PROBABILISTIC MEASURES FOR INTERESTINGNESS OF DEVIATIONS – A SURVEY

Adnan Masood[1] and Sofiane Ouaguenouni[2]

[1] Graduate School of Computer and Information Sciences. 3301 College Avenue, Fort Lauderdale-Davie, Florida 33314-7796, USA
adnan@nova.edu
[2]Oracle Corporation, 800 Royal Oaks Drive Monrovia, CA 91016, USA
Sofiane.Ouaguenouni@oracle.com

ABSTRACT

Association rule mining has long being plagued with the problem of finding meaningful, actionable knowledge from the large set of rules. In this age of data deluge with modern computing capabilities, we gather, distribute, and store information in vast amounts from diverse data sources. With such data profusion, the core knowledge discovery problem becomes efficient data retrieval rather than simply finding heaps of information. The most common approach is to employ measures of rule interestingness to filter the results of the association rule generation process. However, study of literature suggests that interestingness is difficult to define quantitatively and can be best summarized as, a record or pattern is interesting if it suggests a change in an established model.

Almost twenty years ago, Gregory Piatetsky-Shapiro and Christopher J. Matheus, in their paper, "The Interestingness of Deviations," argued that deviations should be grouped together in a finding and that the interestingness of a finding is the estimated benefit from a possible action connected to it. Since then, this field has progressed and new data mining techniques have been introduced to address the subjective, objective, and semantic interestingness measures. In this brief survey, we review the current state of literature around interestingness of deviations, i.e. outliers with specific interest around probabilistic measures using Bayesian belief networks.

KEYWORDS

Interestingness, probabilistic interestingness measures, Bayesian belief network, support vector machines, neural networks, random forests, outlier, rare entities.

1. INTRODUCTION

The concepts of interestingness and outliers are arduous to define and quantify. Study of literature suggests that there is no agreement on formal definition of "interestingness"; this notion is best summarized as, "record or pattern is interesting if it suggests a change in an established model." This multi-disciplinary concept portrays interestingness as an entity that captures the impression of "novel" or "surprising". In search of the question "What's Interesting?", [1] attempts to answer by stating that *"Interestingness depends on the observer's current knowledge and computational abilities. Things are boring if either too much or too little is known about them, if they appear trivial or random."*

A similar multi-disciplinary construct like interestingness manifests the rare class entities in data and is often referred to as *anomaly*. Anomalies are data points or entities that do not agree with

DOI : 10.5121/ijaia.2013.4201

the expected model. In research literature, data mining and machine learning communities, the classification problem of outlier analysis and detection is often referred to with various different terminologies. As noted by Chandola [2] in their anomaly detection survey, it is cited as anomaly, novelty, chance discovery, exception mining, mining rare classes, and, informally, finding the needle in the haystack. Within the context of data mining, anomalies are the data points which not represented by the model, i.e. data points from a never before seen class. Similarly, in statistics, rare class entities are embodied as novelty, deviations, anomalies or outliers.

In this paper, we review the current state of interestingness measures specifically probabilistic interestingness measures, in context of anomalies and unexpectedness. These topics have a large body of academic work devoted although the distinction is often subjective; [3] notes *one man's outlier is another man's novelty*. Consider the definitions used by Popescu [4], Chandola [2], Markou [5] and Tan [6] respectively:

This process of retrieving the areas that are "interesting" for the understanding of the event is called "anomaly detection. [4],

"Anomalies are patterns in data that do not conform to a well-defined notion of normal behavior." [2]

"Novelty detection is the identification of new or unknown data or signals that a machine learning system is not aware of during training." [5]

"An outlier is an observation that differs so much from other observations as to arouse suspicion that it was generated by a different mechanism." [6]

This preface now raises the correlation question; are all interesting entities outliers and vice versa? From literature, we observe a repeated sense that if something is new or different, then it is interesting and embodied as novelty, deviations, anomalies or outliers. It is commonly observed that not all outliers are interesting and not all interesting trends in the data are outliers. Based on the established trend in the study of outlier interestingness literature such as [2], we can nevertheless classify most interesting trends as anomalies.

The differences in the preceding are subtle but important. Historically, outlier detection has been the focus of statistical research as a means of "cleaning" data. Since outliers can significantly affect statistical data measures such as mean or standard deviation, outlier detection was deemed a preprocessing procedure to ensure data quality. There are various ways of detecting anomalies, namely statistical methods, nearest-neighbor methods, classification-based methods, clustering-based methods, information theoretic methods, and spectral techniques.

Our research focus in this paper is to review the knowledge discovery interestingness measures based on unexpectedness, specifically those which rely on probabilistic graphical models. Using graphical models, like belief networks, to represent data as background knowledge, researchers can use both the inherent power of explanation as well as probabilistic inference. Therefore, instead of representing outliers in a black-box manner, they can offer interestingness-based sensitivity measures to explain why an anomaly is *potentially* interesting.

2. AN OVERVIEW OF INTERESTINGNESS SURVEYS

Few surveys of interestingness in production rules have been performed in the past by [7-9] and one most recently by [10]. In first survey in 1999 by [7] on "Knowledge Discovery and Interestingness Measures," the researchers examined an enumeration of 17 measures of rule interestingness, offering a brief description of each rule. These rules range from Agrawal and Srikant's Item-set measures [11, 12], such as "interesting rules exceed a certain threshold of confidence $(P(B|A))$ and support $(P(AB))$ to more complex rules including Piatetsky-Shapiro's Rule-Interest Function [13]. Smyth and Goodman's J-Measure [14], Major and Mangano's Rule Refinement [15], Klemettinen et al. Rule Templates [16], Matheus and Piatetsky-Shapiro's Projected Savings [17], Hamilton and Fudger's I-Measures [18], Silbershatz and Tuzhilin's Interestingness [19], Kamber and Shinghal's Interestingness [20], Hamilton et al. Credibility Generalized [21], Liu et al. General Impressions [22], Gago and Bento's Distance Metric [23], Freitas' Surprisingness [24], Gray and Orlowska's Interestingness [25], Dong and Li's Interestingness [26], Liu et al. Reliable Exceptions [27] and Zhong et al. Peculiarity [28].

This assortment of objective and subjective measures is further classified as either distance-based, probabilistic, or syntactic. [3] provides brief description of these measures as follows.

- **Piatetsky-Shapiro**: Deviation from statistical independence between the antecedent and the consequent: $(P(A|B) - P(A)P(B))$; the higher the deviation, the more interesting is the measure.
- **J-Measure: T**he average information content of a classification rule where given attributes are discrete valued,

$$P(AB)\log\frac{P(AB)}{P(B)} + P\ (A \sim B)\log\frac{P(\sim B|A)}{P(\sim B)}$$

The higher the J-values are, more interesting the measure is.

- **Gaga-Bento**: Distance metric to measure the distance between two rules, where distance is a function of the number of overlapping attributes common to two rules. Rules which are very distant from other rules are more interesting i.e. qualify to be outliers.
- **Zhong -** Peculiarity is a distance metric. In this case if the antecedents to a rule are similar to those of other rules, but its consequents are different, then the rule is interesting.
- **Silbershatz-Tuzhilin -** Measure of the extent to which a soft belief (hypothesis with "low" confidence) is changed in light of new evidence.
- **Freitas -** The explicit search for occurrences of Simpson's paradox, a seemingly self-contradictory statistical occurrence wherein conclusions drawn from a large data set are contradicted by conclusions drawn from subsets of the large data set.
- **Klemettin -** Rule templates are specified to identify the syntactic structure of either desired rules or undesired rules.

The survey performed by [7] provides a combination of both objective and subjective rules creating a good overall survey of researchers' efforts to define the interestingness of association rules. When Hilderman [29] reviewed the field again four years later, an additional 33 rules had been developed due to the field's growth.

It is important that along with analyzing interestingness in general, the key notion of interestingness in deviations originally emerged from [30], in which researchers introduced KEFIR; a discovery system for data analysis and report generation from relational databases. With a health care case study, Piatetsky-Shapiro and Matheus defined their view of interestingness as statisticians' view of an optimal utility function. The algorithm iterates through subsets of data known as "sectors" by comparing trend and normative deviations of all the measures relevant to the top sector (the entire population). The outliers are then ranked by the measure of trend and normative deviation.

The 2005 survey paper on Interestingness Measures for Knowledge Discovery [8] evaluated then-current research literature on the various techniques for determining the interestingness of patterns discovered by the data mining process. During the analysis, McGarry defines objective measures as those that are based upon the structure of the discovered patterns, while subjective measures are based upon user beliefs or biases regarding relationships in the data. This survey identifies the primary disadvantage of a subjective or user-driven approach: that it limits the knowledge discovery process to user’s hypothesis. In contrast, objective patterns are data-driven and therefore may manifest knowledge which is already known. This ultimately poses a research challenge to unify objective and subjective measures. The taxonomy of interestingness measures as noted by McGarry [8] follows.

- Objective
    - Coverage
    - Support
    - Accuracy
- Subjective
    - Unexpected
    - Actionable
    - Novel

McGarry’s [8] survey of interestingness measures for knowledge discovery approaches the topic in terms of data mining and knowledge discovery. Included in the paper as objective measures are standard statistical/information theoretic measures such as Shannon Entropy [31], Lorenz measure, Gini Index, Kullback-Leibler Distance, and the Atkinson Inequality, as well as the measures reviewed earlier by Hilderman [7, 29]. The term "distance" in this context is actually a measure of difference. None of the measures used are distance measures in the geometric sense.

Bourassa [3] notes an important contribution of McGarry’s paper as the coverage of the work of Lenat, Walsh, Zytkow, Gaines, Kulkarni, and Ludwig all of whom were involved in automated scientific discovery as cited in their work [32]. McGarry concludes with future research directions, primarily highlighting the strain between the objective and subjective approaches to finding interesting association rules. As discussed earlier in this paper regarding objective and subjective measure, McGarry states that subjective rules must necessarily constrain rule discovery to what a user expects to find and, consequently, unanticipated rules are indiscoverable. On the other hand, objective measures of interestingness will find rules that are of no interest to the user, since no context guides the discovery process. McGarry identified the resolution of this strain as an open question. A proposed solution is to find measures of interestingness, such as Simpson's Paradox detection explored by [33], that provide a middle ground to both approaches.

McGarry also realizes a second area of research, i.e. capturing the temporal nature of interestingness. Since all measures of interestingness are based on historical data, and since the human experience is that interest varies over time, it is feasible and necessary to develop measures that accommodate changes in interestingness. McGarry cites Bayesian Networks as a potentially fruitful technology for this application.

3. BAYESIAN NETWORKS

McGarry’s [8] survey elucidates on Bayesian classifiers as an efficient probabilistic way of determining class labels. Probabilistic graphical models can perfectly capture the sensitivity, causal inference, and underlying uncertainty among these diverging attributes. Bayesian Belief Network (commonly known as belief network or BBN) is a special type of probabilistic graphical model which provides relationship between various attributes with the representation, usually a directed acyclic graph. A probability table associates each node in the graph. BBNs are particularly useful in capturing existing knowledge; for example, they can incorporate the knowledge of domain experts into a model.

Innately remarkable in handling uncertainty, Bayesian Belief networks are particularly useful when data contains noise or unknown factors. In such instances, two virtually identical data points can be from different classes. Bourassa [3] notes that fitness and diet are generally accepted to be two strong indicators used to gauge life expectancy. However, fit, healthy eaters have been known to live very short lives. Clearly, in such cases, other unknown factors are present.

A Bayesian classifier estimates the probability of a class label for a given data point. This estimate is known as the posterior probability. The assigned label identifies the class with the largest posterior probability. A Naive Bayesian classifier is constructed under the assumption that the attributes of a data point are conditionally independent. Conditional independence means that the attributes have no correlation. [34] noted that this latter assumption makes the use of Bayesian classifiers appealing. Under the assumption of conditional independence, Naive Bayesian classifiers can be constructed using the simple statistics of each attribute. Consequently they can be constructed from relatively little data. They are robust to irrelevant attributes as these will show a uniform distribution with respect to the class label. In some instances, previously established correlations encourage researchers to relax the assumption of independence of attributes.

The next notable and comprehensive survey was performed by [9] for interestingness measures in data mining. This survey identifies interestingness as a broader concept which constitutes of *conciseness, coverage, reliability, peculiarity, diversity, novelty, surprisingness, utility and actionability*. Bourassa [3] noted it as a very thorough review of interestingness measures and their properties. It distinguishes itself from McGarry's work in departing from a data mining context and instead focusing on measure categorization and behavior. The authors point out that they consider their work complimentary to McGarry's work.

Geng and Hamilton [9] classified these interestingness measures based on the fundamental calculation or methodology for each measure (i.e., utilitarian, probabilistic, syntactic, distance). Majority of interestingness measures cited in Geng's survey are probabilistic in nature. Geng's review highlights the scope of the measures available to three types of rules: association, classification, and summaries (rule sets the paper reiterates the absence of a single definition for interestingness. Based on the diversity of measure definitions, the paper has compiled nine rule-interestingness criteria. They are as follows:

1. Conciseness: A pattern is concise if it contains few attribute-value pairs. A concise pattern is easy to understand, remember, and add to a user's knowledge (extends to sets of patterns).
2. Generality/Coverage: The generality or coverage of a pattern is a measure of how large a subset of the data the pattern covers. Patterns that characterize more information are interesting.
3. Reliability: a reliable pattern describes a relationship in the data that applies to a high percentage of the data.
4. Peculiarity: a pattern is peculiar if, by some distance measure, it lies far from other discovered patterns.

5. Diversity: a pattern is diverse if it consists of elements that differ significantly from each other (extends to sets of patterns).
6. Novelty: a pattern is novel if has never been seen before and could not have been inferred from previously seen patterns;
7. Surprisingness (unexpectedness): the property of a pattern which contradicts existing knowledge or expectations.
8. Utility: the utility of a pattern is measured by its usefulness in reaching a goal (e.g. a business can use a sales pattern or market basket analysis to increase profits).
9. Actionability/Applicability: an actionable pattern enables decision making about future actions in a desired domain.

Geng then reviewed 38 objective, 3 subjective, and 2 semantic interestingness measures for association/classification rules according to the nine interestingness criteria. The interestingness measures were apportioned in three categories summarized as follows:

1. Objective Measures
    a. Based on probability
    b. Based on the form of the rules
        i. Peculiarity
        ii. Surprisingness
        iii. Conciseness
            1. Non-redundant rules
            2. Minimum description length
2. Subjective Measures
    a. Unexpectedness
    b. Novelty
3. Semantic Measures
    a. Utility
    b. Actionability

Bourassa [3] noted that all the interestingness measures of the previous surveys are included in the Geng survey [9] and their properties are matched to their most appropriate use in addressing the criteria. Two additional measures have not been seen previously: "conciseness" and "semantic measures." Table #1 provides a complete list of Probability-Based Objective Interestingness Measures for Rules as shown in Geng’s paper.

Now a prevalent technique in text mining, semantic measures were prescribed to detect novelty of text-mined rules using lexical knowledge. Here, similarly, semantic measures were utilized to consider the meaning of the rules. For instance, a typical utility-based measure may assign weights to transactions according to the importance of the transaction in the data set. Some transactions have more utility than others when determining the support and confidence of rules. These differences can be observed in the [35] example of mining weighted association rules.
The second additional measure, Conciseness, is a form-dependent measure which applies to both rule sets and individual rules. [9] highlights two such measures: one to explicitly scour the non-redundant rule sets and the second to encode data based on a given hypothesis following minimum description length principle (MDL). Geng concludes by describing potential areas of research and identifying choice of interestingness measures that reflect real human interest as an open problem.

Since Geng’s work [9], the most recent survey on knowledge discovery interestingness measures based on unexpectedness is by Kontonasios et al [10] which summarizes the primary features of syntactical and probabilistic approaches to interestingness mining. By surveying syntactical approaches, the authors discuss the use of template rules, fuzzy logic, and unexpectedness via contradiction, users' classifications, and users' dynamics. They discuss another group of methods which exploits the structure of more sophisticated background knowledge, i.e. taxonomies and ontologies. Researchers reviewed a multitude of important publications in this area and provided an exhaustive list of probability-based objective interestingness measures. Rather than explicitly enumerating all the interestingness measures covered in the surveys, a holistic view of the operating principles of the most representative measures as specified by Geng is summarized as part of Table 1.

Table 1 List of Probabilistic interestingness measure

Name	**Formula**
Information Gain	$log \frac{P(AB)}{P(A)P(B)}$
Goodman and Kruskal	$\frac{\sum_i max_j P(A_iB_j) + \sum_j max_i P(A_iB_j) - max_i P(A_i) max_i P(B_j)}{2 - maxiP(Ai) - maxiP(Bj)}$
Gini Index	$P(A)[P(B\|A)2 + P(\neg B\|A)2] + P(\neg A)[P(B\|\neg A)2 + P(\neg B\|\neg A)2] - P(B)2 - P(\neg B)2$
Example and Counter Example Rate	$1 - \frac{P(A\neg B)}{P(AB)}$
Coverage	$P(A)$
Cosine	$\frac{P(AB)}{\sqrt{P(A)P(B)}}$
Conviction	$P(A)P(\neg B)\boxed{P(A\neg B)}$
Confidence	$P(B\|A)$
Collective Strength	$\frac{P(AB) + P(\neg B\|\neg A)}{P(A)P(B) + P(\neg A)P(\neg B)}$
Certainty Factor	$\frac{P(B\|A) - P(B)}{1 - P(B)}$
Added Value	$P(B\|A) - P(B)$
Accuracy	$P(AB) + P(\neg A\neg B)$
Zhang	$\frac{P(AB) - P(A)P(B)}{max(P(AB)P(\neg B), P(B)P(A\neg B))}$
Yules Y	$\frac{\sqrt{P(AB)P(\neg A\neg B)} - \sqrt{P(A\neg B)P(\neg AB)}}{\sqrt{P(AB)P(\neg A\neg B)} + \sqrt{P(A\neg B)P(\neg AB)}}$
Yules Q	$\frac{P(AB)P(\neg A\neg B) - P(A\neg B)P(\neg AB)}{P(AB)P(\neg A\neg B) + P(A\neg B)P(\neg AB)}$
Two-Way Support	$\frac{P(AB) \times log2P(AB)}{P(A)P(B)}$
Support	$P(AB)$
Specificity	$P(\neg B\|\neg A)$
Sebag Schoenauer	$\frac{P(AB)}{P(A\neg B)}$
Relative risk	$\frac{P(B\|A)}{P(B\|\neg A)}$

Recall	$P(A\|B)$
Prevalence	$P(B)$
Piatesky Shapiro	$P(AB) - P(A)P(B)$
Odd's Ratio	$\frac{P(AB)P(\neg A \neg B)}{P(A \neg B)P(\neg BA)}$
Odd Multiplier	$P(AB)P(\neg B)\boxed{P(B)P(A\neg B)}$
Normalize Mutual Information	$\frac{\Sigma i \Sigma j P(AiBj) \times log2(AiBj)}{P(Ai)P(Bj) - \Sigma i P(Ai) * log2P(Ai)}$
Loevinger	$\frac{1 - P(A)P(\neg B)}{P(A\neg B)}$
Linear Correlation Co-efficient	$\frac{P(AB) \quad P(A)P(B)}{\sqrt{P(A)P(B)P(\neg A)P(\neg B)}}$
Lift	$\frac{P(B\|A)}{P(B)}$ or $\frac{P(AB)}{P(A)P(B)}$
Leverage	$P(B\|A) - P(A)P(B)$
Least Contradiction	$\frac{P(AB) - P(A\neg B)}{P(B)}$
Laplace Correction	$\frac{n(AB) + 1}{n(A) + 2}$
Klosgen	$\sqrt{\boxed{P(AB)}}(P(B\|A) - P(B))$, $\sqrt{\boxed{P(AB)}}max(P(B\|A) - P(B), P(A\|B) - P(A))$
JMeasure	$P(AB)log(P(B\|A)\boxed{P(B)}) + P(A\neg B)log(P(\neg B\|A)\boxed{P(\neg B)})$
Jacard	$\frac{P(AB)}{P(A) + P(B) - P(AB)}$
IWD	$((\frac{P(AB}{P(A)P(B)})^k - 1) \times P(AB)^m$

4. PROBABILISTIC APPROACHES TO ANOMALY DETECTION

As discussed earlier by [7-9], in the probabilistic methods the background knowledge is usually encoded in a probability distribution of the data, expressed explicitly or specified implicitly. This distribution is often called the background model. [36] notes Bayesian network as a model that encodes probabilistic relationships among variables of interest.

The most significant and pertinent research work related to belief network comprises "Sensitivity analysis of probabilistic graphical models" by [37], research on "Interestingness of frequent item-sets using Bayesian networks as background knowledge" [38, 39] by Jaroszewicz and Simovici of University of Massachusetts Boston, and "Using sensitivity of a Bayesian network to discover interesting patterns" [40] by Rana Malhas and Zaher Al-Aghbari of University of Sharjah. [38, 39] approach was to determine interestingness of an attribute set with respect to a Bayesian network. In their applied belief network embodying work, the authors engaged in performance evaluation and developed algorithms to find interesting item-set and attribute sets.

In the background model approach as discussed by [38, 39], the interestingness is quantified as the deviation between a statistic of the pattern calculated on the empirical data and the one calculated on the background model. For example, one could compare the support of an item-set as measured on the data with the expected support as computed under the background model.
Malhas et al [40] echoed this discovery with an applied interestingness measure titled *"sensitivity"*. These researchers concluded that probabilistic methods have been proven useful for encoding background knowledge obtained either from a human expert or through a dataset. Probabilistic methods allow the researcher to handle not only local but also more global

background knowledge and aggregates, including but not limited to sums over rows and columns of the data matrix representing the data. Using probabilistic approaches, the researcher can also apply general tools such as Monte Carlo sampling approaches and graphical modelling.
Probabilistic methods assign a degree of interestingness to every pattern. [10] observes that such interestingness measures help to rank the resulting patterns. Due to assigned degree of probabilistic interestingness, this approach is deemed more effective than mere filtration.

Silberschatz and Tuzhilin [19, 41] proposed a framework for finding unexpected patterns. Defining unexpectedness in terms of logical contradiction, authors present an algorithm called ZoominUR, which uses a set of user-defined beliefs to seed the search for the patterns that are unexpected relative to these beliefs. Researchers observed that user-defined beliefs can drastically reduce the number of irrelevant and obvious patterns found during the discovery process and that user-defined beliefs are essential to the discovery process in some applications, such as web application log analysis or modern social media graph inquiry.

This framework idea was further formalized by [39] in "*Using a Bayesian Network as Background K*nowledge" and later matured in ""Scalable pattern mining with Bayesian networks as background knowledge" [38]. In this case, interestingness is defined as the absolute difference between the support of an item-set calculated from the actual data and the estimated support inferred from the belief network. Using the foundation set earlier by [38, 39], Malhas and Albaghari [40] introduced IFEMiner algorithm. Dubbed as Interestingness Filtering Engine, the algorithm allowed researchers to mine Bayesian networks for interesting patterns using the networks' relative *sensitivity*. Sensitivity was introduced as a new interestingness measure, discovered by assessing the uncertainty-increasing potential of a pattern on the beliefs of the Bayesian network. Patterns with the highest sensitivity scores are deemed interesting.

In survey of probabilistic approaches, [10] noted the use of taxonomies as background knowledge representation by De Graaf. The interestingness measure proposed in this approach is the smallest deviation between the real support of an item-set in the database and its support as estimated by one generation ancestor item-sets (1GA). The more an item-set deviates from the behaviour of its parents, the more interesting it is considered.

Using User's Interactive Feedback approach, [42] formulated the deviation between the expected and the observed support of an item-set. Xin et al's proposed approach to discover interesting patterns for a particular user comprises a framework which learns a user's prior knowledge from interactive feedback. Researchers studied two model formulations, the log-linear model and the biased belief model, and discuss the strategy to select sample patterns for user feedback.
The technique of swap randomization maintains the first-order statistics of the data. In an iterative data mining setting, swap randomizations were used by [43] to evaluate patterns or encoding properties of previously discovered patterns. Swap randomizations allow the use of generically specified global background information and were modified for encoding more complex kinds of prior knowledge and for real-valued data. Based on Markov Chain–Monte Carlo (MCMC) process, swap randomization requires large number of samples for interesting pattern ranking and is computationally expensive.

Similar to swap randomizations, De Bie [44] demonstrated how the MaxEnt model can be computed remarkably efficiently and can serve the same purpose as swap randomizations. The MaxEnt model considers prior information as constraints on a probabilistic model representing the uncertainty about the data. Researchers represent the prior information by the maximum entropy (MaxEnt) distribution subject to these constraints.

5. CONCLUSIONS

Unexpectedness is defined as a subjective notion of interestingness, while the process of retrieving the areas that are interesting for the understanding of the event is defined as anomaly detection. A pattern is deemed unexpected if it contradicts the user's background knowledge or prior expectations about the data. This paper surveyed various approaches to interestingness with special focus on probabilistic measures. We reviewed and provided highlights for the previous work performed by [7], [29], [8], [9], [2], [3] and [10]

By reviewing both the subjective and objective probabilistic interestingness measures, we conclude that graphical models encode uncertainty in real-world datasets in the most flexible way. By comparing the background knowledge with the input dataset, the interestingness is quantified as the deviation between a statistics of the pattern calculated on the empirical data. We reviewed different probabilistic techniques, including but not limited to swap randomization, maximum entropy method, and use of belief network as background knowledge measuring sensitivity.

Bourassa [3] noted that from the social and cognitive science perspective, an interesting theory challenges audiences' beliefs, but not too much. The relevant conclusion is that interesting is novel, and the degree of novelty indicates the degree of interestingness, defining "trivial" and "random" as two extremes of novelty. In this survey paper, we presented an overview of the proposed approaches in literature in three major groups: the syntactical, probabilistic and semantic approaches. Finally, a brief review of interestingness measures highlights a large body of work devoted to defining and detecting novel or anomalous data. However, we established that anomaly detection has usually focused on identifying patterns for exclusion rather than for interest.

The reviewed approaches to interestingness, anomaly or novelty detection offer diverse definitions for what interestingness is or may be. The definitions often depend on the patterns of the problems being addressed; future research work may seek to establish a correlation between subjective and objective measures which distils the common themes of existing interpretations of interestingness and synthesize a new, unifying definition that can be applied generally to all forms of data analysis.

REFERENCES

[1] J. Schmidhuber, "What's interesting?," 1997.

[2] V. Chandola, et al., "Anomaly detection: A survey," ACM Computing Surveys (CSUR), vol. 41, p. 15, 2009.

[3] M. A. J. Bourassa, "Interestingness: Guiding the Search for Significant Information," Division of Graduate Studies of the Royal Military College of Canada, Royal Military College of Canada, 2011.

[4] A. Popescu, et al., "A remote sensing image processing framework for damage assessment in a forest fire scenario," in Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European, 2012, pp. 2496-2500.

[5] M. Markou and S. Singh, "Novelty detection: a review—part 1: statistical approaches," Signal Processing, vol. 83, pp. 2481-2497, 2003.

[6] P. Dokas, et al., "Data mining for network intrusion detection," in Proc. NSF Workshop on Next Generation Data Mining, 2002, pp. 21-30.

[7] R. J. Hilderman and H. J. Hamilton, Knowledge discovery and interestingness measures: A survey: Citeseer, 1999.

[8] K. McGarry, "A survey of interestingness measures for knowledge discovery," The knowledge engineering review, vol. 20, pp. 39-61, 2005.

[9] L. Geng and H. J. Hamilton, "Interestingness measures for data mining: A survey," ACM Computing Surveys (CSUR), vol. 38, p. 9, 2006.
[10] K. N. Kontonasios, et al., "Knowledge discovery interestingness measures based on unexpectedness," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2012.
[11] R. Agrawal, et al., "Mining association rules between sets of items in large databases," in ACM SIGMOD Record, 1993, pp. 207-216.
[12] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in Proc. 20th Int. Conf. Very Large Data Bases, VLDB, 1994, pp. 487-499.
[13] G. Piateski and W. Frawley, Knowledge discovery in databases: MIT press, 1991.
[14] R. M. Goodman and P. Smyth, "Rule induction using information theory," Knowledge Discovery in Databases (KDD-1991). MITPress, vol. 199, 1991.
[15] J. A. Major and J. J. Mangano, "Selecting among rules induced from a hurricane database," Journal of Intelligent Information Systems, vol. 4, pp. 39-52, 1995.
[16] M. Klemettinen, et al., "Finding interesting rules from large sets of discovered association rules," in Proceedings of the third international conference on Information and knowledge management, 1994, pp. 401-407.
[17] C. J. Matheus, et al., "20 Selecting and Reporting What is Interesting: The KEFIR Application to Healthcare Data," 1996.
[18] H. J. Hamilton and D. R. Fudger, "Estimating DBLearn's potential for knowledge discovery in databases," Computational Intelligence, vol. 11, pp. 280-296, 1995.
[19] A. Silberschatz and A. Tuzhilin, "On subjective measures of interestingness in knowledge discovery," in Proceedings of KDD-95: First International Conference on Knowledge Discovery and Data Mining, 1995, pp. 275-281.
[20] M. Kamber and R. Shinghal, "Evaluating the interestingness of characteristic rules," in Proc. 1996 Int'l Conf. on Data Mining and Knowledge Discovery (KDD'96), Portland, Oregon, 1996.
[21] H. Hamilton, et al., "Machine learning of credible classifications," Advanced Topics in Artificial Intelligence, pp. 330-339, 1997.
[22] B. Liu, et al., "Using general impressions to analyze discovered classification rules," in Proc. of the 3rd International Conference on Knowledge Discovery and Data Mining, 1997, pp. 31-36.
[23] P. Gago and C. Bento, "A metric for selection of the most promising rules," Principles of Data Mining and Knowledge Discovery, pp. 19-27, 1998.
[24] A. Freitas, "On objective measures of rule surprisingness," Principles of Data Mining and Knowledge Discovery, pp. 1-9, 1998.
[25] B. Gray and M. Orlowska, "CCAIIA: Clustering categorical attributes into interesting association rules," Research and Development in Knowledge Discovery and Data Mining, pp. 132-143, 1998.
[26] G. Dong and J. Li, "Interestingness of discovered association rules in terms of neighborhood-based unexpectedness," Research and Development in Knowledge Discovery and Data Mining, pp. 72-86, 1998.
[27] H. Liu, et al., "Efficient search of reliable exceptions," Methodologies for Knowledge Discovery and Data Mining, pp. 194-204, 1999.
[28] N. Zhong, et al., "Peculiarity oriented multi-database mining," Principles of Data Mining and Knowledge Discovery, pp. 136-146, 1999.
[29] R. J. Hilderman and H. J. Hamilton, "Measuring the interestingness of discovered knowledge: A principled approach," Intelligent Data Analysis, vol. 7, p. 347, 2003.
[30] G. Piatetsky-Shapiro and C. J. Matheus, "The interestingness of deviations," in Proceedings of the AAAI-94 workshop on Knowledge Discovery in Databases, 1994, pp. 25-36.
[31] J. Lin, "Divergence measures based on the Shannon entropy," Information Theory, IEEE Transactions on, vol. 37, pp. 145-151, 1991.
[32] S. Colton, et al., "On the notion of interestingness in automated mathematical discovery," International Journal of Human-Computer Studies, vol. 53, pp. 351-375, 2000.
[33] C. C. Fabris and A. A. Freitas, "Discovering surprising patterns by detecting occurrences of Simpson's paradox," Research and Development in Intelligent Systems, vol. 16, pp. 148-160, 1999.
[34] S. B. McGrayne, The theory that would not die: How Bayes' rule cracked the Enigma code, hunted down Russian submarines, and emerged triumphant from two centuries of controversy: Yale University Press, 2011.
[35] S. Lu, et al., "Mining weighted association rules," Intelligent Data Analysis, vol. 5, pp. 211-226, 2001.

[36] P. Garcia-Teodoro, et al., "Anomaly-based network intrusion detection: Techniques, systems and challenges," computers & security, vol. 28, pp. 18-28, 2009.
[37] H. Chan, Sensitivity analysis of probabilistic graphical models: UNIVERSITY OF CALIFORNIA, LOS ANGELES, 2006.
[38] S. Jaroszewicz, et al., "Scalable pattern mining with Bayesian networks as background knowledge," Data Mining and Knowledge Discovery, vol. 18, pp. 56-100, 2009.
[39] S. Jaroszewicz and D. A. Simovici, "Interestingness of frequent itemsets using Bayesian networks as background knowledge," presented at the Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, Seattle, WA, USA, 2004.
[40] R. Malhas and Z. Al Aghbari, "Using sensitivity of a bayesian network to discover interesting patterns," in Computer Systems and Applications, 2008. AICCSA 2008. IEEE/ACS International Conference on, 2008, pp. 196-205.
[41] A. Silberschatz and A. Tuzhilin, "What makes patterns interesting in knowledge discovery systems," Knowledge and Data Engineering, IEEE Transactions on, vol. 8, pp. 970-974, 1996.
[42] D. Xin, et al., "Discovering interesting patterns through user's interactive feedback," in Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, 2006, pp. 773-778.
[43] A. Gionis, et al., "Assessing data mining results via swap randomization," ACM Transactions on Knowledge Discovery from Data (TKDD), vol. 1, p. 14, 2007.
[44] T. De Bie, "Maximum entropy models and subjective interestingness: an application to tiles in binary databases," Data Mining and Knowledge Discovery, vol. 23, pp. 407-446, 2011.

About Authors

Adnan Masood is a doctoral candidate at Graduate School for Computer and Information Science, Nova Southeastern University, FL. His research interests include probabilistic graphical models, machine learning, Bayesian inference and determination of interestingness measures among rare entities using belief networks. Adnan works as a software architect for Green Dot Corporation with expertise in enterprise architecture, distributed systems and application security.

Sofiane Ouaguenouni is an IT Architect at Oracle Corporation. He is an expert in Master Data Management technologies, predominantly data quality tools and algorithms, such as matching and standardization. He holds a PhD in computational physics, and was previously a research scientist at Caltech.