The Interestingness of Deviations

Gregory Piatetsky-Shapiro   Christopher J. Matheus

GTE Laboratories Incorporated
40 Sylvan Rd., Waltham MA 02254
gps@gte.com, matheus@gte.com

Abstract

One of the most promising areas in Knowledge Discovery in Databases is the automatic analysis of changes and deviations. Several systems have recently been developed for this task. Success of these systems hinges on their ability to identify a few important and relevant deviations among the multitude of potentially interesting events. In this paper we argue that related 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. We discuss methods for determining the estimated benefit from the impact of the deviations and the success probability of an action. Our analysis is done in the context of the Key Findings Reporter (KEFIR), a system for discovering and explaining "key findings" in large relational databases, currently being applied to the analysis of healthcare information.

Keywords: knowledge discovery, databases, interestingness, deviations, healthcare

1 Introduction

Many companies and organizations maintain large databases to record transactional events such as sales, expenditures, inventory, etc. The timely analysis of key patterns that arise in these databases is highly desirable and may often provide competitive advantage. As the databases grow larger and competition increases, manual methods of analysis become too costly and time consuming to be effective. This problem has led to the development of automated systems for data analysis and report generation, with the most notable examples coming in the area of supermarket scanner data – c.f. Spotlight [Anand and Kahn, 1992] and CoverStory [Schmitz *et al.*, 1990].

KEFIR is a discovery system for data analysis and report generation from relational databases; its application to the problem of healthcare information analysis and reporting is described in [Matheus *et al.*, 1994]. This system embodies a generic approach based on the discovery technique of deviation detection [Matheus *et al.*, 1993] for uncovering "key findings," and dependency networks for explaining the causes of these findings. The results are compiled into a written report along with recommendations for actions to be taken in response to certain types of findings.

Central to KEFIR's methodology is its ability to rank deviations according to some measure of "interestingness." Interestingness refers to the degree to which a discovered pattern is of interest to the user of the system and is driven by factors such as novelty, utility, relevance, and statistical significance (see [Frawley *et al.*, 1991]). An automated discovery system requires specific interestingness factors which it can measure, as well as a way of combining these factors into a metric that accurately reflects how domain experts judge key patterns. This is a difficult problem. The primary purpose of this paper is to discuss our experience from the development of KEFIR and to outline its current approach to measuring the interestingness of discovered patterns.

In KEFIR, the central type of pattern is a *deviation* between an observed value of a measure and a reference value, e.g. a previous or a normative value. We will argue that to properly judge the interestingness of a deviation, one should examine deviations between an observed value and all relevant reference values, that is all previous values, and all relevant normative values. We call such a set of deviations a *finding*. We will also argue that a good measure of the interestingness of a finding is the estimated benefit that could be realized by taking a specific action in response.

In an earlier work [Piatetsky-Shapiro, 1991] we examined various mathematical and statistical factors for interestingness of rules. Here, we will argue that such objective factors are insufficient and that domain-specific, knowledge-based factors also have to be included.

To set the stage for our discussion on interestingness, we start with a concrete example of data analysis of healthcare information. We then outline how KEFIR is designed to perform this type of task in a generic manner. With this background, we will explore how interestingness is measured and used in KEFIR, and offer ideas for future work in this area.

2 Healthcare Example

With the rapid rise in healthcare costs and the recent emphasis on healthcare reform, timely analysis of healthcare information has become an issue of great importance. Large corporations, hospitals, health-maintenance organizations, and insurance companies all require expert analysis of their data – an endeavor that is both time consuming and very expensive. The coming healthcare reform is likely to increase data analysis requirements. All this presents a real opportunity for automating data analysis and reporting systems, especially because the methods currently employed by healthcare analysts lend themselves well to automation.

Current approaches to healthcare data analysis rely on a set of relatively standard *measures* or *indicators* which assess various aspects of healthcare, such as cost, price, usage, and quality (e.g. `average_hospital_payments_per_capita`, `admission_rate_per_1000_people`, `cesarean_section_rate`). These measures are usually aggregate values taken over populations of individuals. For a corporation, the primary population of interest is its employees and their dependents. Various sub-populations of this group are also of interest to the company, such as separate business units, national regions, union vs. non-union employees, etc. From the healthcare side, sub-populations of interest are defined in terms of standard categories, such as Inpatient/Outpatient, Inpatient Admission Type (medical, surgical, etc.), Major Diagnostic Category (MDC), and Diagnostic Related Group (DRG).

A fundamental question in healthcare analysis is: For a given population, how do the standard measures compare to previous values and to normative or expected values? If a measure for the population has changed dramatically or deviates significantly from the norm, then this is a potentially interesting deviation. The actual interestingness depends on whether there are actions that can be taken in response and on the benefits that might result. For example, a $1,000,000 increase in payments due to an increase in the number of regular pregnancies is much less important than a $200,000 increase in payments due to premature deliveries, for which there are well-established intervention strategies that can save a significant part of the cost *and* improve the quality of care. Thus, the interestingness of a deviation is related to the estimated benefit achievable through available actions. The estimated benefit depends on several factors, including the impact on the bottom line, the trend of the deviation, the difference from the norm, and the success probability of the suggested action.

In addition to uncovering the significant findings, the analyst needs to explain them to the extent possible given the data. The standard procedure for explaining a high-level finding is to "drill down" into the data. In this technique, the cause of a finding is traced to either other significant deviations in smaller sub-populations, or to other measures that drive the value of the finding's measure. The healthcare expert performs this drill-down in a top-down fashion, starting with the entire population and drilling down into smaller and smaller populations until no more significant events are found. The key findings and their explanations are then compiled into a summary report along with recommendations for courses of action.

3 The KEFIR System

KEFIR models the analytic process employed by the expert data analyst. The driving premise of the system is that many of the most interesting patterns to be found in transactional databases can be described as *deviations*. A deviation, in our use of the term, is a difference between an *observed value* V_0 and a *reference value* V_R. In our system, the observed value is taken from the most current snapshot of the healthcare database. Comparing the observed value to one from the previous time period generates a deviation over time. A normative deviation results from a comparison to the normative value for the measure. We note that a normative value may be taken from a normative database or it could be computed from a model (see the Appendix for more detail).

Deviations are powerful because they provide a simple way of identifying interesting patterns in the data. We have studied many knowledge-discovery algorithms with potential for identifying vast numbers of significant patterns from data, but most of these are unable to determine when a pattern is truly interesting to the user [Matheus *et al.*, 1993]. With deviations we have a simple way to identify things that differ from our expectations – since they differ from what we expect, they are by definition interesting at least to some degree. Measuring the degree to which they are interesting is the focus of the latter half of this paper.

In addition to detecting and ordering deviations, KEFIR also attempts to provide explanations for the most interesting deviations, and it uses a rule base to generate recommendations for courses of action to respond to specific types of findings (these aspects are described more fully in [Matheus *et al.*, 1994]). The overall design and process flow of the system is depicted in Figure 1.

Deviation Detection: The deviation space that KEFIR explores is completely specified by predefined *measures* and by predefined *categories* used to create subsets of data. We refer to a these subsets as *sectors*, with the "top sector" representing the entire population covered by the data. KEFIR begins its analysis by evaluating the trend and normative deviations of all the measures relevant to the top sector. New sectors are then created for each of the partitions defined by all relevant categories, and deviations are calculated for each measure in each of these new sectors. This drill down into smaller and smaller populations continues recursively until a pre-specified depth is reached or the size of a population becomes inconsequential. The result of this process is several hundred to several thousand deviations compiled into a set of findings.

Evaluation and Ordering: After the findings are calculated, they are ordered in preparation for selecting the *key findings* to include in the final report. This ranking requires a metric for calculating the relative interestingness of a finding. The details of this metric are described in section 4.

Explanation: KEFIR generates explanations for its key findings whenever possible. An explanation for a given finding can come from

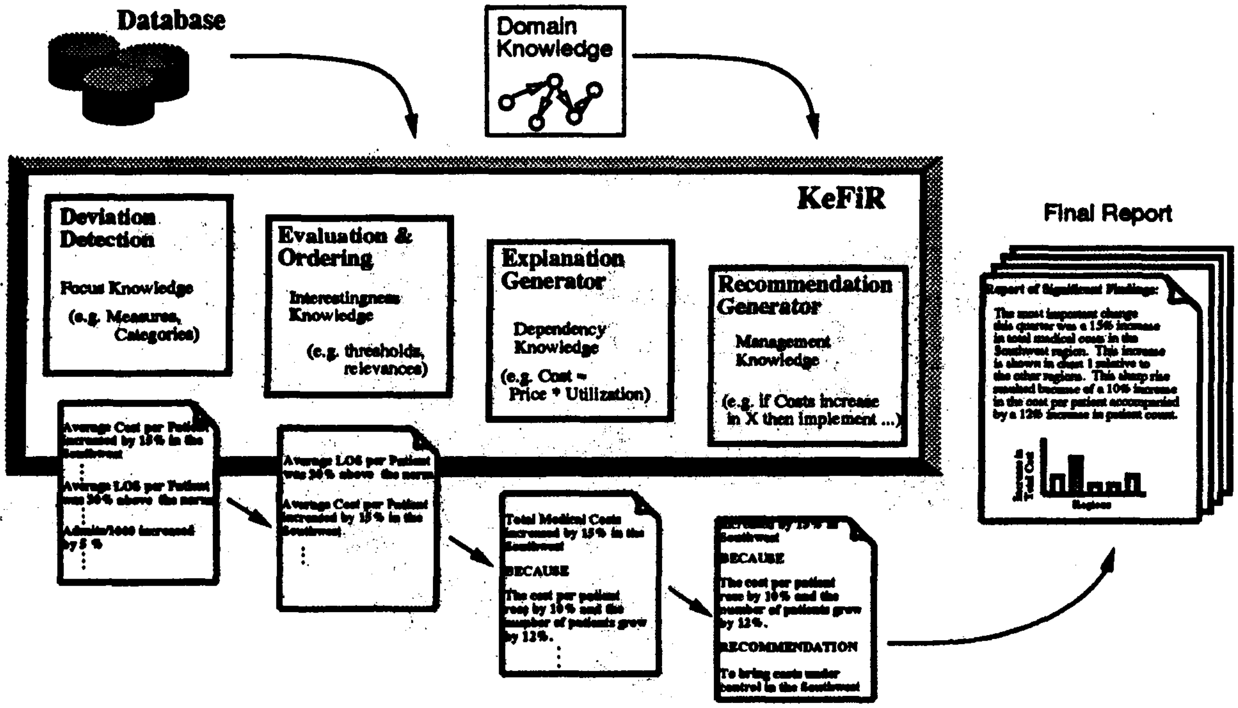


Figure 1: *Overall design and process flow within* KEFIR.

- the decomposition of a measure by a pre-defined formula. For example, the system can use `total_payments = payments_per_case * cases` to infer that the increase in `total_payments` is mainly due to the increase in `payments_per_case`,
- from the breakdown of a sector into sub-sectors. For example, from the relationship
  Admissions = Medical ∪ Surgical ∪ Behavioral ∪ Maternity
  the system might infer that the increase in `total_payments` is mainly due to the increase in payments for Surgical admissions.

KEFIR explains a key finding by first evaluating all other findings affecting it through formulas or breakdowns. It then selects the one finding with the greatest influence and attempts to explain it in the same manner. This recursive process continues until there are no more interesting findings to explain. The final result is a sequence of explanations that chain together a set of interesting findings.

Recommendation: The main purpose for reporting the key findings is to help the user decide what to do to improve the situation. What the user often wants is a set of actions that can be performed in response to the discovered findings. In many cases, the information provided by a finding is sufficient for the system to automatically suggest an appropriate course of action for handling the problem. KEFIR uses a set of rules to identify these situations and to generate recommended actions.

Report Generation: The final output from KEFIR is a written report of the key findings, their explanations, and recommendations. Sentences and paragraphs are generated using simple template matching, with randomized variations to produce more natural sounding text. Descriptive information relevant to the findings also appears in the report in the form of tables, bar charts, and pie charts. The results are produced as a collection of HTML (hypertext markup language) and GIF (graphic interchange file) files for viewing by local or remote WWW (world-wide web) clients such as NCSA's Mosaic.

Implementation: KEFIR is written in tcl [Ousterhout, 1990] and C/C++. The system accesses data through an SQL interface to ensure compatibility with a wide range of database servers. We are currently running the system on a Sparcstation 10 with an Informix DBMS. The healthcare implementation of KEFIR uses the core KEFIR system augmented with healthcare domain knowledge in the form of structures representing the measures, categories, sectors, and recommendations.

4 Analysis of Interestingness

An approximate measure of the right thing
is better than the exact measure of the wrong thing

A critical feature of the KEFIR system is its ability to accurately order findings according to their degree of "interestingness." In this section we consider various aspects of judging interestingness and describe the approaches used in KEFIR, as well as those envisioned for future systems.

Our view of interestingness fits into a statistician's view of an optimal utility function, as defined, e.g. by [DeGroot, 1970]. However, since the potential decisions examined by KEFIR are those of a large company, the KEFIR utility function can be considered to be simply equal to the estimated savings (possibly adjusted by the quality of care multiplier), and various tools of utility theory developed for non-linear subjective utility are not necessary.

In our discussion, we use D to denote database instances and S to denote sectors. A *measure*, denoted $M(S, D)$ (or simply M) is a function that returns a value when applied to a particular sector S and a database instance D. Although measures discussed here are single valued, multi-valued measures are also possible. Not all measures are applicable to all sectors. We also assume, unless noted otherwise, that the desired direction for each measure (from the perspective of benefit to the user) is down, as it is for example in the measure `payments_per_case`.

First, we will examine the interestingness of a single deviation and then show why it is important to combine temporal and normative deviations.

4.1 Impact of a Deviation

The major goal in healthcare information analysis is to identify areas for reducing cost and improving quality. In retail sales analysis, the goal would be to identify areas where sales can be increased. In manufacturing, the goal might be to reduce defective output. The common ground here is identifying deviations which can serve as a basis for useful actions. When dealing with financial data, usefulness can be naturally measured in monetary terms (e.g. potential savings or potential earnings). Other measures, such as quality of care or defect rates, can also be translated to financial terms (although difficult, it is being done on a daily basis by the experts in these areas).

From this perspective, an important aspect of the interestingness of a deviation is its impact on the bottom line. For example, if `payments_per_case` for Surgical admissions in the West Region increased from $14,818 to $23,187 between 1992 and 1993, how can we determine the impact of this change on the bottom line? First, we need to select a measure M_0 that represents the bottom line. This measure should be such that any other measure M_i can be related to M_0 via some function f_i of M_i, i.e. $M_0 = f_i(M_i, D)$. Note that f_i would generally be a function of other measures and also of the database instance. For the healthcare application of KEFIR, M_0 is the total GTE healthcare payments, denoted `total_payments`. The impact should be measured with respect to the overall top-level sector S_0. Usually, S_0 is just the overall population covered by the health plan, i.e. GTE employees and their dependents. However, a regional manager may set S_0 to the population in a specific region.

We are now ready to give a formal definition to our notion of impact.

Definition 1 *The* impact *of the deviation of measure M_i in sector S relative to a reference database D_R and an observation (current) database D_O, denoted $impact(M_i, S, D_O, D_R \| M_0, S_0)$, is the difference between*

- *the value M_0 would have if only the value of M_i for sector S was changed to its observation value $M_i(S, D_O)$, while all other values would be as in D_R, and*
- *$M_0(S_0, D_R)$, the reference value of M_0 in sector S_0.*

The formal equation for impact is thus:

$$impact(M_i, S, D_O, D_R \| M_0, S_0) = f_i(M_i(S, D_O), D_R) - M_0(S_0, D_R)$$

When the values of M_0 and S_0 are obvious they will be omitted.

To give a specific example, let D_{93} be D_O, the current database, D_{92} be D_R, the reference database, S_{surg} be the surgical admissions sector, and S_{other} be the remaining sector where $S_0 = S_{surg} \cup S_{other}$. Recall that `total_payments = payments_per_case * cases`. The function $f(\texttt{payments\_per\_case}, D)$ then becomes

$$\begin{aligned}\texttt{total\_payments}(S_0, D) &= \texttt{payments\_per\_case}(S_{surg}, D) \times \texttt{cases}(S_{surg}, D) + \\ &\quad \texttt{total\_payments}(S_{other}, D)\end{aligned}$$

We can calculate the impact from the change in `payments_per_case` as follows:

$$\begin{aligned}&impact(\texttt{payments\_per\_case}, S_{surg}, D_{93}, D_{92}) \\ &= \texttt{payments\_per\_case}(S_{surg}, D_{93}) \times \texttt{cases}(S_{surg}, D_{92}) + \texttt{total\_payments}(S_{other}, D_{92}) - \\ &\quad \texttt{total\_payments}(S_0, D_{92}) \\ &= \texttt{payments\_per\_case}(S_{surg}, D_{93}) \times \texttt{cases}(S_{surg}, D_{92}) + \texttt{total\_payments}(S_{other}, D_{92}) - \\ &\quad (\texttt{payments\_per\_case}(S_{surg}, D_{92}) \times \texttt{cases}(S_{surg}, D_{92}) + \texttt{total\_payments}(S_{other}, D_{92})) \\ &= (\texttt{payments\_per\_case}(S_{surg}, D_{93}) - \texttt{payments\_per\_case}(S_{surg}, D_{92})) \times \texttt{cases}(S_{surg}, D_{92})\end{aligned}$$

Substituting the values

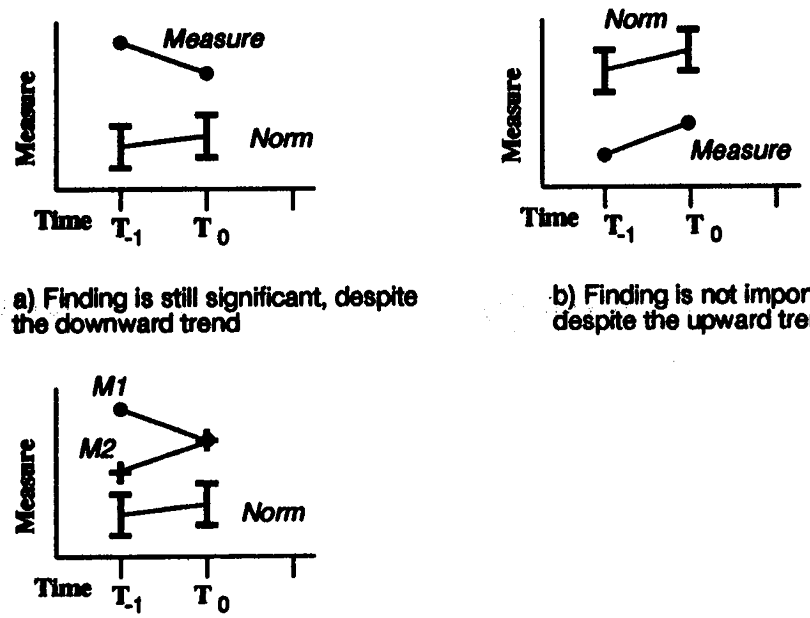
$$\begin{aligned}\texttt{payments\_per\_case}(S_{surg}, D_{93}) &= \$23,187 \\ \texttt{payments\_per\_case}(S_{surg}, D_{92}) &= \$14,818 \\ \texttt{cases}(S_{surg}, D_{92}) &= 149\end{aligned}$$

gives us an impact value of $1,246,981. This means that the $8,369 change in `payments_per_case` in the surgical sector resulted in an increase of $1,246,981 in overall `total_payments`. Equations for computation of the impact in a general case are presented in the Appendix.

In this example the impact represents the potential savings that would be realized if the current value of the measure was brought back to its previous value, i.e. GTE could save $1,246,981 if the `payments_per_case` measure for surgical admissions was brought back to its 1992 level. This approach has two major problems: (1) the old value may be an unrealistic target, and (2) the impact does not indicate the degree of control or *discretion* we might have in changing the measure. These two problems are discussed in the following sections.

4.2 Trend and Normative Deviations

Bringing a measure to its old value may be quite difficult or even impossible. A more realistic target for the measure is its expected normative value. In healthcare, as in other fields, there are tables of norms (computed by medical experts) for many key measures, representing their average or desired values. The appendix describes how these norms can be used to derived expected normative values.



a) Finding is still significant, despite the downward trend

b) Finding is not important, despite the upward trend

c) Although both measure M1 and measure M2 have the same deviation from the norm at T0, their different trends give them different importance.

Figure 2: Different examples of trend and norm deviations

Figure 2 shows two examples of how focusing only on changes can be misleading. In Figure 2a, despite the downward trend in the measure, the finding is significant because the measure's value remains above the norm. In Figure 2b, despite the upward trend in the measure, the finding is not very significant because the measure's value is still below the norm.

This example suggests that perhaps the important issue is simply how a measure compares to its normative value. While this is a better approximation, it is also insufficient, as illustrated by figure 2c. While both measures have equal deviation from the norm at time T_0, the trend suggests that M_2 will have a greater deviation in the future, if no action is taken.

Thus, we see that the normative impact at the present period reflects the "missed savings" and is only an approximation for the real measure of benefit, which is the "potential savings" achievable in the future.

4.2.1 Future Potential Savings

Figure 3 shows two examples of estimating the future potential savings. We need to forecast $M(S, D(T_1))$ (i.e. the expected value of measure M in sector S at the next time period T_1, assuming that no action is taken) and $M(S, Norm(T_1))$ (i.e. the expected normative value of measure M in sector S at T_1).

Figure 3a shows a simple case when only the present and previous values are available. Using

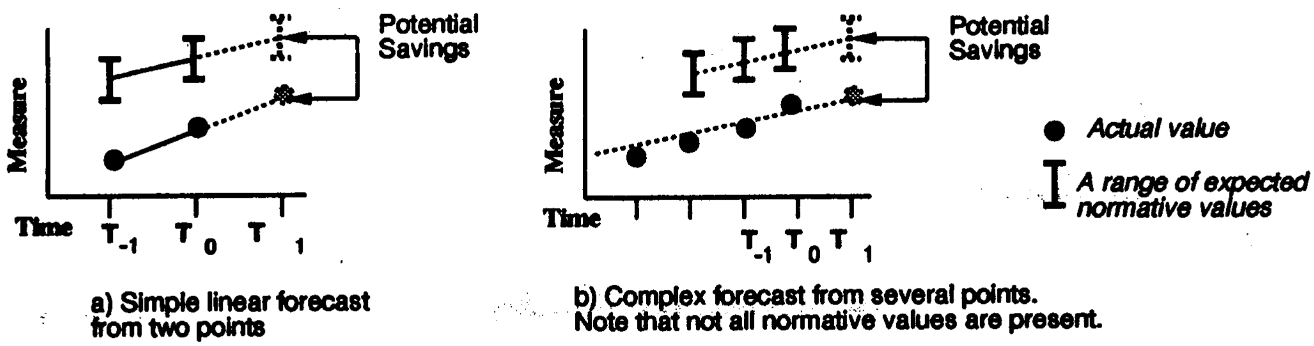


Figure 3: Potential Savings.

a linear trend model, we can forecast

$$M(S, D(T_1)) = 2 \times M(S, D(T_0)) - M(S, D(T_{-1}))$$

$$M(S, Norm(T_1)) = 2 \times M(S, Norm(T_0)) - M(S, Norm(T_{-1}))$$

If more data points are available, as in figure 3b then more complex forecasting strategies are possible, such as fitting the best line or using a rolling average. A further refinement is knowledge-based forecasting [Lee *et al.*, 1990], which can also consider other factors, such as medical inflation trends for price measures, downsizing trends for the number of employees, etc.

Given that we can adequately estimate values for time T_1, we can define the potential savings from the deviation on measure M in sector S as the impact of the difference between $M(S, D(T_1))$ and $M(S, Norm(T_1))$:

$$\text{Potential-Savings}(M, S) = \textit{impact}(M, S, D(T_1), Norm(T_1) \| M_0, S_0)$$

4.3 Discretion

Potential savings is not of much value if it cannot be realized. The degree to which a user has control over obtaining the potential savings is called *discretion*. When a user has total discretionary control, the entire potential savings represented by the impact can be achieved – we then say that the user has 100% discretionary control. More often, a user will have only partial control over the value of a measure in a given sector. In these cases, only a fraction of the potential savings can be expected. A natural way to represent discretion is as a weighting factor between 0 and 1 representing the likelihood of achieving the potential savings.

For each finding we need to derive a discretionary weight. This weight could be associated with the measure, with the sector, an action, or some function of these. For example, a discretionary weight assigned to a particular measure would indicate the relative control the user has over that measure's value independent of the sector or the specific action. The problem with this approach is that the discretion over a measure is usually highly dependent upon the sector in which it occurs. In addition, the likelihood of realizing the potential savings is conditional on there being an action available. It therefore makes more sense to associate discretion with actions.

From a healthcare perspective, the availability of intervening actions makes a finding much more interesting to the manager because these represent real opportunities for savings. For example, managers have no actions for affecting the number of regular pregnancies, but there are several accepted actions for reducing the number of premature deliveries or for improving quality of care

```
if measure=payments_per_case &
   sector=surgical_admission &
   measure value increased by more than 10%,
then recommend:
 A study is suggested for discretionary and high-cost surgery.
success probability: 0.4
```

Figure 4: A sample rule, translated into English.

for chronic illnesses. Consequently, findings that relate to premature deliveries are much more interesting than findings relating to normal deliveries.

Our healthcare domain expert has provided a number of recommended actions for various measures in different sectors. These are encoded in the system as production rules. For each rule the expert estimated the probability of success – how likely the action is to bring the measure back to the norm. An example of one of these rules is given in Figure 4.

For a given finding, the system identifies matching rules and selects the rule with the highest probability of success $p_{success}$. It then computes the estimated (as opposed to potential) benefit using the following formula:

$$\text{Estimated-Benefit(finding)} = \text{Potential-Savings(finding)} \times p_{success}$$

Because the healthcare field is rapidly progressing, the set of available actions will be constantly changing. To adjust to that, and to explicitly account for incompleteness of system's knowledge, a default action of simply reporting the deviation matches any finding; this can be viewed as an encoding for the likelihood that bringing the deviation to the user's attention will lead to some (unknown to the system) corrective action. The success probability of the default action varies for different sectors, but is generally low.

4.4 Statistical Significance

Let us further consider the example of deviations in Surgical **payments_per_case**. The significance of this deviation would be less if the million-dollar-plus increase were attributable to a single extreme case than if it were due to several dozen high-cost cases. The rationale for this reasoning is that a single extreme case is unlikely to re-occur next year, and so there is nothing to be done; several dozen high-cost cases, however, indicates a potentially correctable pattern. Formally capturing this intuition requires analysis of statistical significance.

Estimating the potential benefit of an action as a single number (e.g. estimated benefit = $567,432) has the added problem of giving a false sense of precision. Forecasting is intrinsically an imprecise science and it would be much better to give a range and a confidence (e.g. estimated benefit is between $400,000 and $700,000 with confidence 0.9), or even a central estimate and a standard deviation.

Computing the confidence or a standard deviation requires either knowing the apriori data distribution (impossible in our application and in most real cases), or having a large set of historical data points. In our application, we have huge amounts of data, but they only go back one or two years, and thus we cannot make a reliable annual forecast based only on this data. The lack of historical data and consequent lack of standard statistical measures is, unfortunately, typical for many areas of medical cost analysis today. In the meantime, we are solving the problem by using simple approaches such as disregarding findings based on less than a minimum number of cases, and using heuristic rules for dealing with extreme deviations based on a small number of cases. Better methods of producing statistically reliable estimates, given very incomplete data, are the topic of further research.

5 Conclusions

Several systems have recently been developed for analyzing changes and deviations in large relational databases. Success of these systems hinges on their ability to identify a few important and relevant deviations among the multitude of potentially interesting events. In this paper we argued that interestingness should be based on the estimated benefit from possible actions taken in response to observed deviations. We presented an approach used in KEFIR for doing this based on the notions of impact and discretion. Although this approach makes several simplifying assumptions, the results of the system in practise have shown the merit of the method and encourage further research in this area.

Acknowledgments: Dwight McNeill was a tremendous help and inspiration in clarifying our views on interestingness. We thank the referees for their suggestion. We are very grateful to Shri Goyal and Bill Griffin for their encouragement of our work on discovery in databases.

References

[Anand and Kahn, 1992] T. Anand and G. Kahn. SPOTLIGHT: A data explanation system. In *Proc. Eighth IEEE Conf. Appl. AI*, 1992.

[DeGroot, 1970] M. H. DeGroot. *Optimal Statistical Decisions*. McGraw-Hill, New York, 1970.

[Frawley *et al.*, 1991] William J. Frawley, Gregory Piatetsky-Shapiro, and Christopher J. Matheus. Knowledge discovery in databases: An overview. In *Knowledge Discovery in Databases*, pages 1–27. AAAI/MIT Press, Cambridge, MA, 1991. Reprinted in *AI Magazine, Vol. 13, No. 3, 1992.*

[Lee *et al.*, 1990] J. K. Lee, S. B. Oh, and J. C. Shin. UNIK-FCST: Knowledge-assisted adjustment of statistical forecasts. *Expert Systems with Applications*, 1(1):39–49, 1990.

[Matheus *et al.*, 1993] Christopher J. Matheus, Philip K. Chan, and Gregory Piatetsky-Shapiro. Systems for knowledge discovery in databases. *IEEE Transactions on Knowledge and Data Engineering*, 5(6), 1993.

[Matheus *et al.*, 1994] Christopher J. Matheus, Gregory Piatetsky-Shapiro, and Dwight McNeill. An application of kefir to the analysis of healthcare information. In *Proceedings of AAAI-94 KDD workshop*, July 1994.

[Ousterhout, 1990] John K. Ousterhout. TCL: An embeddable command language. In *Proceedings of the 1990 Winter USENIX Conference*, pages 133–146, 1990.

[Piatetsky-Shapiro, 1991] Gregory Piatetsky-Shapiro. Discovery, analysis, and presentation of strong rules. In *Knowledge Discovery in Databases*, pages 1–27. AAAI/MIT Press, Cambridge, MA, 1991.

[Schmitz *et al.*, 1990] J. Schmitz, G. Armstrong, and J. D. C. Little. CoverStory – automated news finding in marketing. In *DSS Transactions*, pages 46–54, Providence, RI., 1990. Institute of Management Sciences.

Appendix: Calculating Impacts and Expected Values

Some of the aspects of KEFIR's method for determining interestingness may not be immediately obvious. Here we provide implementation details for calculation of impact of a deviation and calculation of the expected normative value for a measure.

Calculating Impact

The bottom line measure M_0 is chosen so that all other measures can be related to it via formulas. Thus, for any measure M_i there is a function f_i such that $M_0 = f_i(M_i, D)$ and the impact of change in M_i is computed simply as

$$impact(M_i, S, D_0, D_R) = f_i(M_i(S, D_0), D_R) - M_0(S_0, D_R) \tag{1}$$

First, let's examine the computation of impact of change in the bottom line measure M_0 for the different sectors. If $S_0 = S_1 \cup S_2 \cup \ldots \cup S_k$, then we can write the old value of $M_0(S_0)$ as

$$M_0(S_0, D_R) = M_0(S_1, D_R) + \ldots + M_0(S_i, D_R) + \ldots + M_0(S_k, D_R) \tag{2}$$

and the value M_0^* that M_0 would have if M_0 would change only in S_i but not in other sectors, is

$$M_0^* = M_0(S_1, D_R) + \ldots + M_0(S_i, D_0) + \ldots + M_0(S_k, D_R) \tag{3}$$

Subtracting these equations we get

$$impact(M_0, S_i, D_0, D_R) = M_0(S_i, D_0) - M_0(S_i, D_R) \tag{4}$$

i.e. the bottom-line impact of M_0 change in S_i is simply the difference between the new and the old values of M_0 in S_i. For example, if the total payments for surgical admissions changed from $2.2 million in 1992 to $3.2 million in 1993, the impact on the bottom line is simply $1 million.

Next, assume we are given a specific sector S and let us examine how to compute the impact of change of a specific measure M_i in just that sector.

Measures are related to other measures by different formulas. In the important special case when these formulas have only additions and multiplications (which is the case for almost all Health KEFIR measures), the function f_i that expresses measure M_0 via M_i can be written as

$$M_0(S_0, D) = A(S, D) M_i(S, D) + B(S, D) \tag{5}$$

where $A(S, D)$ and $B(S, D)$ depend on the sector, the database instance, and other measures, but not on M_i. So, the reference bottom-line value of M_0 is

$$M_0(S_0, D_R) = A(S, D_R) \times M_i(S, D_R) + B(S, D_R) \tag{6}$$

and the value M_0 would have if only M_i would change in D_R would be

$$M_0(S, D_R[M_i = M_i(D_0)]) = A(S, D_R) \times M_i(S, D_0) + B(S, D_R) \tag{7}$$

Subtracting, we get

$$impact(M_i, S, D_0, D_R) = A(S, D_R) \times (M_i(S, D_0) - M_i(S, D_R)) \tag{8}$$

Note that $B(S, D)$ – the contribution resulting from additive terms – drops away completely. This equation allows to compute impact for measures related to M_0 by additions and multiplications simply by keeping track of the multiplicative factor $A(S, D_R)$.

Expected Value of a Measure

Computation and analysis of normative values is a science in itself, practiced by the large number of medical consultants. For healthcare, norms are available for most important measures. Norms vary by region, age, sex, and the DRG (Diagnostic Related Group or, in plain english, disease type), e.g. **payment_per_case** in Northeast USA for DRG=21 is $4,879 while in Southwest USA for DRG=75 it is $25,210.

The normative tables give the average (or best practice) value expected for the typical population. If the examined population differs from the typical one, the direct comparison of the measure value with the norm value may be misleading. Rather, the normative tables should be used to compute the expected value for specified measures, given the particular population.

Healthcare measures can be divided into several broad classes, including *cost* measures (e.g. **payments_per_capita**), *price* measures (e.g. **payments_per_case**), and *use* measures (e.g. **cases_per_1000_people**). While many different criteria affect each type of measure, several simple causal models have been developed in healthcare. In particular, the *use* measures correlate most strongly with age and sex of a person, while the *price* measures correlate most strongly with the DRG distribution of cases.

To compare the value of a *use* measure like **cases_per_1000** for the West Region with the expected normative value, we need to compute the break down of the West Region population into Age/Sex groups (table 1).

Norms		West Region Frequency	
Age Group	Cases per 1000	Age Group	Frequency
M, 0-17	56.5	M, 0-17	9.3%
F, 0-17	53.5	F, 0-17	8.3%
M, 18-34	34.4	M, 18-34	16.5%
F, 18-34	129.8	F, 18-34	24.5%
....		...	
M, 65+	179.9	M, 65+	3.5%
F, 65+	121.2	F, 65+	2.3%

Table 1: Norms table and the population distribution for the West Region.

So, if ASG is the Age/Sex Group, then the expected value of **cases_per_1000** will be

$$\sum_{ASG} Norm_{ASG} \times Frequency_{ASG}$$

Such computation may reveal, for example, that the higher than average value of **cases_per_1000** in the West Region may be due to an unusually large proportion of the **F, 18-34** group, which happens to have a higher than average norm for **cases_per_1000**.

Expected values for other measures are computed in a similar way.