A framework for learning cell interestingness from cube explorations

Patrick Marcel[1[0000−0003−3171−1174]], Veronika Peralta[1[0000−0002−9236−9088]], and Panos Vassiliadis[2[0000−0003−0085−6776]]

[1] University of Tours, Blois, France firstname.lastname@univ-tours.fr
[2] University of Ioannina, Ioannina, Greece pvassil@cs.uoi.gr

Abstract. In this paper, we discuss the problem of organizing the different ways of computing the interestingness of a particular cell derived from a cube in the context of a hierarchical, multidimensional space. We start from an in-depth study of the interestingness aspects in the study of human behavior and include in our survey the approaches taken by computer-science efforts in the area of data mining and user recommendations. We move on to structure interestingness along different fundamental, high level *aspects*, and, due to their high-level nature, we also move towards much more concrete data-oriented definitions of intertestingness aspects.

1 Introduction

Given a cell of a datacube and a user's exploration over this datacube, how to assign to this cell a score reflecting its interestingness for the exploration?

The **significance** of the answer to the above question, cannot be underestimated. A cell is the most granular piece of information in a BI session, thus, in this paper it is the epicenter of our search, both because it can be of value per se, and because the interestingness of groups of cells can be based on the interestingness of individual cells. We need better systems at recommending questions, data and highlights to the users. Understanding what is important for a user is key to this goal and a cell interestingness score is a pre-requisite for this. If we manage to successfully score (i.e., understand) which cells matter more to each user, this would allow to better understand how users navigate cubes by studying logs of user sessions, categorize these cube explorations, and make online recommendations. Apart from the aforementioned practical considerations, from the research point of view, succeeding in structuring the aspects of interestingness will allow to structure our knowledge on existing methods, as well as provide the basis to benchmark and compare such methods and help develop new ones for supporting cube exploration in aspects not successfully covered yet.

To the best of our knowledge, there is a **gap in the literature** in answering the motivating question of the paper. Although interestingness measures have attracted a lot of attention in other communities, like for instance Data Mining, for measuring the interestingness of a pattern [9], or Recommender Systems, for measuring the quality of recommendations [16], to our knowledge there exists no principled study or survey for the interestingness of a cell in cube exploration.

We consider the **context of cube exploration** as follows. We assume a user performing exploratory data analysis over a hierarchical, multidimensional nature of data [15]. After a while, the user acquires an overarching informational goal that their exploration tries to address. During the exploration, the user devises queries to acquire new information. Each such query brings in new data and constructs a new cube that is presented to the user. We call this Q&A a transition, and it constitutes a step in the overall exploration of the user. The user, thus, practically covers areas of the hierarchical multidimensional space at each step (possibly at different levels of granularity), and progressively, for each such area, some kind of expectation on its values is constructed or updated (call it a "model" if you will). In each transition the user makes, each new observation, is (0) relevant or not to the user's informational need and either (1) reinforces the expectation or (2) contradicts it, or (3) just creates expectations for newly explored places where none existed before. Each new observation, is therefore, assessed with respect to its novelty, relevance, surprise and peculiarity. Each such criterion covers a different aspect of interestingness.

Our **contributions** in this paper, are structured as follows: In Section 2 we discuss earlier proposals of interestingness measures and in Section 3, we study the forces that affect interestingness computation and structure them around *high level aspects of interestingness* –specifically, novelty, relevance, surprise and peculiarity. In Section 4, we provide exemplary algorithms and methods for assessing the high level aspects of a cell's interestingness, on the basis of low-level measures, and Section 5 describes the experiments we ran to showcase the framework. Section 6 concludes the paper and suggests open roads for future work.

2 Related work

Although there is little work proposing measures for quantifying the interestingness of a cell in a datacube, several measures can be borrowed from close research areas and adapted to cells. In this subsection we discuss interestingness measures proposed for (i) pattern mining, (ii) cube exploration and summaries, and (iii) recommendation.

Interestingness criteria for pattern mining In [9], the authors point out that interestingness is a broad concept and identify from the literature 9 criteria to determine whether or not a pattern is interesting: conciseness, generality/coverage, reliability, peculiarity, diversity, novelty, surprisingness, utility and actionability/applicability. They categorize these criteria in 3 groups: i) objective measures, based only on the raw data (generality, reliability, peculiarity, diversity, conciseness), like for instance the classical support, ii) subjective measures, considering both the data and the user (surprise and novelty), like for instance the informational content [3], and iii) semantic measures, based on the semantics and explanations of the patterns (utility and actionability), like for instance measures based on user preferences [26]. We note that according to De Bie [3], subjective interestingness is particularly well adapted for *exploratory* data mining, whose goal is to pick patterns that will result in the best updates of the user's belief state, while presenting a minimal strain on the user's resources. One challenge is to define and update the belief of the user. De Bie proposes to model it as a background distribution over patterns representing the belief the user attaches to patterns being present in the data.

Most of the criteria introduced above can be reused in our context, except diversity (that would concern groups of cells) and reliability (since data in cubes are assumed reliable by construction).

Interestingness criteria for summaries In [9], authors also review interestingness measures for what they call summaries, i.e., aggregated cross-tabs corresponding to the result of an OLAP query, where numeric values (i.e., measures) are aggregated by several criteria (i.e., dimensions). Out of the 9 criteria defined for pattern interestingness, 4 are adapted to summaries: diversity (proportional distribution of classes in the summary versus the number of classes), conciseness/generality (level of aggregation), peculiarity (a cell in a summary is peculiar if it is differs from the other cells in the summary) and surprisingness/unexpectedness (a summary is surprising if it deviates from user's expectations). According to the classification of [9], the first three criteria are objective and the last one is subjective. Note that except for peculiarity, and to a lesser extent, conciseness, the criteria concern the interestingness of the whole summary instead of the interestingness of each cell.

To the best of our knowledge, such peculiarity measures are the cornerstone of discovery-driven analysis [22–25] for measuring cell interestingness in the context of cube exploration. Discovery-driven analysis guides the exploration of a datacube by providing users with interestingness values for measuring the peculiarity of the cells in a data cube, according to statistical models, e.g., based on the maximum entropy principle, and leveraging the intrinsic structure of multidimensional information. From an initial user query, the system automatically calculates 3 kinds of interestingness values for each cell in the query result: (i) $SelfExp$ measures the difference between the observed and anticipated values (the latter are calculated statistically by computing the mean of subsets of attributes), (ii) $InExp$ is obtained as the maximum of $SelfExp$ over all cells that are under this cell (those that result from a drill down), and (iii) $PathExp$ is calculated as the maximum of $SelfExp$ over all cells reachable by drilling down along a given path. The DIFF, INFORM and RELAX advanced OLAP operators proposed in [22, 23, 25] use such interestingness values to recommend relevant cells for explaining drops or increases, or for recommending areas of a cube that should surprise the user, based on their history with the cube.

In the context of OLAP, other works propose further measures concerning (or related to) interestingness of a cross-tab, a query result or a set of cells. Without trying to be exhaustive, we mention here some of those works, illustrating the diversity of the proposed measures.

Klemettinen et al. [17] use skewness, as a peculiarity measure of asymmetry in data distribution, for discovering interesting paths and guiding the navigation in a datacube. Given a cuboid, the possible drill-downs are explored, measuring skewness and generating skew-based navigation rules for the more significant paths. Skewness is computed observing the underlying facts (the raw data that is aggregated), looking for outliers or substantial differences with other facts. Based on skewness, Kumar et al. [18] propose interestingness measures based on the unexpectedness of skewness in navigation rules and navigation paths.

Fabris and Freitas [7] defined interestingness measures for attribute-value pairs in a data cube: the I_1 measure reflects the difference between the observed probability of an attribute-value pair and the average probability in the summary and the I_2 measure reflects the degree of correlation among two attributes. Both measures can be seen as value-based conciseness.

Djedaini et al. use supervised classification techniques for learning two interest measures for OLAP queries: *focus*, that indicates to what extent a query is well detailed and related to other queries in an exploration, indicating that the user investigates in details precise facts and learns from this investigation [5], and *contribution*, that highlights to what extent a query is important for an exploration, contributing to its interest and quality [4].

Finally, we mention two recent works [27, 21] that are concerned with detecting the validity of insights gained by users when examining query answers. As other works measuring peculiarity by leveraging the nature of OLAP cubes, this is again achieved by statistical tests comparing data at different levels of details.

Interestingness criteria for recommendations There is a long discussion about interestingness in the area of evaluating recommender systems [14, 11, 16]. We mention [16] as an excellent recent survey on the topic. The survey presents 4 criteria (diversity, serendipity, novelty, and coverage), in addition to the traditional accuracy, for evaluating the quality of a recommendation.

Query recommendation techniques (see e.g., [6, 2]) are usually evaluated with interestingness measures coming from the literature on recommender systems exposed above. We mention the more OLAP-specific *foresight* measure [2], that quantifies how distant is the recommendation from the current point of exploration.

3 Interestingness aspects for cube exploration

How can we define interestingness? To the best of our knowledge, there is no formal definition. Online resources[3] propose "Interest is a feeling or emotion that causes attention to focus on an object, event, or process". In contemporary psychology of interest, the term is used as a general concept that may encompass other more specific psychological terms, such as *curiosity* [19] and to a much lesser degree *surprise* [20] and *novelty* [8].

[3] https://en.wikipedia.org/wiki/Interest_(emotion)

In this section, we derive from our study of the literature the criteria of the interestingness of a cell, by listing what influences them. We can conclude from our study of related work that interestingness is a degree attributed to a piece of information, regarding the curiosity and surprise it generates. This piece of information under consideration may spark the will to continue exploring the source of information to close some knowledge gap, or get novel information. But how can we pass from such a high level description of interestingness, to a more concrete one? Our approach is a two level modeling. At the first level, we discuss *high-level aspects* of interestingness, like the ones deduced from the study of human behavior. Second, we provide *data-oriented measures* of interestingness, substantiating the aforementioned high-level aspects, on the grounds of the available information. This section presents the first level, while next section provides examples of concrete measures (the second level of our approach) and describes their computation. A proof of concepts implementing some measures is described in Section 5.

3.1 Interestingness aspects

We now present 4 fundamental, high-level interestingness aspects: relevance, novelty, surprise, and peculiarity.

Relevance as a measure for the user's curiosity Curiosity is the main driver of knowledge acquisition. Data exploration, especially in an environment of Business Intelligence, is primarily related to the answering of an open question. So, it is realistic to assume that the user comes with a question for a particular subset of the multidimensional space, and her exploration has to do with "a walk" within this sub-space in order to answer the question. We will call the aspect of interestingness that pertains to curiosity as the *relevance* of the cell with respect to the exploration and its underlying user goal.

The main force, thus, of the assessment of relevance is the modeling of the user intentions. Basically, we can discriminate between (a) the case where a description of the user intention is given vs. (b) the case where no such knowledge is available. In the former, we deal with an expression of the user's interest as the space of a user goal. In the latter, we need to learn the user goal from the history of past activity, which, in turn, relies on the availability of the coordinates of the cells of the queries in the exploration and the schema of the cube.

Novelty Novelty is also an aspect of interestingness that mainly pertains to the need of users to learn information previously unknown. The simple reporting of data that have not been previously reported might increase their interestingness.

The main force behind novelty is the existence of a history. A lesser influence is the availability of results (cell coordinates are sufficient to understand if the cell have never been seen). Without the knowledge of the history of the user's queries, novelty is practically a wild guess. When dealing with novelty, we are not primarily interested in the intention of the user, although it can affect the attention that a user pays to a particular cell (in other words, we assume all cells being equally probable to have been observed by the user).

Surprise Not surprisingly, surprise is a major aspect of interestingness. Surprise occurs when our previous beliefs are disconfirmed or contradicted. This can happen either directly, when the expected value of an event proves to be significantly different than the actual value, or implicitly, when the disconfirmation of a certain fact deduces the disconfirmation of a dependent fact.

Clearly, the main prerequisite for evaluating surprise is the existence of a previous belief of the user. Without the existence of a structured model for the estimation of the previous beliefs, the assessment of surprise is impossible; for this case, it is only possible to measure some objective peculiarity intrinsic to the data (see below). Surprise can be measured using models leveraging the history of the user with the datacube, for instance to estimate belief.

Peculiarity Consistently with the literature on cubes, we use peculiarity to denote an intrinsic property of the data, i.e., the cell's value, when considered together with other cells related to it.

Peculiarity of a cell cannot be assessed in vacuum. Most typically, it can be assessed against the cells of the same query. Taken to extremes, it can also be evaluated by comparing the cell to all the previous cells of the history of the exploration – or even, to all the cells of the full history of the user with the datacube, i.e., including past explorations. Finally, peculiarity may also be calculated with respect to the unseen cells of the cube. The full instance, i.e., with measure values, of cells considered are prerequisites for this criteria.

3.2 Definition of interestingness

We define interestingness of a cell as a vector of scores, defined over a set of interestingness measures.

Definition 1 (Cell interestringness). *Given a user's exploration over a datacube, the interestingness of a cell of this exploration is a tuple of scores for a list of interestingness measures.*

We intentionally do not differentiate between high-level and data-oriented criteria. We support an extensible approach towards which criteria would an interestingness assessment tool include, especially as we cannot provide any completeness proof on our list of high-level interestingness aspects.

4 Detecting interesting cells in an exploration

In this Section, armed with the tools of the previous sections, we revisit the originating question of our introduction: How do we compute the different aspects of the interestingness of a cell? To this end, and without trying to be exhaustive, we provide some alternatives per high-level aspect and discuss their computation.

4.1 Relevance

Assessing the relevance of a cell practically answers the question: *how close is this cell to the subset of the multidimensional space that the user intents to explore*? Two fundamental notions hide behind this formulation of the problem, the specification of an area of interest and the understanding of the user's intention.

As already mentioned, we define the *space of a user goal* as the framing of a subspace of the multidimensional space (either intentionally via selection predicates, or explicitly, at the extensional level, as a set of cells) for which the user wants to obtain information. In the former case, we refer to the *intentional specification of a user goal* whereas in the latter to refer to the extensional *area of interest* of the goal, with the explicit set of cells defined by this framing. Then, given a specific exploration, with a user goal as its underlying motive, we define *relevance* as the degree to which the cell overlaps with the area of interest of the exploration's motivating goal.

Concerning the user intentions, as already mentioned, we discriminate between (a) the case we have no such information, and, (b) the case we have an expression of the user's intentions. Let us proceed in exploring both cases.

Relevance without knowledge of the user's intent Let's start with the case where no model for the user's intent is given a priori. To assess the relevance of a cell, we need to quantify how "close" or "central" the cell is to the subspace induced by the exploration of the user. Practically speaking, we need an algorithm that enumerates the cells that have been visited by the user during her exploration. Due to the hierarchical nature of the space, the easiest way to compare cells is by referring all cells to a common level of granularity (i.e., the node in the lattice of group-by's [13] that is (a) dominated by all the nodes to which history queries correspond, and, (b) the highest among all the candidates of (a)). For simplicity, in this paper, we assume this is the lowest possible node of the group-by lattice, i.e., the level of the facts, that we call C^0.

Now, we need an algorithm that computes the area of interest, starting with its most detailed form, at the level of C^0 (see Algorithm 1). The input to this algorithm is the history of user queries of an exploration. The output is the detailed area of interest. Basically, for every aggregate cell that is part of a query result, the algorithm detects its detailed cells, increases a score for each of the times this cell has contributed to the computation of a query result and adds it to the detailed area of interest, returned by the algorithm.

Having computed the detailed area of interest of a user goal, we can now proceed to answer the question "What is the relevance of a cell c to an area of interest, say S?" Let S be the area of interest of the session, and $\mathbf{S}^0 = \{c_1^S, \ldots, c_k^S\}$ be the set of cells corresponding to the cells of S at the detailed cube C^0. Let $c = \langle a_1, \ldots, a_n,\ v_1, \ldots, v_m \rangle$ be the cell we are interested in and $\mathbf{c}^0 = \{c_1^c, \ldots, c_l^S\}$ be the set of descendant cells corresponding to c at the most detailed level. Then, $relevance(c \mid S)$ is a function f_R that calculates the percentage of $\mathbf{c}^0$ that also lies within $\mathbf{S}^0$ (see Algorithm 2).

Algorithm 1: ComputeDetailedAreaOfInterest

Data:
a history of user queries Q
a basic cube C^0
a set of dimension hierarchies defining the multidimensional space set of models $\mathcal{D}$
Result:
a Detailed Area of Interest S^0, with all its cells annotated with a relevance indicator

begin
  for *every query* $q \in Q$ **do**
    for *every cell* $r \in q.cells$ **do**
      Let $\mathbf{r}^0$ be the set of descendants of r at the most detailed level, $\mathbf{r}^0 \subseteq C^0$;
      for *every detailed cell* $r_i^0 \in \mathbf{r}^0$ **do**
        increase $r_i^0.score$ by 1;
        $S^0 = S^0 \cup r_i^0$;
  return the detailed area of interest S^0

Variants. A more liberal definition of relevance can compute a distance function of the two sets. A more strict definition might take the frequency of the visits of the user to each member of $\mathbf{S}^0$ during the exploration. Then, each cell is weighted by how many times it has been visited by the user during the exploration. Then, *relevance* is defined as the fraction of the sum of the weights of the common cells of the two sets over the sum of weights of the cells of $\mathbf{S}^0$.

A side-effect problem, that we leave aside for the moment concerns the most concise description of $\mathbf{S}^0$ by rolling up regions of C^0 completely covered by cuboids at an ancestor level at the lattice of group-by's.

Relevance in the presence of knowledge of the user's intent Assume now that we have the expression of a user goal. Here, we do not discriminate between an induced goal by a user profile, or a deliberate expression of the goal by the user. We assume that *the goal is expressed as a boolean predicate* ϕ (typically -but not obligatorily- expressed as the conjunction of simple atomic selection formulae). There are several ways to compute the relevance of a cell c to ϕ. Note that ϕ may not be part of the query that retrieves c. The user may (a) compare cells within the area of the original goal with similar / peer cells, or, (b) put the values she observes in context by rolling-up in a way that produces aggregate values broader than the original goal's selection condition.

Variants. The simplest way is to see whether c satisfies the goal ϕ. To do that, both c and ϕ must be converted to the same level of detail – again to their highest common descendant in the lattice of group-by's. Then, *relevance* in its simplest form is Boolean and evaluates to true or false if all descendants of c satisfy ϕ, or numerical, if a percentage is computed. In these variants, the

Algorithm 2: ComputeSimpleRelevance

Data:
a cell c
a history of user queries Q
a basic cube C^0, and a set of dimension hierarchies defining the multidimensional space set of models $\mathcal{D}$
Result:
the relevance of c to Q computed via S^0

```
begin
    S^0 = computeDetailedAreaOfInterest(Q);
    Let c^0 be the set of descendants of c at the most detailed level, c^0 ⊆ C^0 ;
    return relevance(c|Q) = |S^0 ∩ c^0| / |c^0| ;    /* Other variants of the
        formula can be envisaged */
```

history of queries is not taken into consideration – only the intentional space of the user goal.

If we want to assess relevance given the history too, we can resort to the computation of the previous subsection that did not take the user goal into consideration. Assume now that we convert ϕ to the lowest possible level and obtain ϕ^0 [10]. In this case, we can isolate the subset of the explored space that is relevant to the user goal, via $\sigma_{\phi^0}(\mathbf{S}^0)$, with $\mathbf{S}^0$ as previously defined, and then search for its (simple or weighted) intersection with $\mathbf{c}^0$ (also as previously defined).

4.2 Novelty

As already mentioned, novelty refers to the second facet of curiosity, obtaining new knowledge, and for all practical concerns, it deals with whether the user has seen a cell before or not. Due to the hierarchical nature of the multidimensional space, novelty does not only concern the previous appearance of a cell per se, but also, whether the user has been exposed to ancestor or descendant cells too.

Given a datacube C, the history Q of queries of an exploration, and the set H of the cells of the queries of Q, we have several alternatives for the evaluation of novelty, which in all cases is a function f_N assessing $novelty(c \mid H)$ or $novelty(c \mid H, C)$.

1. We define the *strict novelty* of a cell c as its absence from H or not. Thus, the strict novelty is Boolean, and refers to the cell per se, in the context of the exploration's query history.
2. We define the *coverage novelty* of c based on the fraction of cells of the datacube C covered by c (e.g., all the descendants of c) that the user has seen during the exploration: $1 - \frac{|cov(c,H)|}{|cov(c,C)|}$, where $cov(c, S)$ denotes the cells of S covered by c.
3. We define the *inferred novelty* of a cell c as the extent of overlap of c with the cells of H, even via ancestor or descendant relationships. For each cell of H,

say c^H, that is related via an ancestor or descendant relationship with c, we count the complement of the weight of c^H over c. This can be done in many ways, and here we mention the simplest ones. Assume c^H is an ancestor of c, then the respective weight is the fraction of the cell's measures, if the aggregate function is distributive (i.e., not avg). Alternatively, the fraction can be the inverse of the cardinality of c^H's descendants at the level of c. The roles are inverted if the relationship is a descendant rather than an ancestor one. In all these cases, the inferred novelty is a real number that can easily be normalized in the range [0 .. 1]

4. We can also define *inferred novelty at the detailed level* by comparing the detailed descendants of c and the descendants of the members of H, say H^0 at the level of C^0. The percentage of descendants of c at the detailed level that also belong to the H^0 define the inferred novelty of c at the detailed level.

4.3 Surprise

Surprise is a fundamental aspect of interestingness. Where relevance describes the general area of data within which the user wants to walk around, and has to do with *why* he is interested in a cell, surprise relates to the divergence of what she sees with her *previous belief* of what she expected to find. Surprise instigates further searches or actions, in order to adapt our challenged beliefs to the new data, and opens new ways of looking at the data. The fundamental premise upon which surprise can be computed is the modeling of the user's previous beliefs.

How then do we structure a model of beliefs for the cells of a multidimensional space? Fundamentally, there are two ways of handling beliefs: (a) the objective way, where there is a function that assigns an expected value to a measure, independently of what the user has seen in her exploration, and, (b) the subjective way, where the expectation of a cell's value is dependent upon the previous cells that the user has seen in her exploration. The objective evaluation is very demanding, in the sense that it requires that the user has full knowledge of the cube - or even, the sub-cube that she explores and some way to express this knowledge as a potential value. The subjective mechanism is more dynamic: it can start with the user being tabula rasa and, progressively, as cells are observed, her beliefs for the next cells that are related to the previously seen ones are updated.

Surprise assessment. We give two indicative ways to compute surprise, one objective and one subjective.

The *value-based surprise* for a cell c, $surprise(c)$ is the difference between the actual value of a measure M of the cell, say m, compared against its expected value, for instance $\overline{m}$.

The *probability-based surprise* of a cell. Assume a probability distribution P over the set of all potential values for the cells of C. This distribution is used to represent a user's belief, i.e., for a cell $c = \langle a_1, \ldots, a_n, v_1, \ldots, v_m \rangle$ the probability that the user attaches to the statements "the i^{th} measure of c is

v_i". The surprise brought by c is a function over this probability, for instance $surprise(c) = -log(P(c))$.

A fundamental aspect of a model for user beliefs is belief refreshment. As the exploration unravels, the beliefs of the user are updated with every new cell he observes. A mechanism for belief update is out of the scope of this paper, but could follow the general principle given in [3]. However, this does not fundamentally alter the mechanism for interestingness assessment that we propose, as, at any time point, when a cell appears, we can assume that the user has an expected value for it.

4.4 Peculiarity

Peculiarity is an intrinsic property of the data: it makes a particular cell to be set apart from its peers, typically due to the divergence of its measure values from a typical value distribution. Peculiarity can be used to estimate surprise in the absence of any other model for the user (e.g., if we know nothing about what the user expects to see, we can possibly assume that very small or high values in the sales, i.e., outliers, could be interesting). Peculiarity is not restricted to naive outlierness, as it can be due to a more complex pattern (e.g., how a cell evolves over time).

Assessing peculiarity can be performed in a plethora of ways (e.g., via isolating extreme values, assessing how close a value is to its "neighboring" values, performing clustering of the values, information theoretic approaches) [1]. It is beyond the scope of this paper to discuss outlier detection methods, either simple or advanced. We refer the interested reader to [1, 12] for an extensive coverage.

5 Experiments

This section showcases our framework through preliminary experiments over a small set of real user explorations.

5.1 Experimental setup

In our experiments, we reuse the dataset described in [5], consisting of navigation traces collected in the context of a French project on energy vulnerability. Traces consist of logged OLAP sessions[4] of volunteer students of a Master degree in Business Intelligence, answering some high-level information needs defined by their lecturer, using Saiku[5] to ask the queries and see the results. In the present paper, we analyzed 11 sessions, whose sizes range from 12 to 69 queries, 411 queries in total, with an average of 37 queries per session, and an overall of 14,384 cells. Both queries and sessions were manually inspected and labelled by the lecturer. Queries were assigned a binary label regarding their focus on the

[4] We do not distinguish between the terms session and exploration in what follows.

[5] https://www.meteorite.bi/products/saiku

phenomenon analyzed by the student during the session. The term focus is used as in [5]: "When focused, an analyst would expect more precise queries, related to what she is currently analyzing. On the contrary, when exploring the data, the analyst would prefer more diverse queries, for a better data space coverage."

Sessions were graded from A (lowest) to D (highest grade) with respect to the combination of two characteristics, specifically, (a) the extent to which the queries of the session are semantically linked to their previous query (and not ad-hoc) and (b) the progressive stabilization of an area of interest in the multidimensional space (as opposed to everlasting, ad-hoc explorations of the space). Among the 11 sessions analyzed, 4 sessions were labelled B, 3 labelled C and 4 labelled D.

We have developed a prototype *session analyzer* to analyze the logs of the users. Our prototype loads the sessions of each user, and for each of them evaluates the queries one by one, in order. Each time a query is evaluated, the user history is updated, the detailed area of interest (cf. Algorithm 1) is refreshed and the cell interestingness measures are computed. We implemented the extraction of 4 basic measures, one per high level aspect described in the previous section: (i) simple relevance, as of Algorithm 2, (ii) strict binary novelty, i.e., the cell is previously seen or not, (iii) a limited form of surprise, called positional surprise, computed as minus log of the product of the member's probability of appearance in the user history[6], and (iv) simple peculiarity hereafter called outlierness, calculated as z-score w.r.t. the rest of the cells in the query result to which it belongs. Our goal is to confront the measures with the labels assigned to the sessions and queries, looking for correlations between interestingness, user focus, and session quality.

Our prototype is written in Java 8 and ran on a MacBook Pro Core I5 with 16GB RAM running MacOS Mojave 10.14.3. The average processing time per cell is 1071.55 ms, with a minimum of 376 ms, a maximum of 10663 ms and a standard deviation of 248.11. The computation of relevance constitutes by far the majority of the computation time. The average processing time *per query* is 37.18 seconds, with a standard deviation of 85.02. Comparatively, the average consideration time (i.e., the time the user took between two consecutive queries) is 29.42 seconds, with a standard deviation of 65.59.

5.2 Lessons learned

Our first experiment investigates whether the queries with a higher focus obtain higher values for these interestingness measures compared to the queries with less focus.

The first result comes from Table 1. We average all focused vs non-focused cells and compare the values. The focused category consistently demonstrates higher values for all the measures, with novelty having a 15% difference in the

[6] In this implementation, the user belief is agnostic of measure values, and the metric therefore characterizes how surprising it is that the user visits this particular cell.

	Relevance	Novelty	Surprise	Peculiarity
Not focused	0.68 (0.43)	0.56 (0.50)	0.77 (0.25)	0.61 (0.90)
Focused	0.78 (0.31)	0.71 (0.46)	0.82 (0.26)	0.66 (0.78)

Table 1. Average and Standard deviation (in brackets) of measures per query labels

values and relevance a 10%, even though this is nuanced by the standard deviation.

Then, one can refine the above result by assessing whether there is any difference in their behavior of these measures during the progression of the sessions. As session lengths are different, for each query we compute the percentage of progress with respect to the session, as an indicator of how deep the analyst was in her search during that session. To reduce the visual clutter, we organize the demonstration by ranges of 10 steps, where the average value is shown for each category.

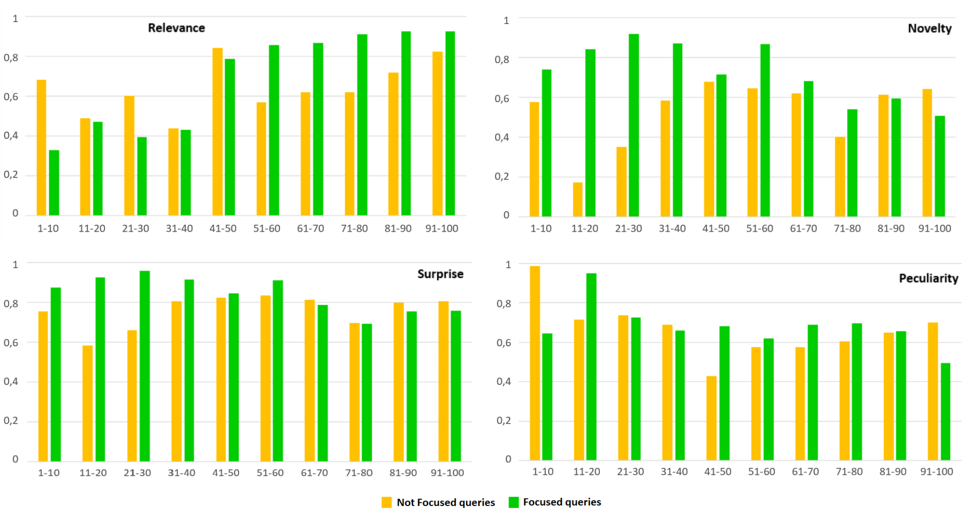

Fig. 1. Evolution of the four interestingness measures (y-axis) with respect to the % progress in a session (x-axis) for focused vs non-focused queries

Figure 1 shows how the four measures evolve along the progression of the sessions, distinguishing by query labels. Concerning novelty, we see that focused queries soon demonstrate higher amounts of novelty compared to non-focused ones (which seem to revolve around the same cells). Only very later in the session is this difference equalized or surpassed (and indeed at low levels of novelty anyway). So overall, focused queries demonstrate more novelty than the non-focused ones. The same phenomenon is observed for surprise, but with less variations. Concerning relevance, as already mentioned, we measure relevance as the subset of the detailed multidimensional space that is revisited, as an indicator of what the user is looking at. Practically, this is acting as the counterpart of novelty, albeit here we are found in the detailed multidimensional space rather than the space of the actual aggregated cells. Here, we observe that the non-focused queries, due to the repetition, obtain higher values than the focused ones. Only later in the session, when the focused queries are returning to the well-established area of exploration to finalize conclusions is the situation reversed. For peculiarity, things are pretty much equal throughout the entire session, apart from a few cases where focused queries contain a little bit more outlier cells than non-focused ones. This justifies the small 5% advantage they have in the total scoring of Table 1.

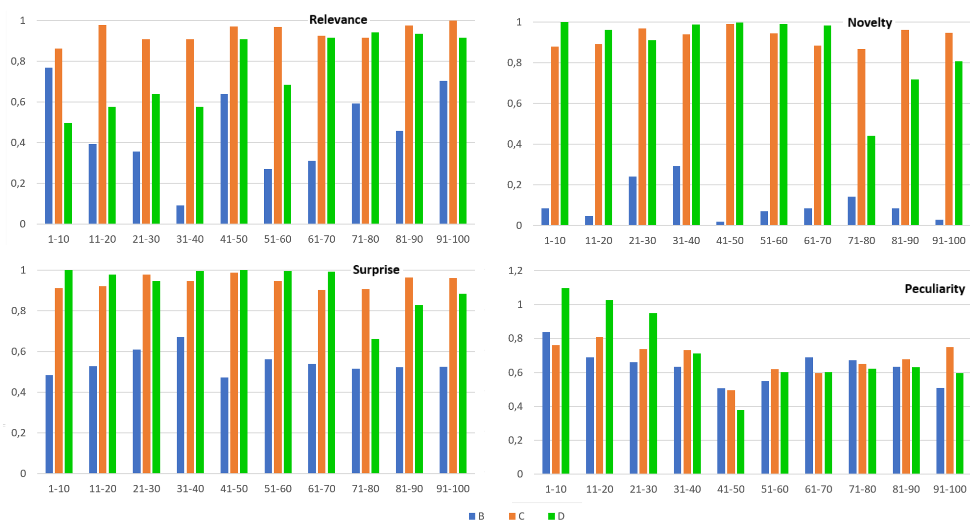

Fig. 2. Evolution of the four interestingness measures (y-axis) with respect to the % progress in a session (x-axis) for session labels

Figure 2 shows how the four measures evolve along the progression of the session arranged by session label. The following general behaviors can be observed:

– B sessions are erratic, and novelty is low, one could say they are not really analyzing, in that users are merely comparing with novel facts.
– In C sessions, all measures are high, there is too much movement, indicating that they are focused, but not enough. The fact that novelty and relevance are high at the same time is not contradictory: users stay in the same detailed area, but keep rolling-up, drilling-down. In other words, they keep investigating, but seem inconclusive, which is corroborated by the fact that

those sessions are often longer than D sessions, that get straight to the point. And also by the fact that outlierness tends to increase in the end.
- In D sessions, relevance keeps increasing, novelty is high then collapses, like surprise, and then start increasing again. This indicates that the sessions are more focused in the end. Outlierness is very high in the beginning, which could have sparked the session.

6 Conclusions

This paper has addressed the problem of measuring the interestingness of the cells of a data cube, analyzed by a user during a session of data exploration. We have assumed a hierarchically-structured multidimensional space and, within this context, we have proposed criteria of interestingness at both a high-level and a data-oriented level.

We have kept our discussion independent from the particular model of OLAP operations that can be applied to the data, or from technological aspects influencing it. We believe that the paper opens the road for a more directed research of interestingness assessment and recommendation algorithms with specific targets among the high-level interesting aspects discussed here. Our experiments provide a proof of concept in this direction, showing how even simple measures can help the analysis of user behavior. Extending the framework beyond the realm of clean, simply structured multidimensional spaces, in the realm of an arbitrarily structured and populated database schema, is a clear path for future work.

References

1. Aggarwal, C.C.: Data Mining - The Textbook. Springer (2015)
2. Aligon, J., Gallinucci, E., Golfarelli, M., Marcel, P., Rizzi, S.: A collaborative filtering approach for recommending OLAP sessions. Decision Support Systems **69**, 20–30 (2015)
3. Bie, T.D.: Subjective interestingness in exploratory data mining. In: Proceedings of IDA. pp. 19–31 (2013)
4. Djedaini, M., Drushku, K., Labroche, N., Marcel, P., Peralta, V., Verdeau, W.: Automatic assessment of interactive OLAP explorations. Inf. Syst. **82**, 148–163 (2019)
5. Djedaini, M., Labroche, N., Marcel, P., Peralta, V.: Detecting user focus in OLAP analyses. In: ADBIS. pp. 105–119 (2017)
6. Eirinaki, M., Abraham, S., Polyzotis, N., Shaikh, N.: QueRIE: Collaborative database exploration. IEEE Trans. Knowl. Data Eng. **26**(7), 1778–1790 (2014)
7. Fabris, C.C., Freitas, A.A.: Incorporating deviation-detection functionality into the OLAP paradigm. In: SBBD. pp. 274–285 (2001)
8. Förster, J., Marguc, J., Gillebaart, M.: Novelty categorization theory. Social and Personality Psychology Compass **4**(9), 736 – 755 (2010)
9. Geng, L., Hamilton, H.J.: Interestingness measures for data mining: A survey. ACM Comput. Surv. **38**(3), 9 (2006)

10. Gkesoulis, D., Vassiliadis, P., Manousis, P.: CineCubes: Aiding data workers gain insights from OLAP queries. Inf. Syst. **53**, 60–86 (2015)
11. Gunawardana, A., Shani, G.: A survey of accuracy evaluation metrics of recommendation tasks. Journal of Machine Learning Research **10**, 2935–2962 (2009)
12. Han, J., Kamber, M., Pei, J.: Data Mining: Concepts and Techniques, 3rd edition. Morgan Kaufmann (2011)
13. Harinarayan, V., Rajaraman, A., Ullman, J.D.: Implementing data cubes efficiently. In: SIGMOD. pp. 205–216 (1996)
14. Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.: Evaluating collaborative filtering recommender systems. ACM Trans. Inf. Syst. **22**(1), 5–53 (2004)
15. Jensen, C.S., Pedersen, T.B., Thomsen, C.: Multidimensional Databases and Data Warehousing. Synthesis Lectures on Data Management, Morgan & Claypool Publishers (2010)
16. Kaminskas, M., Bridge, D.: Diversity, serendipity, novelty, and coverage: A survey and empirical analysis of beyond-accuracy objectives in recommender systems. TiiS **7**(1), 2:1–2:42 (2017)
17. Klemettinen, M., Mannila, H., Toivonen, H.: Interactive exploration of interesting findings in the telecommunication network alarm sequence analyzer (tasa). Information and Software Technology **41**(9), 557 – 567 (1999)
18. Kumar, N., Gangopadhyay, A., Bapna, S., Karabatis, G., Chen, Z.: Measuring interestingness of discovered skewed patterns in data cubes. Decision Support Systems **46**(1), 429 – 439 (2008)
19. Litman, J.: Curiosity and the pleasures of learning: Wanting and liking new information. Cognition and Emotion **19**(6), 793–814 (2005)
20. Reisenzein, R., Meyer, W.U., Niepel, M.: Surprise. In: editor), V.S.R. (ed.) Encyclopedia of Human Behavior. Elsevier, 2nd ed. edn. (2012)
21. Salimi, B., Gehrke, J., Suciu, D.: Bias in OLAP queries: Detection, explanation, and removal. In: SIGMOD. pp. 1021–1035 (2018)
22. Sarawagi, S.: Explaining differences in multidimensional aggregates. In: Proceedings of VLDB. pp. 42–53 (1999)
23. Sarawagi, S.: User-adaptive exploration of multidimensional data. In: Proceedings of VLDB. pp. 307–316 (2000)
24. Sarawagi, S., Agrawal, R., Megiddo, N.: Discovery-driven exploration of OLAP data cubes. In: EDBT. pp. 168–182 (1998)
25. Sathe, G., Sarawagi, S.: Intelligent rollups in multidimensional OLAP data. In: Proceedings of VLDB. pp. 531–540 (2001)
26. Yao, Y., Chen, Y., Yang, X.D.: A measurement-theoretic foundation of rule interestingness evaluation. In: Foundations and Novel Approaches in Data Mining, pp. 41–59 (2006)
27. Zhao, Z., Stefani, L.D., Zgraggen, E., Binnig, C., Upfal, E., Kraska, T.: Controlling false discoveries during interactive data exploration. In: SIGMOD. pp. 527–540 (2017)