closed*.

First, consider a Drug-ADR association $\mathcal{R} \equiv \mathcal{D} \Rightarrow \mathcal{A}$, if $\mathcal{R}$ is *closed* then there does not exist an $\mathcal{R}_i$ such that $\mathcal{R}_i$ has additional items beyond $\mathcal{R}$ and is contained by the same set of ADR reports as $\mathcal{R}$. There are two possibilities causing such non-existence: (1) no report exists that contains more items than $\mathcal{D} \cup \mathcal{A}$ which makes $\mathcal{R}$ *explicitly* supported; (2) $\mathcal{D} \cup \mathcal{A}$ is an intersection of multiple reports and all $\mathcal{R}_i$ with additional items are of course also contained in less reports; If there is a report among them that contains the exact same items in $\mathcal{R}$ then $\mathcal{R}$ is *explicitly* supported, otherwise it is *implicitly* supported.

Second, if $\mathcal{R}$ is *explicitly* supported, either (1) there exists no report with additional items in $\mathcal{R}$ which makes $\mathcal{R}$ *closed* because there is no $\mathcal{R}_i$ with additional items that can be learned from the reports; or (2) in addition to the report(s) that contain the exact items in $\mathcal{R}$, there are reports with more items; But this will make the R_i with additional items be contained by less amount of reports than $\mathcal{R}$; Therefore, $\mathcal{R}$ is *closed*. If $\mathcal{R}$ is *implicitly* supported, it contains the interaction of multiple reports, then all the R_i with additional items are contained by less reports; Therefore $\mathcal{R}$ is *closed*. □

We use Lemma 3.5 as theoretical foundation to efficiently identify non-spurious Drug-ADR associations.

3.3 Contextual Association Cluster

Table 1: Example of a Contextual Association Cluster of $\mathcal{R}$

$\mathcal{R}$	[Furosemide] [Isosorbide] [Aspirin] ⇒ [Myocardial Infarction]
$\tilde{\mathcal{R}}^2$	$\tilde{\mathcal{R}}^2_1 \equiv$ [Furosemide] [Isosorbide] ⇒ [Myocardial Infarction]
	$\tilde{\mathcal{R}}^2_2 \equiv$ [Furosemide] [Aspirin] ⇒ [Myocardial Infarction]
	$\tilde{\mathcal{R}}^2_3 \equiv$ [Isosorbide] [Aspirin] ⇒ [Myocardial Infarction]
$\tilde{\mathcal{R}}^1$	$\tilde{\mathcal{R}}^1_1 \equiv$ [Furosemide] ⇒ [Myocardial Infarction]
	$\tilde{\mathcal{R}}^1_2 \equiv$ [Isosorbide] ⇒ [Myocardial Infarction]
	$\tilde{\mathcal{R}}^1_3 \equiv$ [Aspirin] ⇒ [Myocardial Infarction]

Next, we introduce how MARAS measures non-spurious Drug-ADR associations that contain multiple drugs to signal MDARs. Existing measures [7] including *support*, *confidence* and *lift* (RR) evaluate the strength of the association between two set of items. However, they lack the ability to verify whether this strong association is already implied by a subset of the antecedent. Such a domination from a subset of the drug antecedents would weaken the MDAR signal. For example, if the ADRs are already highly associated with an individual drug in the given combination of drugs of the association, it means that the ADRs are likely caused by this particular drug or subset of drugs instead of the larger MDAR.

To measure this notion of *exclusiveness* of the association between drugs and ADRs, any association between a subset of drugs and the ADRs needs to be considered. These related associations are henceforth referred to as the **contextual** associations of the target association.

Definition 3.6. A Drug-ADR association $\mathcal{R}_i \equiv \mathcal{D}_i \Rightarrow \mathcal{A}_i$ is a **contextual association** of a Drug-ADR Association $\mathcal{R}_j \equiv \mathcal{D}_j \Rightarrow \mathcal{A}_j$ if and only if $\mathcal{D}_j \subset \mathcal{D}_i$ and $\mathcal{A}_i \equiv \mathcal{A}_i$.

Based on Definition 3.6, we define the **Contextual Association Cluster** (CAC) of a target Drug-ADR association.

Definition 3.7. A **Contextual Association Cluster** $C \equiv \{\mathcal{R}, \tilde{\mathcal{R}}_1,...,\tilde{\mathcal{R}}_n\}$ includes an explicitly or implicitly supported Drug-ADR association $\mathcal{R} \equiv \mathcal{D} \Rightarrow \mathcal{A}$ and its contextual associations such that $\bigcup_{i=1}^{n} \tilde{\mathcal{D}}_i \equiv \mathbb{P}(\mathcal{D}) - \{\emptyset, \mathcal{D}\}$ where $\tilde{\mathcal{D}}_i$ is antecedent of the contextual association $\tilde{\mathcal{R}}_i$ and $\mathbb{P}(\mathcal{D})$ is the power set of $\mathcal{D}$. $\mathcal{R}$ is called **target** association.

Table 1 shows an example of the CAC of a target Drug-ADR association $\mathcal{R}$ which represents the MDAR signal. The CAC is organized based on the cardinality of the antecedent. The number n in $\tilde{\mathcal{R}}^n$ refers to the number of drugs in the association. In this example, $\mathcal{R}$ has 3 drugs. Hence, there are 6 contextual associations in CAC. MARAS uses CAC to evaluate the interestingness of the target Drug-ADR association that contains multiple drugs in terms of signaling the most severe MDARs.

3.4 Contrast Score for MDAR Signal

To measure if a Drug-ADR association encodes a strong signal that indicates a severe MDAR, two factors need be taken into consideration. First, how strong the association of ADRs is with the drug combination and second, how strong the association of ADRs is with the individual or subset of drugs. As explained in Section 3.3, if ADRs are caused by the interaction of a drug combination then not only the ADRs must be strongly associated with the drug combination but also any subset of these drugs should only be weakly associated with the particular ADRs.

For the first factor, MARAS adopts the *confidence* model (Formula 4) that represents a maximum likelihood estimate of the conditional probability $P(\mathcal{A}|\mathcal{D})$ for a Drug-ADR association $\mathcal{R}$. It models the strength of the association between the antecedent and consequent. High *confidence* indicates strong association while low *confidence* indicates weak association. For the second factor, we first defined the CAC introduced in Section 3.3. A CAC includes a target association that represents the MDAR signal along with all its contextual associations that represent the associations between the target ADRs and the subsets of the target drugs. The MDAR signal is strongest if the target association has high *confidence* and all of its contextual associations in the cluster have low *confidence*. To quantify such a contrast that captures the intuition of the MDAR phenomenon, as discussed in Section 3.3, we propose the *contrast* measure.

Let $C \equiv \{\mathcal{R}, ..., \tilde{\mathcal{R}}_j^i, ...\}$ represent a CAC, with $\mathcal{R}$ the target association and $\tilde{\mathcal{R}}_j^i$ its contextual associations where i denotes the number of drugs in the association and j is used to distinguish between different contextual associations with the same amount of drugs i. $\mathcal{P}_c(\mathcal{R})$ denotes the *confidence* of an association $\mathcal{R}$. The MDAR signal is strong if the *confidence* of $\mathcal{R}$ is significantly higher than any *confidence* of its contextual associations.

$$contrast_{max}(C) = \mathcal{P}_c(\mathcal{R}) - max(\mathcal{P}_c(\tilde{\mathcal{R}}_j^i)). \quad (7)$$

A negative $contrast_{max}$ value means that a subset of drugs is more likely to cause the ADRs then the actual target set. This idea is similar to the *improvement* measure proposed by Bayardo *et al.* [17]. However, only considering the contextual association with the highest *confidence* deprives us of the opportunity to differentiate more complex cases. For example, even if two MDAR signals share the same *contrast* value, the one with more higher *confidence* contextual associations may be less interesting than the other one because more drugs may cause the same ADRs showing a weaker sign of the MDAR. To utilize the full context in the evaluation of the MDAR signal, an alternative solution would be to measure the difference between the *confidence* of the target association and the average *confidence* of its contextual associations:

$$contrast_{avg}(C) = \mathcal{P}_c(\mathcal{R}) - \frac{1}{|C|-1} \sum_{i=1}^{n} \sum_{j=1}^{m} \mathcal{P}_c(\tilde{\mathcal{R}}_j^i). \quad (8)$$

The shortcoming of this solution is that it falsely weakens the negative effect of any contextual association with a high *confidence*. For example, let us consider two CAC cases $C_1 \equiv \{\mathcal{R}, \tilde{\mathcal{R}}_1^1, \tilde{\mathcal{R}}_2^1\}$ and $C_2 \equiv \{\mathcal{R}_2, \tilde{\mathcal{R}}_1^1, \tilde{\mathcal{R}}_2^1\}$ where the *confidence* of each association in the CAC are C_1:{1,0.2,0.8} and C_2:{1,0.5,0.55}. Using the measure defined by Formula 8, C_1 scores higher than C_2 (0.5 > 0.475). However, intuitively the contextual association in C_1 with 0.8 *confidence* indicates that the ADRs are more likely to be caused by one of the individual drugs. In this example, C_2 should score higher since all of its contextual associations have relatively lower *confidence* as compared to the target association. To overcome this, we now introduce the coefficient of variation to penalize the CAC with diverse contextual associations w.r.t their *confidence*:

$$contrast_{cv}(C) = contrast_{avg}(C) \times G(C - \mathcal{R}), \quad (9)$$

where

$$G(\mathcal{S}) = (1 - \theta \cdot C_v(\mathcal{S})), \quad (10)$$

$C_v(\mathcal{S})$ computes the coefficient of variation of the *confidence* set of a set of associations $\mathcal{S}$, while θ denotes a user-specified parameter ($0 \leq \theta \leq 1$) that controls the effect of this penalty. Using the previous example with $\theta = 0.75$, then $contrast_{cv}(C_1) = 0.18$ and $contrast_{cv}(C_2) = 0.45$ where $contrast_{cv}$ is in favor of C_2 now.

A drug-safety evaluator is typically knowledgeable about the individual drugs but may be less experienced with unknown MDARs. To expose more complicated cases, MARAS assigns more weight to the contextual association with less drugs. For example, if there are 3 drugs in the target association, the weak association between each individual drug and the ADRs is more important than the weak association between any 2 of the drugs and the ADRs. By considering this, the CAC that involves more drugs should get higher score so that it is pointed out to the drug-safety evaluator. Therefore, we design the final *contrast* score as below:

$$\frac{1}{n} \sum_{i=1}^{n} \frac{1}{m} \sum_{j=1}^{m} (\mathcal{P}_c(\mathcal{R}) - \mathcal{P}_c(\tilde{\mathcal{R}}_j^i)) \times H(i,n) \times G(\{\tilde{\mathcal{R}}^i\}), \quad (11)$$

where $H(i,n)$ is a weighting function that is inversely proportional to the number of drugs in an association, i the number of drugs in $\tilde{\mathcal{R}}_j^i$, n the number of drugs in $\mathcal{R}$, and $\{\tilde{\mathcal{R}}^i\}$ denotes the set of contextual associations with the same number of drugs (i). In our experiment, $H(i,n)$ is chosen to be a linear decay function where $H(i,n) = (1 - (i-1)/n)$, though other functions are possible.

4 THE MARAS SYSTEM

We have designed the MARAS framework to efficiently signal MDARs from ADR reports using ARL foundation.

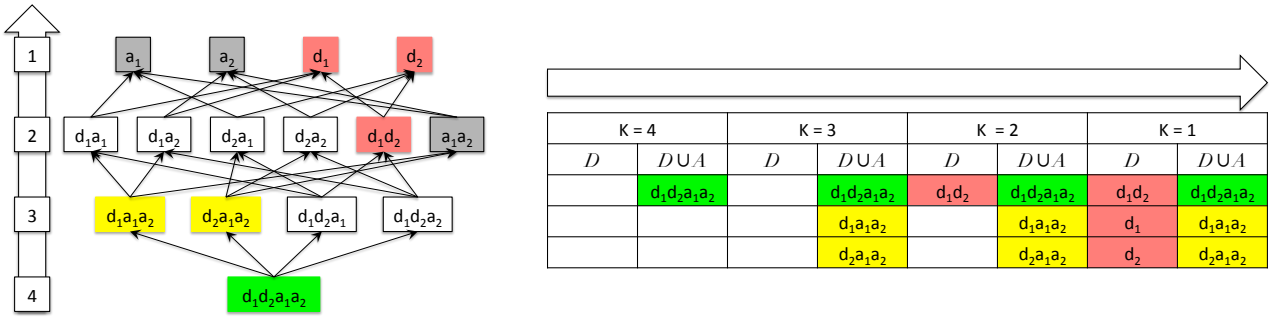

Figure 2: An illustration of CAC generation process.

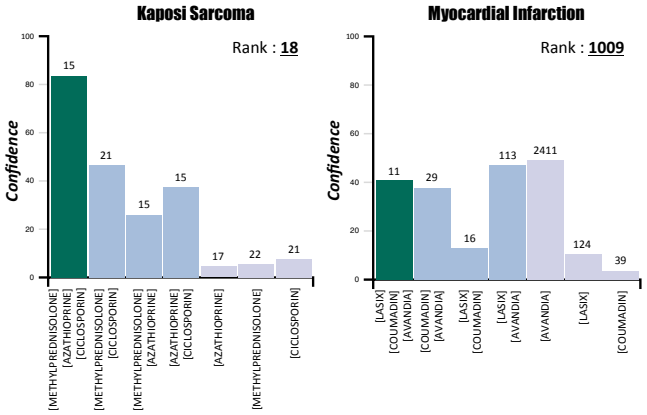

Figure 3: Visualizing CAC with defined *contrast* scores.

Algorithm 1: MDAR Signaler

A: **Itemset Generator**
Algorithm `Itemset Miner()`
k = 2, $\mathcal{L}_1$.add(items) //frequent items
while $\mathcal{L}_k \neq \emptyset$ **do**
C_k = apriori-gen($\mathcal{L}_{k-1}$)
for $t \in D$ **do**
for $c \in C_k$ **do**
Increase(c,t)
TagExplicitSupp(c,t)
$\mathcal{L}_k = \{c \; in \; C_k - c.count \geq min_supp\}$
for $l \in \mathcal{L}_{k-1}$ **do**
RemoveIrrelevant(l)
TagImplicitSupp(l, $\mathcal{L}_k$)
k++
return $\mathcal{L}$

B: **CAC Generator**
[1] **Algorithm** `MCAC()`
k = height($\mathcal{L}$)
while *k ¿ 0* **do**
for $l \in \mathcal{L}_k$ **do**
if *l is explicitly or implicitly supported* **then**
CACs = Mine($\mathcal{L}$,l)
k−−
return CACs
[2] **Procedure** `Mine(`*Itemset l*`)`
C = parents of l **while** $C \neq \emptyset$ **do**
for $l \in C$ **do**
FormCAC(l)
C_{sub}.add(parents of l)
$C = C_{sub}$
return CAC

4.1 Data Processor

The Data Processor extracts and cleans the drug names and adverse events from a given ADR report and prepares this extracted data into the required transactional format for association learning by the MDAR Signaler. Any duplicate reports are removed. Each transaction is linked to its original report via a case id so that the drug-safety evaluator can trace back to the original reports that contain the interested MDAR association for further investigation.

4.2 MDAR Signaler

The MDAR Signaler learns the non-spurious Drug-ADR associations along with their contextual associations needed to form the CACs. It then computes the *contrast* score for each CAC to measure the strength of the MDAR signal. The process involves itemset generation and CAC generation.

In the first step, MARAS adapts an *Apriori* [5] like method with pruning to generate the necessary itemsets and highlight the ones that can form non-spurious Drug-ADR associations. These itemsets are maintained in a lattice structure (Figure 2(left)). The itemset generation algorithm (IGA) (Algorithm 1.A) starts from evaluating the frequency of the itemsets with the smallest cardinality and extends upward the lattice to the larger itemsets.

Itemset Pruning. According to the *confidence* model in Formula 4, two types of itemsets are needed to form an association, namely, itemsets that contain both drugs and ADRs ($\mathcal{D} \cup \mathcal{A}$) and itemsets that only contain drugs ($\mathcal{D}$). Therefore, itemsets that only contain ADRs are pruned immediately after forming larger itemset candidates. According to Lemma 3.5, the itemset that contains all items in non-spurious Drug-ADR associations has to be closed. An itemset is closed if there exists no immediate superset that has the same frequency [18]. Since a MDAR signal involves at least two drugs, all closed itemsets with at least two drugs and at least one ADR have to be identified. Once the larger itemsets are generated, IGA follows their subset links to find and tag the smaller itemsets that satisfy the above constraints. If an itemset contains the exact drugs and ADRs in a report, it is tagged as "explicitly supported". Otherwise, it is tagged as "implicitly supported" if it is closed.

In the second step, using the constructed lattice, the CAC generation algorithm (CGA) (Algorithm 1.B) generates the CACs and calculates their corresponding *contrast* score. It again traverses the lattice in a bottom-up fashion and finds itemsets with an "explicitly supported" or "implicitly supported" tag. For each such

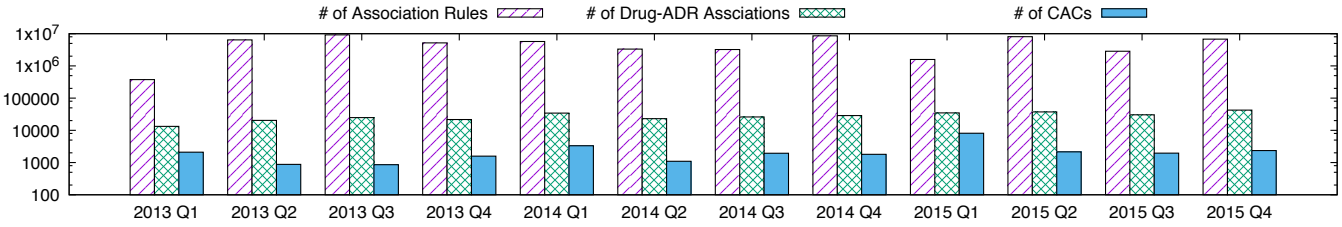

Figure 4: Number of regular associations vs. Drug-ADR associations including spurious ones vs. CACs.

itemset, CGA (Algorithm 1.B.[2]) follows the subset links and recursively generates the target Drug-ADR association and all its contextual associations. For example, in Figure 2, $d_1d_2a_1a_2$ is a closed itemset and thus forms the target association. Following the subset links to level 3, $d_1a_1a_2$ and $d_2a_1a_2$ form the basis for two contextual associations because they contain the identical ADR as the target association but fewer drugs. Moving up to level 2, d_1d_2 is the antecedent of the target association. d_1 and d_2 on level 1 are the antecedents of the contextual associations. This process is illustrated in Figure 2. Once a CAC is generated, its *contrast* is computed using Formula 11.

4.3 MARAS Explorer

Many visuals are possible in MARAS, the default visualization of each signal represented by a CAC is a bar chart depicted in Figure 3. Each bar represents a Drug-ADR association. The height of the bar is the *confidence* of the Drug-ADR association. The darkest bar represents the target association and the lighter the color the less drugs compose the contextual association. The number on the top of each bar shows the number of reports that contain it. The display links to the original reports for further investigation. The CAC is also indexed by the drugs and ADRs so that the drug-safety evaluator can quickly filter and search interesting signals.

5 EXPERIMENTAL EVALUATION

The FAERS Data Source. We work with ADR reports from FAERS, a reporting system and database maintained by the FDA as a part of its post-marketing drug safety surveillance program. It contains million of records about adverse events and medication errors. To ensure the **reproducibility** of this experiment, we used the public version of the FAERS [3] data available quarterly from 2013-15. We selected the mandatory reports submitted by manufacturers marked as expedited (EXP). Each quarter has 100k - 160k reports, 30k - 37k reported drugs and 9k - 10k reported ADRs.

5.1 Number of the MDAR Signals

Figure 4 shows that the amount of signals generated by MARAS is greatly reduced as compared to the overwhelming number of associations that would be produced by state-of-the-art ARL methods such as *confidence* or *RR* (Formula 5). With our effective pruning strategy, MARAS avoids generating irrelevant and spurious associations and thus produces a relatively small number of key signals represented as CACs.

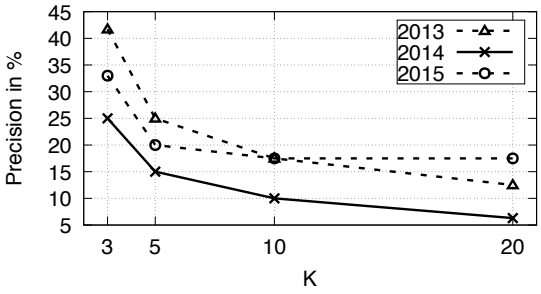

Figure 5: Precision of top K MARAS MDAR signals.

5.2 Quality of MDAR Signal

The main purpose of MARAS is to alert the drug-safety reviewers about possibly unknown MDAR cases collected through the post-market surveillance programs. There is no benchmark database that can be used to systematically evaluate how one should most effectively signal MDARs using ADR reports i.e., no "golden standard". Therefore, one of our evaluation strategies is to evaluate the effectiveness of MARAS by measuring the precision in terms of a hit of a known MDARs. The two sources we used are Drugs.com [1], a FDA recommended resource for obtaining information on known MDARs and DrugBank [11], a drug database that contains comprehensive biochemical and pharmacological information providing insights on MDARs. Figure 5 shows the precision of MARAS within the top k results. Precision is defined by the ratio of the number of hits to the number of the signals. "Precision at K" measures the accuracy of MARAS for signaling the known MDAR as well as the effectiveness of the *contrast* measure for ranking the returned signals. The precision of MARAS for each year is an average precision on 4 quarters data. There are relatively more hits in the higher ranked results, thus proving the effectiveness of our ranking strategy.

5.3 Case Study

Here, we report a case study on three top signals detected by MARAS. The goal of our case study using FAERS ADR reports is to validate the top ranked MDARs identified by MARAS through domain knowledge resources.

Case I: Eliquis and Ibuprofen (Detected and ranked 2^{nd} by MARAS in 2014-Q2 dataset). *Eliquis (Apixaban)*, an anticoagulant for the treatment of venous thromboembolic events is used to prevent platelets in the blood from sticking together and forming a blood clot. *Ibuprofen* is a nonsteroidal anti-inflammatory drug used to reduce inflammation and pain in the body. According to Drugs.com and DrugBank, using these two drugs together may increase the anticoagulant activities of *Apixaban*, lowering the body's

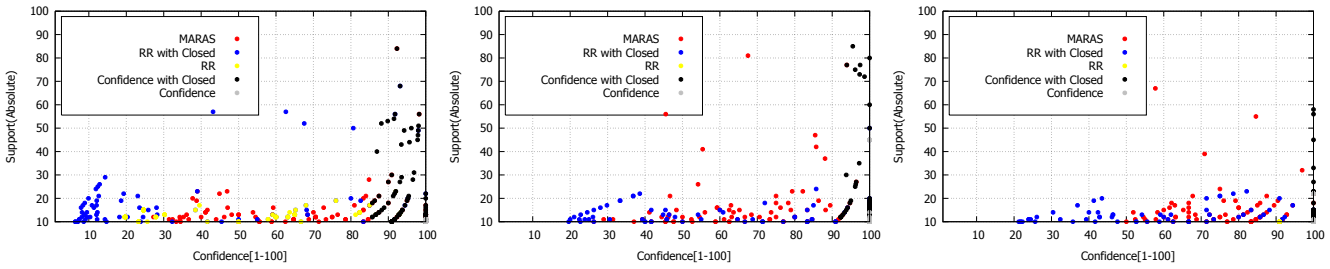

Figure 6: Top 100 signals by different methods in the space of $support$ and $confidence$ of the 4^{th} quarter of 2013, 2014 and 2015.

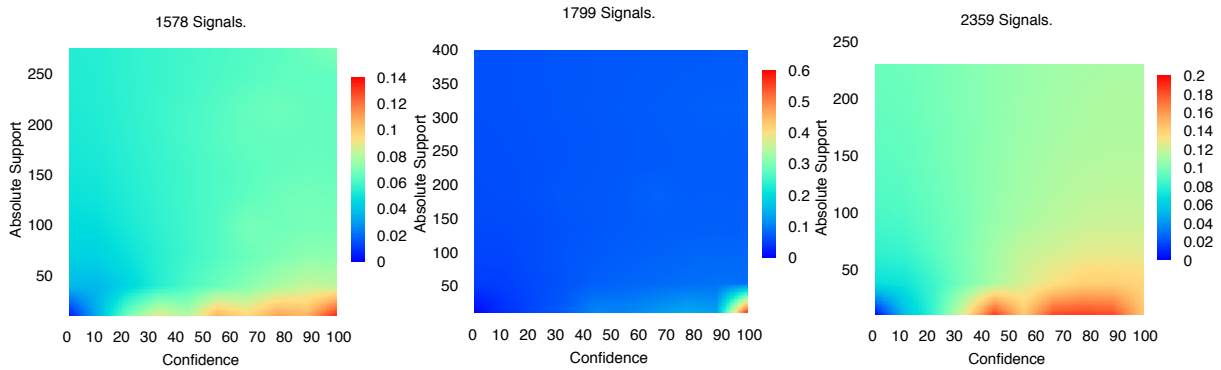

Figure 7: Heatmap of the MDAR signals by MARAS in the space of $support$ and $confidence$ of the 4^{th} quarter of 2013, 2014 and 2015.

ability to form clots and may cause increased bleeding, including severe and sometimes fatal hemorrhage.

Case II: Ondansetron and Lithium (Detected and ranked 1^{st} by MARAS in 2014-Q3 dataset). *Ondansetron* is used to prevent nausea and vomiting that may be caused by surgery or by medicine to treat cancer. *Lithium* is used to treat the manic episodes of bipolar disorder. According to DrugBank, "*Lithium* may increase the neurotoxic activities of *Ondansetron*". Neurotoxicity occurs when the exposure to natural or man-made toxic substances (neurotoxicants) alters the normal activity of the nervous system [2]. According to Drugs.com, "using the two drugs together can increase the risk of a rare but serious condition called the serotonin syndrome, which may include symptoms such as confusion, hallucination, seizure, extreme changes in blood pressure, increased heart rate, fever. Severe cases may result in coma and even death".

Case III: Abilify and Ramipril (Detected and ranked 1^{st} by MARAS in 2015-Q3 dataset). *Abilify (Aripiprazole)*, an antipsychotic medication is used to treat the symptoms of psychotic conditions such as schizophrenia and bipolar I disorder. *Ramipril*, an ACE inhibitor is used to treat high blood pressure or congestive heart failure. According to Drugs.com and DrugBank, these two medications taken in combination can have an additive effect in lowering blood pressure and can cause headache, dizziness, fainting, and/or changes in pulse or heart rate.

5.4 Comparison to State-of-the-Art Baselines

Table 2 shows top 5 MDAR signals generated each from 2015 Q3 data by three different methods namely *Confidence* [22], *Reporting Ratio* [14] (Lift) and MARAS as depicted in the columns one, two and three respectively. The first two columns show the associations between drugs and ADRs ranked by their *confidence* and *RR* values respectively. These two methods do not filter spurious associations. As a result, there are many similar redundant and possibly misleading signals.

In contrast, top ranked signals generated by MARAS are more diverse as compared to those produced by the first two methods. Worse yet, the top ranked signals produced by MARAS signals on interaction between *Rampiril* and *Abilify* as verified via a case study is ranked $2{,}436^{th}$ by *confidence* and $16{,}984^{th}$ by *RR*. Similarly, the second top ranked association by MARAS that shows interaction between *Xgeva* and *Prednison* can lead to osteonecrosis of jaw is ranked $2{,}166^{th}$ by *confidence* and $9{,}312^{th}$ by *RR*. Thus by using the *Confidence* or *Reporting Ratio* (RR) we would risk important findings staying hidden in the association set. Hence we can deduce that MARAS successfully detects non-spurious and non-redundant MDARs, which other methods fail to detect.

Next, we plot the *support* and *confidence* of the top 100 signals generated by different methods. In Figure 6, "RR with Closed" and "Confidence with Closed" refer to signals generated using non-spurious Drug-ADR associations with *RR* and *confidence* measure respectively instead of the *contrast* measure. In general, the signals including the spurious associations are located within a small region (sometimes a single overlapping location) because many redundant associations (associations with very similar items) exist within the top scored signals. Signals using the non-spurious associations tend to spread out over the space. As discussed above, similar results are observed from Table 2, where MARAS captures a diverse set of MDARs compared to other methods. *RR* measure produces more low *confidence* signals where as MARAS's signals are located in between "RR with Closed" and "Confidence with Closed" in terms of their *confidence* values. Figure 7 shows the heatmaps of the

Table 2: Top 5 MDAR signals from 3^{rd} Quarter of 2015.

Rank	Confidence		Reporting Ratio		MARAS	
1	Procyclidine	Bradycardia	Citalopram	Suicidal Ideation	Abilify	Drug Interaction
	Amlodipine		Fluoxetine	Inhibitory Drug Interaction	Ramipril	
	Doxazosin		Zoladex			
2	Procyclidine	Fall	Citalopram	Inhibitory Drug Interaction	Xgeva	Osteonecrosis of the Jaw
	Amlodipine		Fluoxetine	Depressive Symptom	Prednison	
	Doxazosin		Zoladex			
3	Procyclidine	Fall	Citalopram	Suicidal Ideation	Lisinopril	Neutrophil Count Decreased
						Influenza
	Amlodipine	Bradycardia	Zoladex	Inhibitory Drug Interaction	Prednisolone	White Blood Cell Count Decreased
				Depressive Symptom		Blepharitis
						Lower Respiratory Tract Infection
4	Procyclidine	Bradycardia	Citalopram	Suicidal Ideation	Methadone	Enterococcal Infection
	Amlodipine		Zoladex	Inhibitory Drug Interaction	Olanzapine	
5	Procyclidine	Bradycardia	Citalopram	Suicidal Ideation	Ibuprofen	Suicide Attempt
	Doxazosin	Fall	Zoladex	Depressive Symptom	Nifedipine	

MARAS signals displayed in the space of *support* and *confidence*. The color represents the *contrast* score of the signal. Highly scored signals are usually located from 40% to 100% in *confidence* values.

6 RELATED WORK

MDARs. [20, 21] used statistical methods to find interactions among drug classes. However, these methods are typically designed for a particular class of drugs or ADRs only. Hence, they do not consider all reported drugs and ADRs crucial for drug-surveillance. Unsupervised methods in particular association rule mining has been used in the medical domain to explore drug related ADRs [12, 13, 16]. These methods considered the identification of ADRs related to a single drug, rather than a combination of drugs.

ARL for Signaling MDAR. [14, 15] used ARL with *Reporting Ratio* (RR) and Proportional Reporting Ratio (PRR) respectively to find drug interactions triggering a set of ADRs. However, these approaches do not consider the association of individual drugs with the ADRs within a drug combination therefore providing many false positive signals. Cai et al [8] uses ARL and defines interestingness based on causal relation between two interacting drugs and ADRs. Moreover, none of these approaches remove spurious or misleading rules as introduced by our work.

Interestingness in ARL. Various attempts have been made in the literature to reduce the number of the generated rules and rank the most interesting ones [6, 19, 24]. However the majority of these measures are either for classification rules or are subjective measures that need domain specific knowledge to define interestingness. Sub-rules based interestingness has been studied by [10], where interestingness is defined as an unexpected confidence among a neighborhood. The interestingness based on sub-rule's confidence known as improvement [14] ensures that for every rule none of its simplifications offer any predictive advantage over it. None of these methods capture the most interesting associations among multiple drugs and ADRs.

7 CONCLUSION

In this work we have designed the **MARAS** technology that signals interesting MDAR using contextual information. We defined the non-spurious association that is appropriate for MDAR signals, and proposed the *contrast* measure to find the most severe MDAR signals. When compared with state-of-the-art methods, **MARAS** clearly detects an accurate and diverse set of non-spurious MDAR signals, as confirmed by our case study on FAERS ADR reports data.

8 ACKNOWLEDGMENT

We are grateful to Dr. Marni Hall, Sr. Program Director, Regulatory Science, OSE, FDA for introducing us to Pharmacovigilance in general and the ADR problem in particular, and to Deputy Director, Regulatory Science, OSE, FDA, Suranjan De for assisting us with the FAERS system.

REFERENCES

[1] Drugs.com. http://www.drugs.com. [Accessed 2016-04-20].
[2] National institute of neurological disorders and stroke. http://www.ninds.nih.gov/disorders/neurotoxicity/neurotoxicity.htm. [Accessed 2016-10-23].
[3] Openfda. https://open.fda.gov/drug/event. [Accessed: 2016-04-20].
[4] FDA adverse event reporting system (FAERS). http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ucm082193.htm. [Online; accessed 11-March-2016].
[5] R. Agrawal, T. Imieliński, and A. Swami. Mining association rules between sets of items in large databases. In *SIGMOD Record*, volume 22, pages 207–216. ACM, 1993.
[6] Y. Bastide, N. Pasquier, R. Taouil, G. Stumme, and L. Lakhal. Mining minimal non-redundant association rules using frequent closed itemsets. In *Computational Logic*, pages 972–986. Springer, 2000.
[7] R. J. Bayardo Jr and R. Agrawal. Mining the most interesting rules. In *SIGKDD*, pages 145–154. ACM, 1999.
[8] R. Cai, M. Liu, Y. Hu, B. L. Melton, M. E. Matheny, H. Xu, L. Duan, and L. R. Waitman. Identification of adverse drug-drug interactions through causal association rule discovery from spontaneous adverse event reports. *Artificial Intelligence in Medicine*, 2017.
[9] T. Chan. Adverse interactions between warfarin and nonsteroidal antiinflammatory drugs: mechanisms, clinical significance, and avoidance. *The Annals of pharmacotherapy*, 29(12):1274–1283, 1995.
[10] G. Dong and J. Li. Interestingness of discovered association rules in terms of neighborhood-based unexpectedness. In *PAKDD*, pages 72–86, 1998.
[11] V. L. et al. Drugbank 4.0: shedding new light on drug metabolism. *Nucleic Acids Research*, 42(Database-Issue):1091–1097, 2014.
[12] D. M. Fram, J. S. Almenoff, and W. DuMouchel. Empirical bayesian data mining for discovering patterns in post-marketing drug safety. In *SIGKDD*, pages 359–368. ACM, 2003.
[13] M. R. Hacene, Y. Toussaint, and P. Valtchev. Mining safety signals in spontaneous reports database using concept analysis. In *Artificial Intelligence in Medicine*, pages 285–294, 2009.
[14] R. Harpaz, H. S. Chase, and C. Friedman. Mining multi-item drug adverse effect associations in spontaneous reporting systems. *BMC Bioinformatics*, 11(S-9):S7, 2010.
[15] H. Ibrahim, A. Saad, A. Abdo, and A. S. Eldin. Mining association patterns of drug-interactions using post marketing fdafis spontaneous reporting data. *Journal of biomedical informatics*, 60:294–308, 2016.
[16] H. Jin, J. Chen, H. He, G. J. Williams, C. Kelman, and C. M. O'Keefe. Mining unexpected temporal associations: Applications in detecting adverse drug reactions. *IEEE Trans. Information Technology in Biomedicine*, 12(4):488–500, 2008.
[17] R. J. B. Jr., R. Agrawal, and D. Gunopulos. Constraint-based rule mining in large, dense databases. In *ICDE*, pages 188–197. IEEE, 1999.
[18] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal. Discovering frequent closed itemsets for association rules. In *ICDT*, pages 398–416. Springer, 1999.
[19] S. Sahar. Interestingness measures - on determining what is interesting. In *Data Mining and Knowledge Discovery Handbook, 2nd ed.*, pages 603–612. 2010.
[20] N. P. Tatonetti, G. H. Fernald, and R. B. Altman. A novel signal detection algorithm for identifying hidden drug-drug interactions in adverse event reports. *JAMIA*, 19(1):79–85, 2012.
[21] N. P. Tatonetti, P. Y. Patrick, R. Daneshjou, and R. B. Altman. Data-driven prediction of drug effects and interactions. *Science translational medicine*, 4(125):125ra31–125ra31, 2012.
[22] L. Wei and J. Scott. Association rule mining in the us vaccine adverse event reporting system (vaers). *Pharmacoepidemiology and Drug Safety*, 24(9):922–933, 2015.
[23] I. H. Witten and E. Frank. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005.
[24] M. J. Zaki. Generating non-redundant association rules. In *SIGKDD*, pages 34–43. ACM, 2000.