A decision analysis approach to financial risk management in strategic outsourcing contracts

Bonnie K. Ray · Shu Tao · Anatoli Olkhovets · Dharmashankar Subramanian

Abstract Managing risks and the associated decisions surrounding them is an activity often undertaken by organizations within an enterprise on an ad hoc basis, using document templates or spreadsheets to track issues or concerns and associated actions. Such an approach, while worthwhile, does not provide a consistent foundation for data collection, analysis, or reporting to enable data-driven decision making. We describe a framework consisting of (1) data specification and collection, (2) predictive modeling, and (3) reporting for managing service provider risks associated with the execution of strategic outsourcing (SO) contracts and show how such a framework has been leveraged at IBM for managing financial risks associated with IBM's SO business. In particular, we provide details on the construction of quantitative models for proactive risk management and discuss the decision process driven by the model results.

Keywords Business decision making · k-nearest neighbors · Predictive modeling · Risk analysis

Introduction

A large part of IBM's services business is associated with management and delivery of enterprise information technology (IT) services for clients, otherwise known as strategic outsourcing (SO) engagements. Delivery for client environments is complex; service delivery teams typically support a large variety of requests (e.g.,

B. K. Ray (✉) · S. Tao · D. Subramanian
IBM T.J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598, USA
e-mail: bonnier@us.ibm.com

A. Olkhovets
IBM Corporate Headquarters, Armonk, NY, USA

problem tickets, change requests, maintenance requests), with contractual service level agreements associated with each request specifying target response times. To complicate matters, the details of the client environment, such as the number of servers, operating systems, and applications, may not be entirely known when the contractual agreement is finalized between IBM and the client. These issues drive financial risk to IBM associated with delivering the agreed upon services at the price determined at the start of the project. While technical and business risk assessments (a mix of qualitative and quantitative questions) are carried out prior to signing a contract, as well as periodically throughout the life of the contract, these risk assessments are not tightly linked to the impact the identified risks may ultimately have on the financial success of the engagement.

To address this issue, we have developed a risk and decision analysis framework for managing financial risk over the life cycle of an IBM SO contract, along with a software tool that embodies key elements of the framework. The framework consists of (a) a data layer to define and capture in a standardized way, attributes of each contract that may be important for identifying patterns of performance, categorized risk factors (root causes) contributing to ongoing contract financial performance, actions taken in response to predicted and/or observed risks, as well as other associated information (possibly unstructured text) about a contract; (b) an analytic layer consisting of quantitative models to predict the risks a contract may experience and their associated impacts on contract profitability, and (c) a reporting layer to communicate analytical results to the project executives and overall services business leaders. These reports provide views of an individual engagement as well as a portfolio of contracts, and drive decisions regarding risk mitigation actions for ongoing deals and financial feasibility assessments prior to the start of a new engagement. In this paper, we focus primarily on the analytics layer of the framework, providing details on the predictive modeling approach used, how the approach has been validated, and how results of the approach drive business insight and decision making.

While some literature exists regarding managing risk in outsourcing contracts, much of it focuses on risks from the outsourcing client perspective, not necessarily from the viewpoint of the outsourcing provider. See, for example, McIvor et al. (2009), which provides a framework for developing important outsourcing performance management considerations, such as critical success factor (CSF) methodologies, internal performance analysis prior to outsourcing, cost analysis, benchmarking, and performance measurement and management throughout the outsourcing relationship. From a provider perspective, Mojsilovic et al. (2007) use predictive models to estimate the likelihood of revenue erosion in a large outsourcing engagement as a function of contract attributes that change over time, while Goo et al. (2007) investigate factors that influence the duration of IT outsourcing relationships. Aunde and Mathew (2009) analyze offshore IT outsourcing risks from the perspective of service providers and find that relationship maturity, nature of contract, nature of service or project, and nature of client influence the degree of risk. They do not estimate the likelihood and impact of a broad set of risk factors on contract financial performance.

Alternatively, we may think of managing risk in SO contracts as analogous to risk management in large, complex, multi-year projects. Much previous literature on project risk management exists. See, for example, Loosemore et al. (2006). However, much of it focuses on estimating risks associated with schedules, costs, or resources. Works that look at estimating risks associated with financial performance of a project typically rely on direct linkage of a project's attributes to financial outcomes, or prediction of future performance from current financial performance in the case of ongoing projects (Labbi 2005; Ratakonda et al. 2010). Other work, such as that of Deleris et al. (2007), focuses on updating risk likelihoods as information changes over a project's life cycle. Our work is different in that we predict likely risks for a contract using a nearest neighbors approach, i.e., we find examples of historical contracts that best match a contract of interest along a set of contract attributes. The risks observed in the matching contracts and their associated financial impacts are then used to estimate the risk for the given contract. Leung et al. (1998) describe a rule-based system for identifying potential project risks in engineering projects, which has similarities with our work from a data capture perspective but not from a modeling point of view. While the analytical techniques used in our approach are not necessarily new, the work is novel in that it represents the first instance of a risk and decision support system driven by advanced analytics to manage SO contract performance.

The rest of the paper is organized as follows. In "SO financial risk and decision management framework", we provide additional details on the components of the risk and decision management framework briefly described above, including discussion on the data and analysis methods used in the SO Financial Risk Analytics solution. "Application of approach to managing strategic outsourcing financial risk: results and business impact" presents the results of applying the new risk management-driven decision process within IBM's SO business to date, including the challenges encountered and the business impact. "Conclusion and directions for future work" summarizes and discusses future research directions.

SO financial risk and decision management framework

IBM is engaged in multiple SO deals with clients simultaneously, each in different stages of deployment and with potential risks that could impact their profitability. To reduce these risks, decisions must be made regarding actions to mitigate these risks. There is a need to track and manage the performance of both individual contracts and the entire portfolio of contracts over time. The portfolio under management may span the organization and consist of contracts of varying strategic intents and operational complexity. Financial targets (or plans) are typically pre-established at both the contract and portfolio levels, with business success defined and measured by attainment of targets for both. For instance, revenue and cost represent commonly used financial targets. Customer satisfaction may also represent a relevant target for a services contract. However, we focus on financial targets in this paper. No matter the specifics of the target metrics, the challenge is to decide

how to optimally balance resource investment across the entire portfolio of current and potential contracts to ensure that the targets are achieved.

Many project management tools exist, e.g., Microsoft Project (2010), but these are focused primarily on project planning and tracking of project deliverables, which, while related to project risk management, do not enable explicit prediction, prioritization, and management of financial risks. More often, tracking and management of risks in large project initiatives is carried out using spreadsheet or presentation templates that are passed around among the team, with little upfront investment in common data definitions, formats, or structured data collection systems. While this type of management process supports ongoing discussions centered on current issues, it does not enable the business to clearly identify patterns of risks arising for subsets of the initiatives or to easily retrieve and structure information that might be useful for anticipating risks and making decisions as to how to mitigate them. It also does not support quantification of the impact of different risks on performance targets. It is well known that the prediction of risk events by experts tends to exhibit multiple types of bias (Tversky and Kahnaman 1974), such as anchoring bias or recency bias, in which the likelihood of future risk event occurrence is predicted to be greater for those events that are under discussion and have occurred most recently in the past. Other factors also adversely influence decision making, such as the amount of time, money, and effort already spent on a project. Juliusson et al. (2005) find that people tend to continue to make risky decisions when they feel responsible for these so-called sunk costs. Systematic collection and analysis of data pertaining to contract performance, including actions taken to control ongoing performance, is critical to enable more quantitative, fact-based, and pro-active management of outsourcing contracts.

IBM's SO financial risk management framework was designed to orient the relevant business processes toward a more fact-based and analytics-driven approach. As outlined in "Introduction", the foundational elements of this fact-based approach are (1) data specification and collection, (2) risk analytics, and (3) reporting. Data specification consists of creating a structured taxonomy for classification of factors that impact contract performance, along with a set of high-level characteristics (or descriptors) of a contract that are known prior to contract start and are potentially useful for predicting patterns of risk over a contract's life cycle. The impact of each risk factor on contract performance must also be captured. The collected data are used to construct analytic models for predicting the likelihood of risk occurrence and its associated impact. Analysis results, such as prioritized risks based on likelihood and/or impact, are provided to business stakeholders through standard reporting mechanisms to drive risk mitigation actions that can reduce the likelihood and/or impact of a predicted risk. Taken together, these steps provide a foundation upon which predictive and pro-active risk management activities can be built. We provide additional detail on each of these elements for the SO financial risk analysis framework in the following subsections.

Data specification and collection

A well-defined taxonomy of contract risk factors is foundational to data collection. A taxonomy allows discrete events affecting performance to be conceptualized,

classified, and compared across contracts and over time. See, for example, Chapman (2011), for a discussion of risk taxonomies for enterprise risk management. Additionally, a set of high-level characteristics (or descriptors) of a contract that are known prior to the contract start and are potentially useful for predicting patterns of risk over a contract's life cycle is required. Finally, predicting the impact a risk factor is likely to have on a contract's financial performance (in quantitative terms) requires capturing information on how total deviation from a financial target is attributable to specific risk factors. While it is fairly straightforward to collect values of key attributes associated with a particular contract, this latter step requires a bit more care. We provide more discussion on it later in this section.

Periodically throughout the life of the SO contract (typically every 6–12 months), thorough reviews are conducted to assess the health of the contract, resulting in documented project management reviews (PMRs). These PMRs, conducted by domain experts, assess various aspects of contract execution, including financial performance relative to targets, deliverables, customer relationship, etc. An important outcome of this evaluation is the determination of whether a contract is in a "healthy" or "troubled" state. This status is largely driven by the financial performance of the contract, i.e., a contract is deemed healthy if it has met the financial plans set before contract signing. If a contract is considered troubled, additional investigations into contract issues are conducted by contract management experts, resulting in a list of root causes identified as primary contributors to the contract's unsatisfactory financial performance. A set of common causes of unhealthy contracts has been developed by contract management experts over a number of years, based on domain knowledge and defined best practices as to how to diagnose the root causes of financial underperformance. The descriptions of root causes follow a standardized structure and language, which form the basis of our risk taxonomy, i.e., a common language used to identify performance issues encountered in the SO engagement. Through examination of historical PMRs and discussions with experts, we synthesized and reconciled the root cause information to form a comprehensive taxonomy that is applicable to data capture across multiple types of SO contracts. While some of these root cause definitions may lack clarity and/or measurability, they have the advantage that they are part of IBM's quality assurance (QA) tool today, and were defined by field practitioners. The fact that the terminology is familiar to the SO contract leaders makes it easier to obtain their buy-in of the corresponding analysis. Using existing terminology also makes integration of the analytics into the QA tool easier.

For our work, a root cause was defined and included in our taxonomy only if the corresponding issue had been experienced in the context of a historical contract. Additional potential issues never yet experienced could also be included based on input from subject matter experts, but would not be predictable from a data-driven analysis. An initial taxonomy can continue to be refined over time to reflect new and changing categories of risk factors, as long as the historical data set of observations is mapped onto the updated taxonomy. Our performance factor taxonomy has been constructed such that it can easily be expanded to include new factors over time.

An illustrative example of this taxonomy is shown in Fig. 1. As can be seen from the figure, the taxonomy follows a tree structure. Each leaf node represents a root
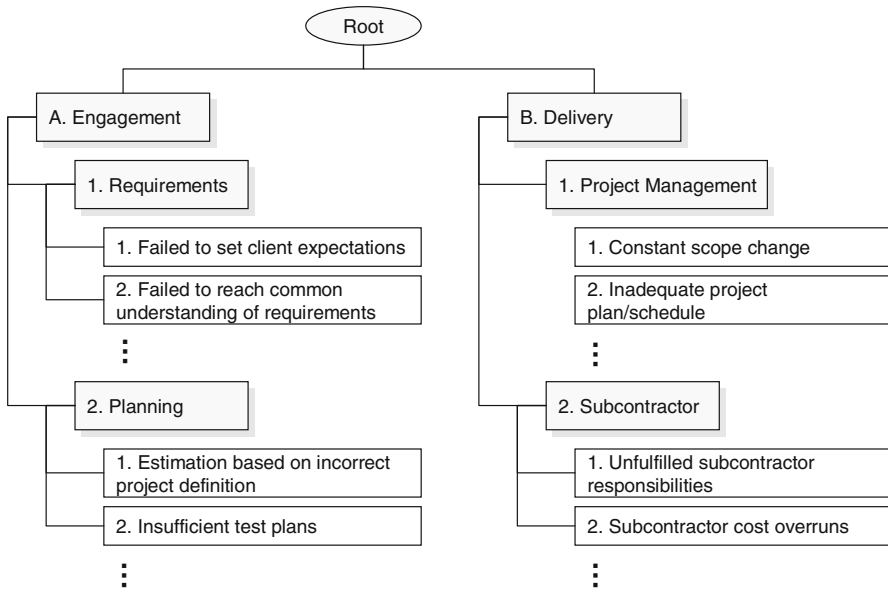
Fig. 1 Example of risk taxonomy derived from the root causes identified in project management reviews (PMRs)

cause, with a unique reference path from the root. For instance, risk or root cause *A.1.1*, "Failed to set client expectations", is part of subcategory *A.1*, Requirements, indicating that client expectations were not set appropriately with respect to contract requirements. The category identifier, *A*, Engagement, indicates that that the issue is one that occurred during the period in which IBM was engaging with the client prior to contract signing. A hierarchical structure for root cause capture allows us to collect information at some level of the hierarchy even if the detailed root cause cannot be identified. From a business perspective, the belief is that project leaders are so familiar with a contract that they will be able to indicate definitively whether a specific risk has occurred. However, they may not observe the issue at the lowest level of the risk tree. In this case, risk occurrence is recorded at the finest level of granularity in the risk tree that can be specified with confidence by the project executive. Due to the hierarchical nature of the taxonomy tree, a risk factor occurrence that is recorded at its finest granularity at a node r in a given tree also has an implicit interpretation as an occurrence at each parent node of r. This feature of the data enables analysis at any chosen level, or depth, in the taxonomy tree. We focus discussion in this paper on predicting root causes at the leaf nodes of the risk taxonomy. However, a predictive model of financial risks can be developed for different levels of the risk taxonomy depending on the information at hand, the desired accuracy of the predictive model, and the desired specificity of the risk mitigation recommendations.

While we have assumed that the identified factors are the drivers of financial performance, they may not, in fact, be the underlying cause of the observed

performance. However, these root causes were developed in conjunction with subject matter experts, starting with issues that had been experienced in past contracts and known to have impacted financial performance. Validation of causality can be obtained through observing the impact of mitigation actions taken to address identified performance factors, to determine whether the action works to reduce the impact of the performance issue on financial performance over time.

For each contract, we also need to understand the impact on contract financial performance of each identified risk factor. Specifically, we collect (1) the inception-to-date financial performance (e.g., measured as gross profit percentages) when each PMR is conducted, including both the actual performance and planned target; and (2) the root causes identified (i.e., a subset selected from the risk taxonomy) in each PMR, when financial performance was not satisfactory. For ongoing contracts, these financial impact attributions are elicited directly from the project delivery executives responsible for the deal as part of the PMR process, through interviews where the executives are asked to allocate the gap between actual contract financial performance and planned financial performance to individual root causes. The allocation can be specified as a percent of total gap relative to the planned financial target (% of target) attributable to a root cause, or directly in terms of gross profit percentage. In the first case, the weights are constrained to sum to 100 %, whereas in the second case, the sum of the values must equal to the overall gap to target.

Measuring project performance relative to its revenue target in percentage terms is the most common way that project financial health is discussed, putting all projects on an equal footing in terms of performance. Quantitative elicitation of weights has been discussed in other studies, for example, Murray and Lopez (1996), who describe elicitation of disability weights for use in summary health impact measures defining the severity of a disease. Our approach can also be thought of as similar to asking related questions from which a quantitative value can be derived, in that we do not elicit absolute dollar impact amounts for each root cause, rather percentage deviations, with the total constrained to sum to 100 %. See, for example, O'Leary et al. (2009), for a discussion of indirect elicitation. We did not elicit qualitative information on the uncertainty associated with the allocation as a complement to the quantitative information as prescribed in Sluijs et al. (2004), as the constraint that the quantities sum to 100 % makes capturing the spread appropriately a complex task. Note that IBM deal executives are not compensated based on their explicit responsibilities, but rather on overall profitability of a contract, even if the factors driving that profitability are not directly under their control. This mitigates the potential for bias in attributing underperformance to a factor that might have been controllable by the deal executive's team.

In cases where an expert does not feel confident about allocating the gap to specific risk factors, the impact can be uniformly distributed among them. Details on the use of these weights to compute risk impact estimates are presented in the following section. If a factor is not observed, it is assumed that the risk did not occur. For simplicity, we have assumed that risks occur independently. However, one could consider an approach in which combinations of performance factors driving over- or underperformance are treated as separate, additional performance

factors. Of course, for even a moderate number of individual factors, the number of combined factors becomes large rather quickly and may not be feasible in practice.

Note that the contracts analyzed in our application had information from an existing QA process associated with them, focused entirely on root causes of underperformance. Thus we chose to also focus on underperformance of contracts. That said, the approach described here is applicable to managing the performance of a contract, good or bad, assuming that root causes of good performance are also captured systematically. In our experience, however, experts have more difficulty pinpointing factors associated with improved performance.

Root causes captured in a standardized risk taxonomy and their associated impacts on contract financial performance form only one set of data needed for risk management. For pro-active risk management, information that can be used to predict potential risks prior to the start of a contract is also needed. In the case of SO contracts, the information collected in the technical and business risk assessments conducted as part of due diligence prior to contract signing form the basis of this information. These so-called deal fingerprints include items such as the total contract value, the planned duration of the deal, hardware and software dependencies, length of the transition period, etc.

The premise of our analytical approach is that certain types of projects exhibit a significant propensity for certain types of performance-related risk factors. For example, examination of historical contract information may indicate that those contracts that relied on geographically dispersed delivery teams had a much greater deviation from their financial targets. In this case, the makeup of the delivery team can be determined prior to the start of a contract and appropriate actions taken to mitigate the anticipated risk factor. In the case of SO contracts, risk assessments are carried out as part of the due diligence process and typically cover the following aspects.

– Business: evaluation of customer requirements, customer environment, solution definition, scope of service, contract financials, terms and conditions, etc.
– Technical: overall evaluation of solution design (accuracy, complexity, etc.) and verification, technical dependencies, resource and planning, as well as detailed assessments of specific technical areas, such as hardware, software, maintenance, helpdesk, etc.

These assessments are often conducted in the form of questionnaires. We treat the answer to each question in the questionnaire as an attribute of a contract in the corresponding business or technical area. In the existing risk assessment, the answers to each question are coded into different numerical values between 0 and 5 to represent different levels of uncertainties to which a contract is subject. For example, if a contract requires a standard technical solution that the service provider has successfully delivered many times in the past, then it may receive a value of 0 (risk free) for this attribute; If a contract requires a highly non-standard solution that has never been delivered before, then it may receive a value of 5 (extremely high risk).

Collectively, all answers to these questions form a vector of attributes, x_c, that characterizes a contract c in the engagement phase. We call this vector the contract

fingerprint. The goal of our predictive model is to predict the potential risks and their impact on a contract to be signed, based on the contract fingerprint data collected during engagement.

Risk and impact quantification

Our method for risk prediction is based on the concept of "similar" contracts. Namely, for any new contract, if we can identify a set of contracts that are "similar" to it and have been executed in the past, we can then apply the learnings (including the real issues observed and the corresponding root causes) from these contracts to predict what is likely to happen to the new contract.

The measures we want to predict are threefold. First, we want to predict how a new contract c will perform during delivery, i.e., its performance "class". We do this by measuring the similarity, $sim(c, c')$, between the new contract and each historical contract, c', based on their fingerprints. We then use a simple k-nearest neighbor (k-nn) classifier (see, e.g., Cover and Hart 1967) to make the prediction. Specifically, we take the top k historical contracts most similar to the new contract and do a majority voting based on their performance classes, i.e., "healthy" if a contract's financial performance met the planned target or "unhealthy" if it did not. We weight the vote from each historical contract with the inverse of its similarity with c to mitigate the situation in which the overall population of contracts has contracts falling predominantly into a single class.

This method of estimating the risk likelihood stems from classical nonparametric regression theory, which estimates the value of the function $f(\mathbf{x}$ at point $\mathbf{x}_i$ by identifying points within some defined neighborhood of $(\mathbf{x})$ and averaging their corresponding y_i values. The average can be a weighted average, where the weights are based on a kernel function $K(\mathbf{x} - \tilde{\mathbf{x}})$. Such kernel functions typically result in weights that decrease as the distance between $\mathbf{x}$ and $\tilde{\mathbf{x}}$ increases. In the case of Eq. (1) below, the similarity measure of $\mathbf{x}$ and $\tilde{\mathbf{x}}$ plays the role of the kernel function.

Second, we want to predict the risks that are mostly likely to cause problems in a new contract c. Formally, let S_c be the set of historical contracts similar to c. For contract $c' \in S_c$, denote the set of risks (or root causes) observed during contract delivery as $R(c')$. We estimate the likelihood of risk r occurring for contract c as

$$P(r) = \frac{\sum_{r \in R(c'),\, c' \in S(c)} sim(c, c')}{\sum_{c' \in S(c)} sim(c, c')}. \tag{1}$$

That is, the predicted likelihood of risk r occurring in contract c is estimated as a weighted average of the proportion of times risk r occurred in similar historical contracts c', with weights equal to $sim(c, c')$.

Third, for each risk predicted for contract c, we also estimate its potential impact on the financial performance of the contract. As discussed in "Data specification and collection", the impact of each contract root cause $r \in R(c')$ is elicited from an expert familiar with the contract. For instance, if contract c' had a gross profit margin target of $\alpha_{c'}$ (a percentage), but only achieved an actual margin of $\beta_{c'}$, it

missed its target by $\delta_{c'} = \alpha_{c'} - \beta_{c'}$. The experts are asked to attribute a portion of $\delta_{c'}$, $\delta_{c'}(r)$, to each root cause r, with the constraint that

$$\sum_{r \in R(c')} \delta_{c'}(r) = \delta_{c'}. \tag{2}$$

The potential impact of risk r on the financial performance of contract c is estimated as a weighted average of the impacts attributable to risk r across historical contracts, again with weights equal to $sim(c, c')$. Note that if the financial target of interest is revenue, the impact estimates are computed in terms of percentage deviation from target revenue instead of raw dollars, so as to avoid bias arising from widely disparate revenues sizes across contracts.

All the above predictions depend on the notion of similarity between two contracts, which is based on the fingerprints of the two contracts. A naive way to measure similarity between contracts is to use a simple Euclidean distance measure. However, this has the disadvantage that differences between values along every dimension of the contract fingerprint are given the same significance. In our work, we use a weighted distance metric to gauge contract similarity. Specifically, we define the similarity between two contracts as the inverse of the distance between them:

$$sim(c, c') = 1/d(c, c'), \tag{3}$$

where the distance between contract c and c', $d(c, c')$ is defined as

$$d(c, c') = d_A(c, c') = ||x_c - x_{c'}||_A = \sqrt{(x_c - x_{c'})^T A (x_c - x_{c'})}. \tag{4}$$

Here, x_c represents the n-dimensional fingerprint vector of contract c and A is a positive semi-definite, diagonal transformation matrix, i.e., $A \succeq 0$ and $A = diag\{-A_{11}, A_{22}, \ldots, A_{NN}\}$. This is essentially the same as applying weights to each dimension of the fingerprint vector. To determine A, we use the method developed in Xing et al. (2002). From our training data set, we obtain sets of similar contract pairs, $\mathcal{S}$, and sets of dissimilar pairs, $\mathcal{D}$. Two unhealthy contracts are deemed similar if the distance between them is smaller than a threshold θ. More formally, contracts c, c' are similar if

$$d_J(c_i, c_j) = 1 - \frac{R_{c_i} \cap R_{c_j}}{R_{c_i} \cup R_{c_j}} < \theta. \tag{5}$$

In addition, if c_i is a healthy contract and c_j a troubled contract, then (c_i, c_j) is considered to be a dissimilar pair in $\mathcal{D}$. With the above definition of similar and dissimilar pairs, we apply the Newton–Ralphson method to optimize A so that the distances between similar pairs of contracts are minimized, while the distances between dissimilar pairs are maximized.

Note that while the scale used to form the contract fingerprint is assumed to be linear in the simplest case, a nonlinear scale could also be accommodated through appropriate modification of the distance measure defined in Eq. (4). For example, one could modify the assigned distance between different values of the deal characteristic to reflect that a difference between value 4 for one deal and value 5 for

a different deal is of greater importance than a difference between the values 2 and 3. In other words, a nonlinear mapping from the 0–5 scale can be used to reflect the logrithmic nature of the assessment scale. Since the QA tool currently in use assumes a linear 0 to 5 rating scale, our analysis also follows this (potentially flawed) assumption.

Insight generation

The third component of the risk and decision management system is focused on reporting, i.e., providing information to project executives that are insightful and actionable. Performance reporting is a crucial step in ensuring that all parties have access to the same information in the same format. For the system developed within IBM, we have defined a set of reports providing different views of performance, both for individual contracts and for portfolios of contracts. Project executives who need to access detailed information regarding a contract can view reports containing contract-specific risks and mitigation actions, while business executives may prefer to see an overview of performance of a set of contracts, by industry or geography, for example. Figure 2 provides an exemplary report for a specific contract. Key attributes of the contract are shown in the top portion of the report, with the top 15 predicted risks, as measured according to potential impact on gross profit margin, shown in the bottom portion, with the impact values depicted as horizontal bars. A business user can click on an individual bar to be shown additional information about the risk, such as its predicted likelihood of occurrence and the list of historical contracts determined to be similar to the selected contract. Presentation of the similar contracts and the use of a simple k-nn classifier based on the contract similarity may provide an advantage over other, more complex analytic techniques in terms of driving analytics-based decision making for risk management, as the analytic results may be easier for business executives to understand and accept. In our experience at IBM, a decision maker's understanding/confidence in the analytics underlying a reported prediction significantly impact their use for decision making. While research such as that of Bharati and Chaudury (2003) finds that information quality and system quality influence decision-making satisfaction more than information presentation, the impact of model understanding on decision-making satisfaction is a question for future research.

Additional reports might show recommended mitigation actions to address the top risks. A business analyst or initiative leader might choose to view such a report after observing that the initiative is expected to underperform against its target, for example, and would like to understand why and what might be done to prevent this from happening. Risk status is included in reporting and is tracked over time. That is, on a regular basis, previously reported risks are reviewed by relevant stakeholders—which risks are resolved and how, which risks remain influential and what has been/could be done to address the risks. As a result, best practices and lessons learned for addressing specific risks are systematically culled, providing various business benefits such as guiding mitigation planning.
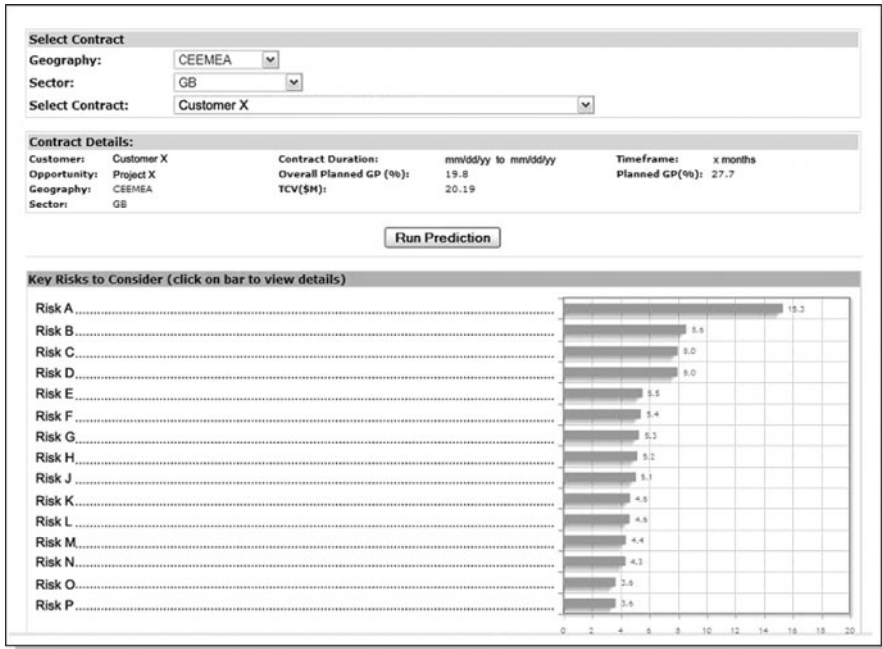
Fig. 2 A screen shot of the top 15 predicted risks for an anonymized contract taken from the SO financial risk analysis tool implemented at IBM Research

Application of approach to managing strategic outsourcing financial risk: results and business impact

Modeling results using historical SO contracts

We applied the risk prediction method introduced in the previous section to roughly 56 historical contracts, for which we could obtain both financial performance data and the detailed PMRs. The effectiveness of our model was evaluated on all these contracts, using leave-one-out cross validation. In other words, for each contract, we use all historical contracts except the one to train a predictive model; we then use the trained model to predict for the selected contract and evaluate how well the model performs. The method of Xing (2002) with $\theta = 0.8$ was used to compute the optimal value of A for classifying contracts as healthy or troubled, where k was set to equal 56 in the k-nn classifier. In other words, a weighted average of the performance of each of the contracts was used for prediction, with weights as described in "Risk and impact quantification".

For each contract, we first evaluate how well our model is able to predict its future health, i.e., will it be "healthy" or "unhealthy". Then, for each troubled contract, we predict the top risks that are most likely to turn into issues, and compare the predicted risks with the actual root cause analysis results documented in the PMR. We then rank the predicted risks based on the financial risk exposure

driven by the predicted risk, defined as the predicted likelihood of a risk multiplied by its predicted impact, i.e.,

$$\text{Risk exposure} = \text{likelihood} \times \text{impact}, \qquad (6)$$

as in Condiman et al. (2007, Chapter 1). Out of 56 contracts, 16 (29 %) were troubled. Using the nearest neighbor approach, we were able to correctly identify 14 of the 16 troubled projects with a precision of 87.5 %, while predicting trouble for one project that was actually healthy i.e., a recall of 97.5 %. On average then, the overall accuracy for predicting which contracts will be financially troubled is 95 %. These results indicate that healthy (or unhealthy) contracts can be clearly differentiated on the basis of the contract fingerprint and therefore a simple k-nn classifier works well. In contrast, a naive model that predicts the incidence of troubled projects at the same rate as in the analyzed sample would result in an accuracy of 71 %.

For those contracts predicted to be troubled, we focus on the top 15 predicted risks for evaluating the effectiveness of the risk prediction method. The selection of the top 15 is a trade-off between model effectiveness and practicality: too short a list will fail to identify the actual risks, while too long a list will make it infeasible to track down every predicted risk.

We have assessed the accuracy of predicting the root causes of troubled contracts by focusing on the top 15 root causes (out of 133). Our measure of accuracy looks at the number of troubled contracts in which our top 15 predicted risks contained at least one of the actual root causes of trouble. In this case, we were able to correctly predict at least one of the actual risks for 77 % of the contracts predicted to be troubled. A typical contract has one to three root causes reported. For a contract with only one root cause, the probability of predicting it in 15 random guesses is $15/133 = 0.11$, assuming all root causes are equally likely. For a contract with two root causes, the probability of predicting at least one based on 15 random guesses is $1- (131/133)*(130/132) *...* (117/119) = 0.21$. For a contract with three root causes, the probability is around 0.30. While not all of the root causes are equally likely in practice, overall our result of 0.77 provides a significant improvement over random guessing. Furthermore, around 40 %, on average, of the actual root causes observed in a contract were included in the list of top 15 predicted risks. Given that we have a list of 133 potential risks from which to predict, we consider these results very encouraging.

The main factor driving the lower than desirable accuracy of our model is the fact that we have relatively few unhealthy contracts in our training data. Hence, a number of root causes were not observed frequently enough to be properly accounted for in the model. We believe these shortcomings can be overcome as we accumulate more training data in the future.

Impact of financial risk management in practice

An important factor in gaining acceptance of new capabilities for risk management is that they be well integrated into an existing process. The steps in the end-to-end SO contract risk management process are shown in Fig. 3, with the Risk and Impact Prediction box representing a new step based on the financial risk prediction
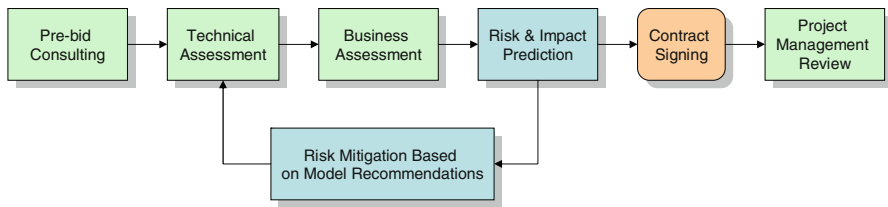
Fig. 3 Integrating predictive analytics into the end-to-end risk management process

models. The process begins with pre-bid consulting, i.e., the service provider engages with the customer to better understand their requirements, develop a technical solution in response to the customer's requests, and propose a structure for the overall contract. Then, two types of assessments, technical and business, are conducted, as described in "Data specification and collection". These assessments provide the contract fingerprint as input to our risk prediction tool, which predicts the financial outlook and the top-15 risks for the new contract, as shown in Fig. 2.

Based on the results of the new predictive analytics, the risk manager now reviews and prioritizes the risks based on the predictions and decides on mitigation actions as recommended by the tool. Prior to the creation of the predictive models, risk assessment and mitigation actions were based on a risk manager's experience in evaluating output from the technical and business assessments, which is primarily qualitative. Note that the "assess–predict–mitigate" steps can be conducted iteratively, as the engagement process develops and the risk manager continues to mitigate the identified risks. Finally, when the risk manager feels comfortable with the proposed contract, he provides his recommendation to the decision maker, who has ultimate authority to sign the contract, after which the contract delivery process begins.

The new approach has been in place at IBM to manage risks in a set of critical initiatives for only about 6 months, with information from more than 500 contracts incorporated into the system. To date, the approach has resulted in two types of value to the end user: (1) prediction of risk exposure before project start, and (2) visibility into ongoing issues during project operation, with explicit risk tracking until resolution. More specifically, the system is used to predict, prioritize, and quantify risk exposure for individual risks as well as total contract and contract portfolio risks. This information has been used to either develop appropriate mitigation plans before contract start, or to support a decision not to pursue a particular contract if deemed too risky. Prioritization of predicted risks is especially important in light of Pennington and Tuttle (2007), who find that information overload can negatively affect risk assessment in software projects. Hence, accurate identification of at-risk projects for in-depth review is critical to improved risk management. Creation of mitigation plans has been greatly facilitated by users having access to a database of typical issues/actions for each risk that they can conveniently reference. The database was initially created based on expert input, but will be continuously refined based on observation of effectiveness of particular mitigation actions.

Another source of value has been generated from providing explicit visibility into current risks, such as understanding top 15 risks within a particular geography or set of contracts of a particular size. Coupled with explicit risk tracking until resolution, it provides a disciplined approach to improve performance. Additionally, it enables historical trend analysis, such as highlighting top risks across multiple PMRs and their evolution in terms of relative importance. This has helped inform strategic adjustments to mitigation actions. While the introduction of the new system has required investment in process transformation and user training, the benefits of the implemented performance management system have already been found to outweigh the costs. Risk managers are comfortable enough with the model results to prioritize and implement action plans based on the predicted risks and their expected impacts.

Note that as the results of the predictive modeling are deployed into the risk management process, these will affect risk management behavior, ultimately changing the outcome of the contracts to which these are applied. We expect that the new data collected after model deployment will be different from the existing training dataset, in that many of the predicted risks will be mitigated early in the engagement and hence never become issues in delivery. We also anticipate that the service environment will constantly change. As a result, there will always be new risks emerging and old risks that become obsolete. All these require the risk taxonomy, as well as the predictive model to be periodically updated with the new training data, so that the model can remain effective in the ever-changing risk management environment.

Conclusions and directions for future work

Regular checkpoints of portfolio health with performance updates from constituent projects have been an integral part of business management systems for a long period of time. However, the new risk management approach described here has reshaped the checkpoint process by instilling significantly more structure and analytical rigor into the status review. First, as we have described, the new approach puts risk factors into perspective by utilizing the structured risk taxonomy. The importance of having a common language cannot be overemphasized in knowledge development and management. Second, the new approach seeks expert opinions to quantify and differentiate the impact of individual performance factors, which collectively explain an observed deviation from the target. Impact quantification forces objectivity and facilitates rigorous analytics. Finally, the new process requests that mitigations be reported for all identified risks and tracks risk status and mitigation efficacy over time.

Critical success factors for the implementation of an analogous risk and performance management approach for other types of business initiatives include (1) Importance of the problem to the end user, (2) sufficient amount and quality of data, and (3) team skills and composition. With regard to the first point, process transformation is difficult, and one needs to address the key business problem to generate consistent support throughout the organization. Additionally, one needs to

have a sufficient number of data points to generate meaningful predictions. Also, each data point needs to be of sufficient quality. Initially, some judgment and data clean-up may need to be applied. Over time, process and system improvement for consistent data capture may be needed. Finally, the team needs to have a combination of general consulting skills (to understand the business problem and propose process and other changes), analytics skills (to develop predictive algorithms), and IT implementation skills (to build prototype tools and systems for data capture and visualization). This combination of skills allows rapid prototyping and iteration to design a solution addressing the needs of the particular problem at hand. Examples where the general methodology may be applicable include sales engagement with external customers (clients or partners), large internal software development initiatives, recruitment management, and business partner management. Initial benefits can be reaped from development and application of the data capture and reporting pieces of the system alone, with additional benefits generated from predictive modeling and impact assessment over time. These benefits include strengthening accountability in risk management and facilitating rapid identification of best practices to be used by other contract delivery teams facing similar risks. Not only does the new approach enable more informed business decisions, but it also transforms the decision-making process from "sense and respond" to "predict and act."

While the current system at IBM is already providing value, we are continuing to develop analytic and system capabilities to provide additional benefits as data continue to accrue and business users become more accustomed to and accepting of the new process and system. For example, temporal prediction, i.e., prediction not only of risk occurrence but also of the time period in which a risk is likely to occur, can support decision making regarding how to allocate risk mitigation effort over time, as can quantitative assessment of mitigation action impact based on, e.g., statistical intervention analysis (Box et al. 2008, Chapter 13). Additionally, underperformance against targets often leads to oversight of positive factors impacting contract performance. Extending the taxonomy to allow for tracking and predicting root causes of positive financial performance relative to financial targets will enable learning of positive patterns of performance over time. More formal methods for updating the deal descriptor set over time, and the predictive models in general, will be required to address the evolution of contract characteristics and/or observed risks over time. We plan to develop approaches to address these challenges in future work.

References

Aundhe MD, Mathew SK (2009) Risks in offshore IT outsourcing: a service provider perspective. Eur Manag J 27(6):418–428

Box G, Jenkins G, Reinsel G (2008) Time series analysis: forecasting and control, John Wiley & Sons

Condiman L, Louisot J-P, Naim P (2007) Risk quantification: management, diagnosis and hedging, John Wiley & Sons

Chapman RJ (2011) Simple tools and techniques for enterprise risk management. John Wiley & Sons

Cover T, Hart P (1967) Nearest neighbor pattern classification. IEEE Trans Inf Theory 13(1):21–27

Deleris LA, Katircioglu K, Kapoor S, Bagchi S, Lam R, Buckley S (2007) Adaptive project risk management analytics. In: Proceedings of the 2007 Winter Simulation Conference 2234–2240

Goo J, Kishore R, Kichan N, Rao HR, Song Y (2007) An investigation of factors that influence the duration of IT outsourcing relationships. Decis Support Syst 42(4):2107–2125

Juliusson EA, Karlsson N, Garling T (2005) Weighing the past and the future in decision making. Eur J Cogn Psychol 17(4):561–575

Labbi A (2005) Handbook of integrated risk management for E-Business: measuring, modeling, and managing Risk. J Ross Pub

Leung HM, Tummala VMR, Chuah KB (1998) A knowledge-based system for identifying potential project risks. Omega 26(5):623–638

Loosemore M, Raftery J, Reilly C, Higgon D (2006) Risk management in projects, Routledge

McIvor R, Humphreys PK, Wall AP, McKittrick A (2009) Chapter 6-evaluation of the outsourcing framework: a study of performance measurement in the outsourcing decision, CIMA Publishing, Oxford, 291–296

Mojsilovic A, Ray B, Lawrence R, Takriti S (2007) A logistic regression framework for information technology outsourcing lifecycle management. Comput Oper Res 34(12):3609–3627

Murray CJ, Lopez AD (1996) The global burden of disease. Harvard University Press, Cambridge

O'Leary R, Low Choy S, Murray J, Kynn M, Denham R, Martin T, Mengersen K (2009) Comparison of three expert elicitation methods for logistic regression on predicting the presence of the threatened brush-tailed rock-wallaby *Petrogale penicillata*. Environmetrics 20:379–398

Pennington R, Tuttle B (2007) The effects of information overload on software project risk assessment. Decis Sci 38(3):489–526

Pratyush B, Chaudhury A (2004) An empirical investigation of decision-making satisfaction in web-based decision support systems. Decis Support Syst 37(2):187–197

Ratakonda K, Williams R, Bisceglia J, Taylor RW, Graham J (2010) Identifying trouble patterns in complex IT services engagements. IBM J Res Develop 54(2):1–9

Sluijs JP, Craye M, Funtowicz S, Kloprogge P, Ravetz J, Risbey J (2005) Combining quantitative and qualitative measures of uncertainty in model-based environmental assessment: the NUSAP system. Risk Anal 25:481–492

Tversky A, Kahneman D (1974) Judgment under uncertainty: Heuristics and biases. Science 85(4157):1124–1131

Xing E, Ng A, Jordan M, Russell S (2002) Distance metric learning, with application to clustering with side-information. In: Proceedings of advances in neural information processing systems 15:505–512 (MIT Press)