Mining for Gold: How to Predict Service Contract Performance with Optimal Accuracy based on Ordinal Risk Assessment Data

Sinem Güven
Tsuyoshi Idé
Sergey Makogon
Mathias Steiner
Amit Paradkar
Alejandro Venegas

IBM T. J. Watson Research Center
IBM Global Technology Services

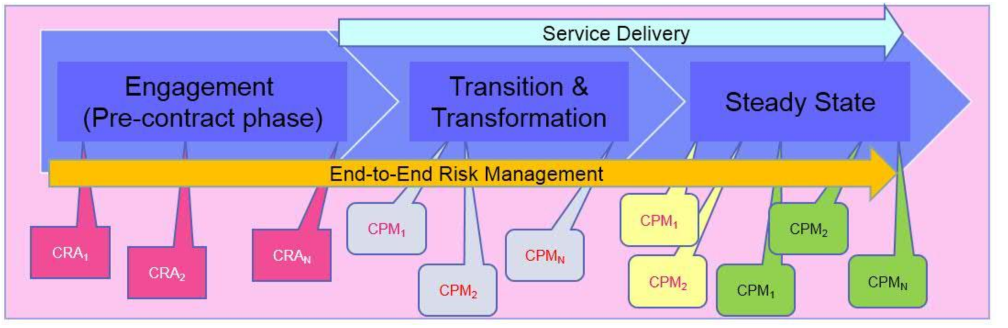
28 June 2014

IEEE SCC’14, 27 June - 2 July, 2014, Anchorage, Alaska

Introduction and Motivation

- Proactive management of service contract risks ahead of contract signing is becoming increasingly important for IT service providers

- Within an end-to-end risk management process, various **Contract Risk Assessments (CRAs)** are performed at multiple stages before a service contract is signed
  – Based on CRA data, service providers seek to have proactive contract risk prediction models

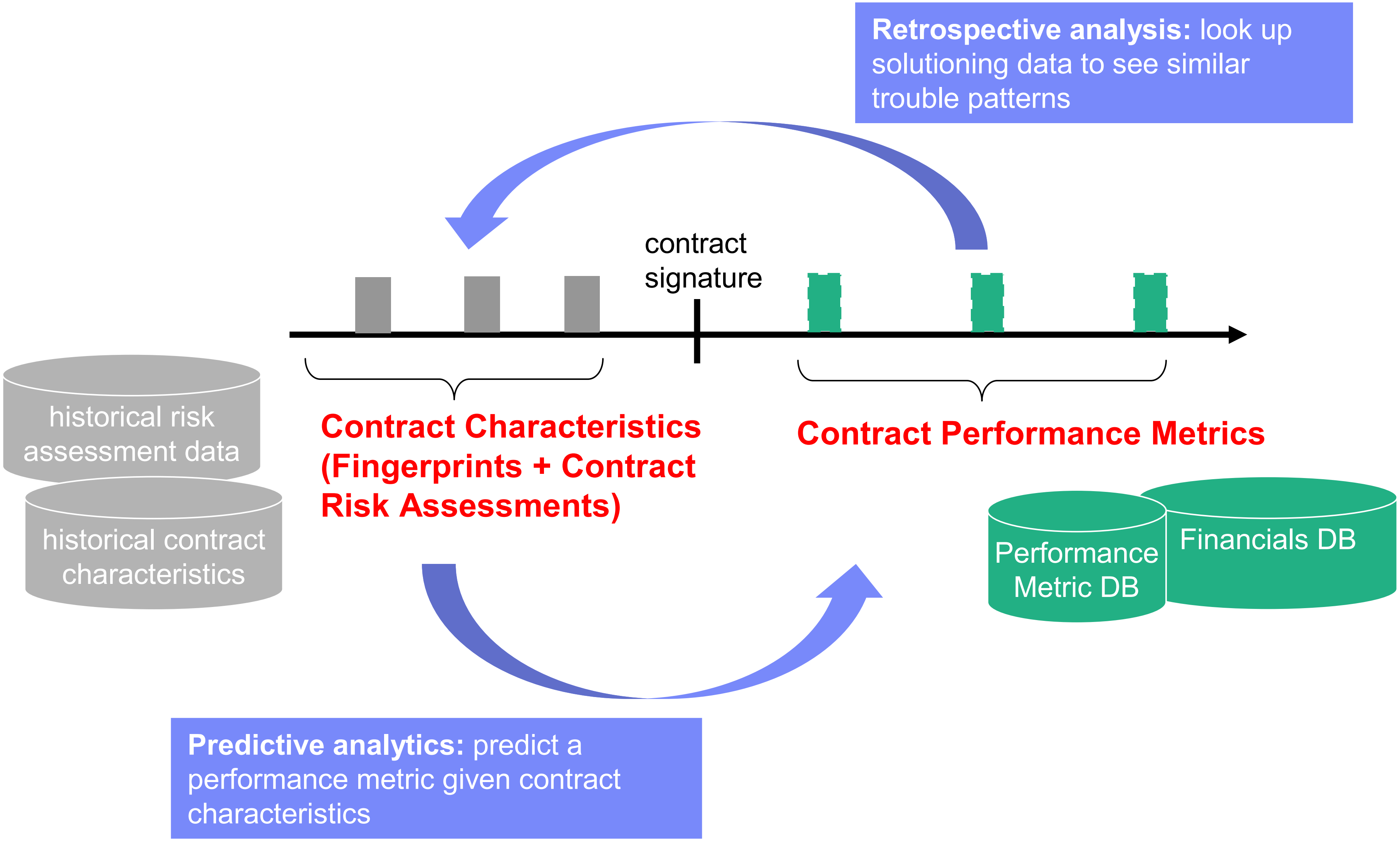


- The performance of the project is evaluated by various reviews to produce **Contract Performance Measures (CPMs).**

- Naïve statistical modeling approaches, such as linear regression, are not readily applicable to such data sets due to:
  – High dimensionality -- wide range of risk assessments,
  – Sequential nature – assessments and reviews are performed some multiple times

- This work describes an analytical methodology that enables optimal risk prediction for service contracts with ordinal CRA data

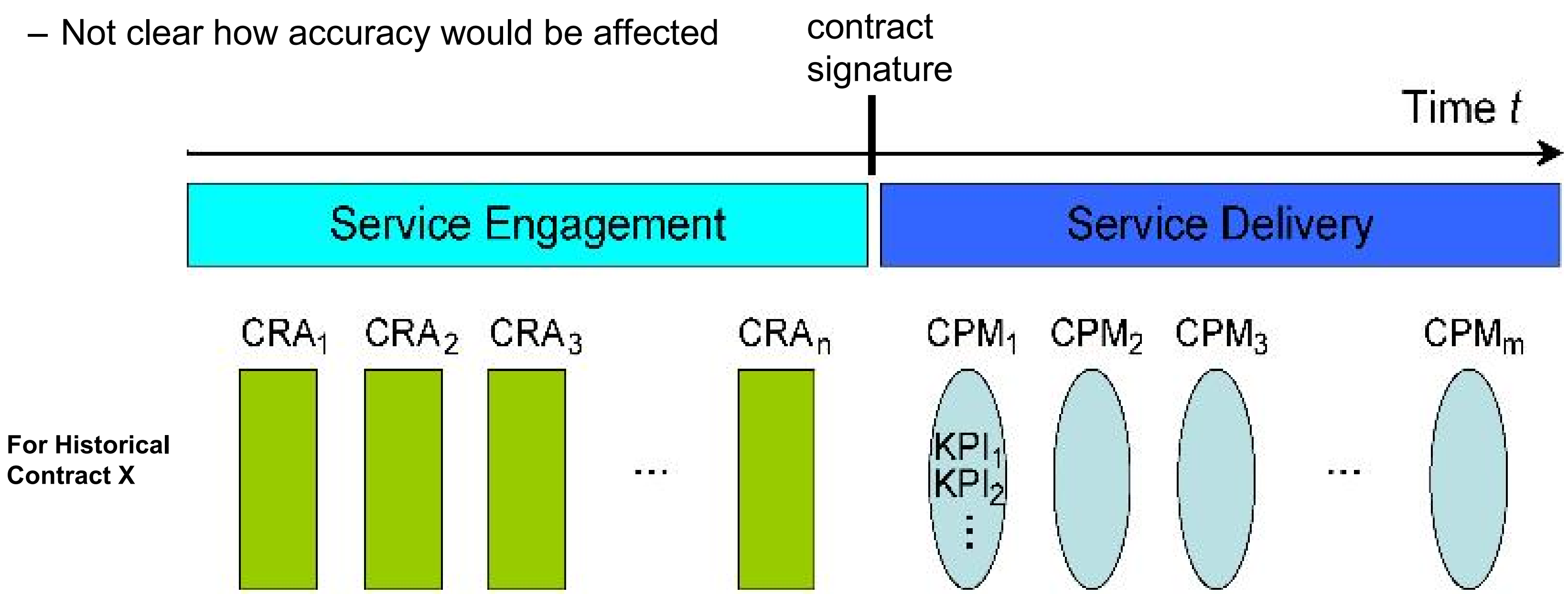
Training Service Contract Prediction model:
Linking pre- and post-signature data to predict future troubled contracts



Data Selection Challenge

- Need to predict one or more Delivery KPIs (Contract Performance Measures; CPMs) at Engagement time
  - Given a wide range of combinations, which data set to use?

- Data selection criteria should be applied to narrow down scope
  - Unique characteristics of service contract data set render data selection non-trivial
    - Variable time delay
    - Evolving data
  - No clear, well-defined strategy known for such a complex data set
  - Not clear how accuracy would be affected



Initial Approach: Determine Data Selection based on Business Insights

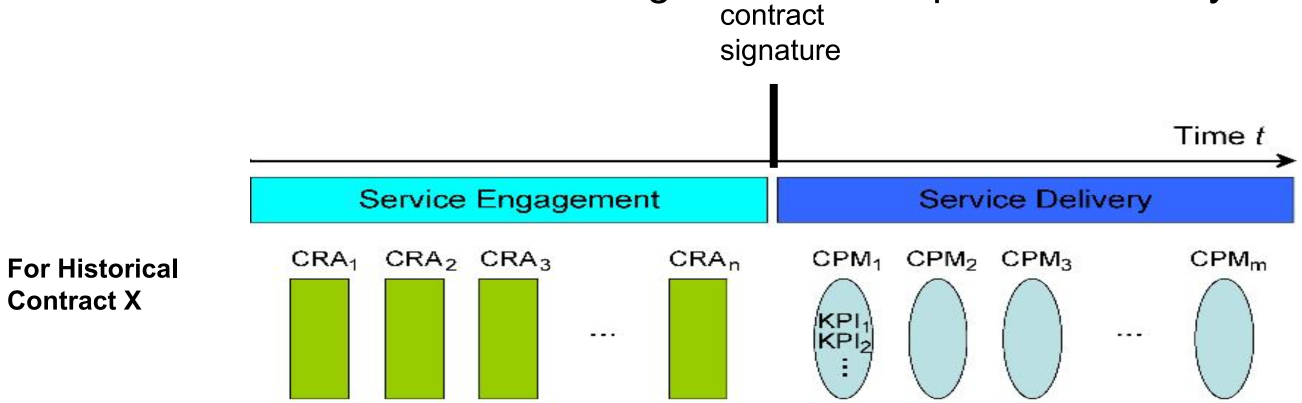
- Goal:
  - Predict contract profitability at Engagement based on Risk Assessment data

- Determined a training data set based on business insights:
  - **<u>Input:</u>** Latest Risk Assessment
    - best representation of the risks in a new opportunity
  - **<u>Target:</u>** The KPI (plan-actual difference in gross profit) measured closest to the end of T&T
    - best representation of T&T performance

- Accuracy Metrics:
  - Overall Directional Accuracy
  - Non-profitable Contract Prediction
  - Profitable Contract Prediction

Metric	Overall prediction accuracy	Non-Profitable Contract Prediction	Profitable Contract Prediction
Accuracy	59%	71%	52%

- Need a data-driven selection of training data set for optimal accuracy



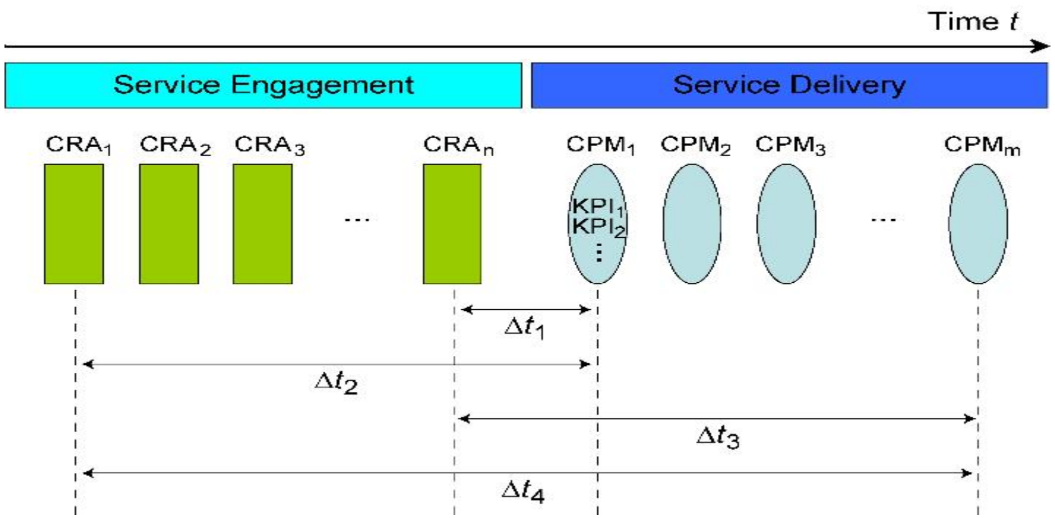
New Approach: Determine Data Set Selection through Data Optimization

- Determine if selected parameter (*time delay)* has any significance in predictive performance

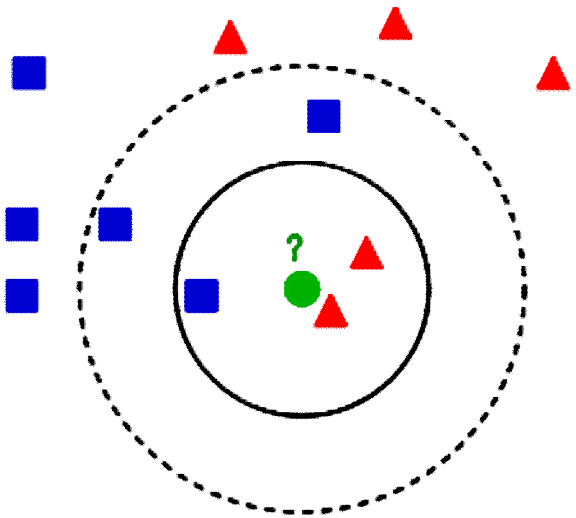
- If selected parameter (time delay) does have significance, develop a systematic approach to selecting the *optimal data set combination (optimal time window)* in the data set
  - Create all data combinations
  - Train / test with a predictive model

- Compare different classification algorithms by training a predictive model using this data set to maximize prediction accuracy

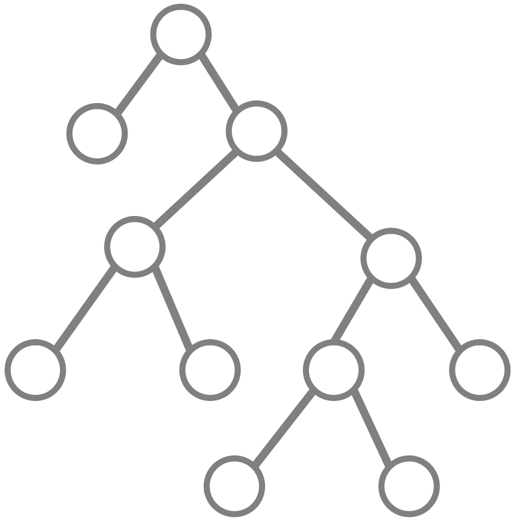
For Historical Contract X



k-Nearest Neighbors

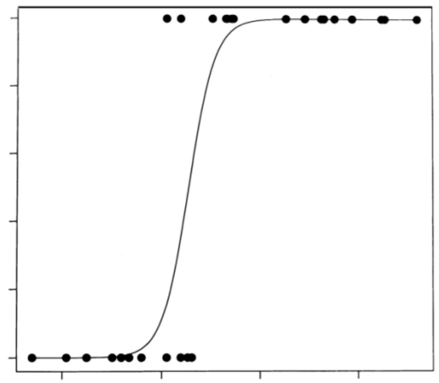


Decision tree (C5.0)



Logistic regression

$$P(Y=1|X)=\frac{1}{1+\exp(w_0+\sum_{i=1}^{n} w_i X_i)}$$



Impact of Time Delays in Accuracy

- Tested over last/first selection criteria for ΔGP
  - Significant dependency on the time delay Dt_i that characterizes the data sets.

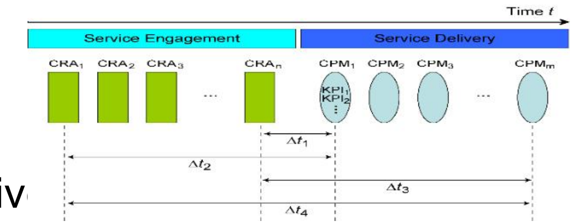
Data Set		Mean/Median Time Delay (months)	C5.0 Classifier Training/ Testing Accuracy (%)	Logistic Regression Training/ Testing Accuracy (%)
Dt_1	last-first	6.5 / 5.2	89 / 49	79 / 59
Dt_2	first-first	10.8 / 9.9	85 / 59	80 / 62
Dt_3	last-last	22.1 / 21.9	81 / 67	79 / 69
Dt_4	first-last	26.3 / 26.0	67 / 74	75 / 69

* 30% of randomly selected samples are used for testing

Modeling results for K(ΔGP) obtained with data sets characterized by different time delays

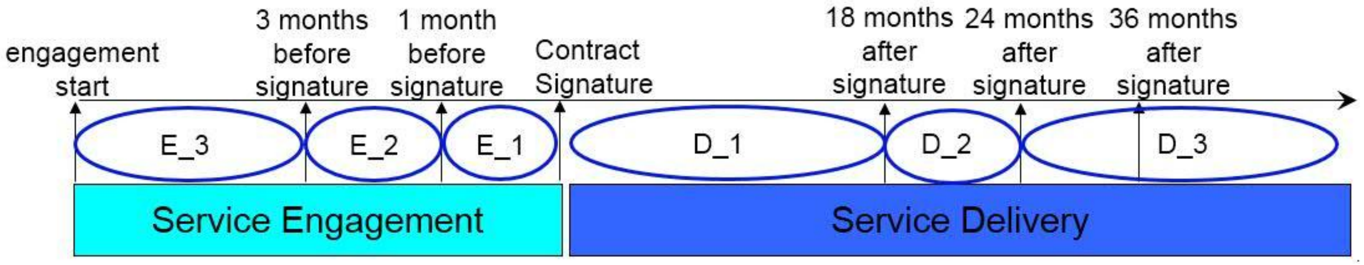
- Data-driven selection method confirms the dependence of classification accuracy on time delay between CRA and CPM.

- How about the combinations other than first/last?
  - Need a strategy to select optimal data set for training predictiv



Optimal Data Set Selection

- Further divide and conquer



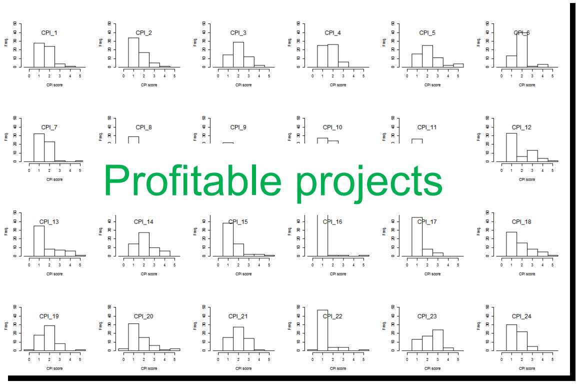
Engagement and Delivery time windows in the SO lifecycle

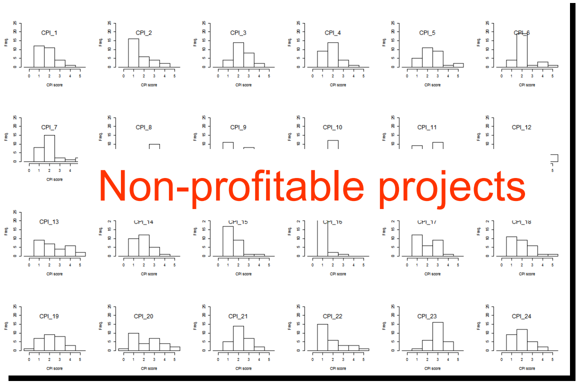
- Generate training samples by taking a combination of {Engagement X Delivery} windows.
  - Yields 9 *Time Window Combinations (TWC)* and associated data sets to train the model with
  - Matches N^{th} Engagement window with the M^{th} Delivery window, and so on.

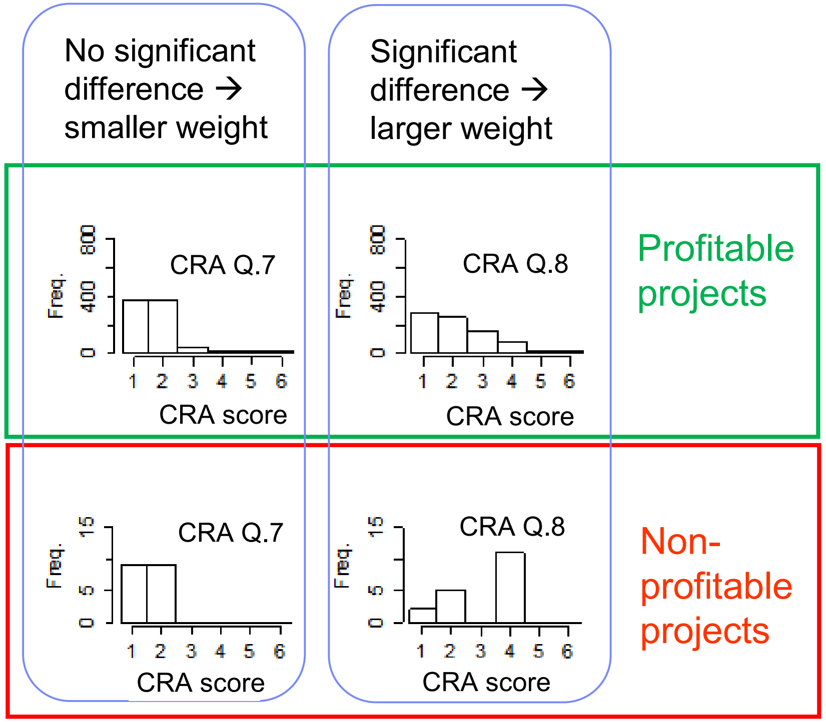
- Once these data sets are generated, we determine the optimal {Engagement X Delivery} TWC by evaluating the “informativeness” of each TWC

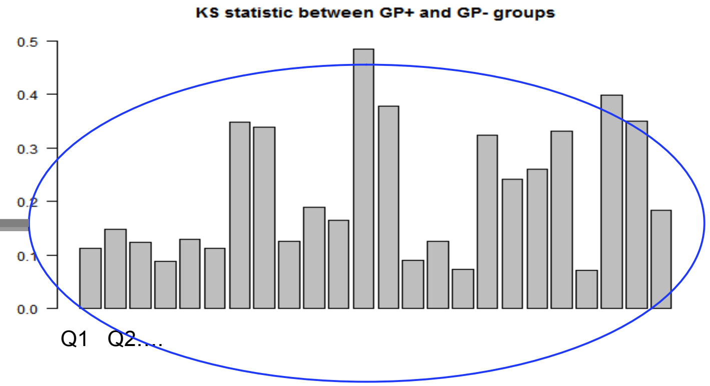
- Developed a method based on statistical two-sample tests (see next page).

Computing informativeness of TWCs using Kolmogorov–Smirnov statistics









Average over all of the CRA questions for a choice of bucket combination

Results: Improved Accuracy

- The results of KS statistics revealed that the most optimal TWC is {E_2 X D_3}

- Also consistent with the time delay study
  - Optimal {E_2 X D_3} TWC belongs to the Dt_4 window, which was determined to be the most optimal for our data set.

Initial Data Set

Metric	Overall prediction accuracy	Non-Profitable Contract Prediction	Profitable Contract Prediction
Accuracy	59%	71%	52%

Optimal Data Set

Run-time Window	Overall prediction accuracy	Non-Profitable Contract Prediction	Profitable Contract Prediction	Engagement Training Data	Delivery Training Data
E_1	71%	72%	70%	E_2	D_3
E_2	76%	86%	68%	E_2	D_3
E_3	74%	81%	68%	E_2	D_3

Different run-time scenarios tested with optimally trained (E_2 and D_3) model

Conclusions and Future Work

- Described a novel methodology for building a financial performance prediction model with enhanced accuracy using ordinal risk assessment data as model input.

- In particular, investigated how the time delay between contract risk assessments and contract performance measures within the IT service delivery lifecycle affects the accuracy of contract risk models in the IT outsourcing domain.

- Results showed that variations of the median time delay between contract risk assessments and the contract performance measures accounts for prediction accuracy variations as large as 25%.

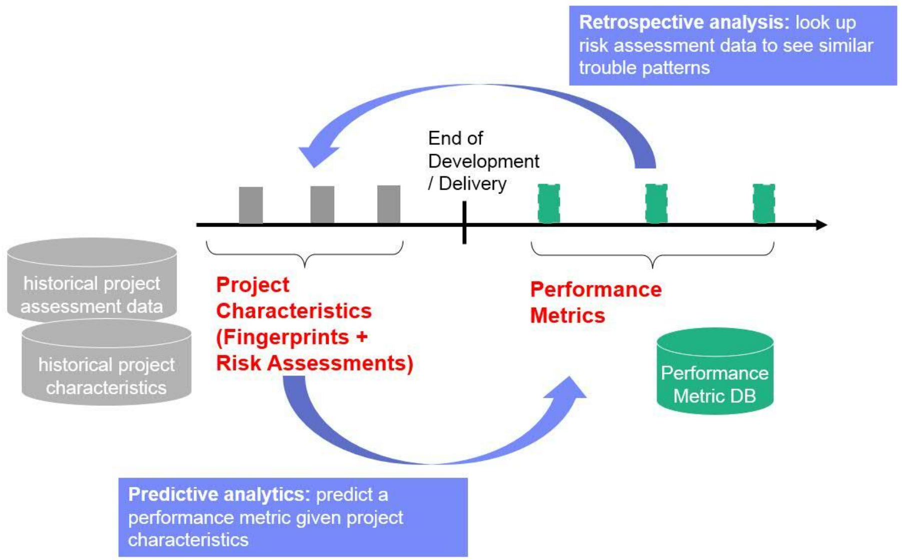
- Statistical modeling strategies such as linear regression fall short when it comes to handling sequential and ordinal data sets, which are characteristics of the IT outsourcing domain.
  - Data mining and machine learning approaches ensure selection of optimal model parameters, thereby maximizing accuracy of risk prediction models.

- Conclude that identification of relevant data selection criteria, such as *time delay* between risk assessment and performance measurement, is key for optimizing prediction accuracy in data-driven, predictive risk modeling.

Thank You!

Applicability to Other Domains

- Project Management
  - Our methodology could be used to select the best representation of a project in terms of its risk assessments and performance measurements to train a model to predict whether a new similar project is going to be successful
    - E.g., Software development project
    - **Several assessments** during development → project risks
    - **Several performance tests** before release + through actual usage → project performance outcomes
    - For a given new software development project similar to some historical ones in terms of project features and risks, need to predict whether it will be successful in terms of a performance metric (e.g., no crashing)
    - To train the predictive model, need to understand which {assessment x performance} pairs best represent a historical project
    - Our methodology can be used to obtain the optimal data set selection (optimal pairing)



Applicability to Other Domains

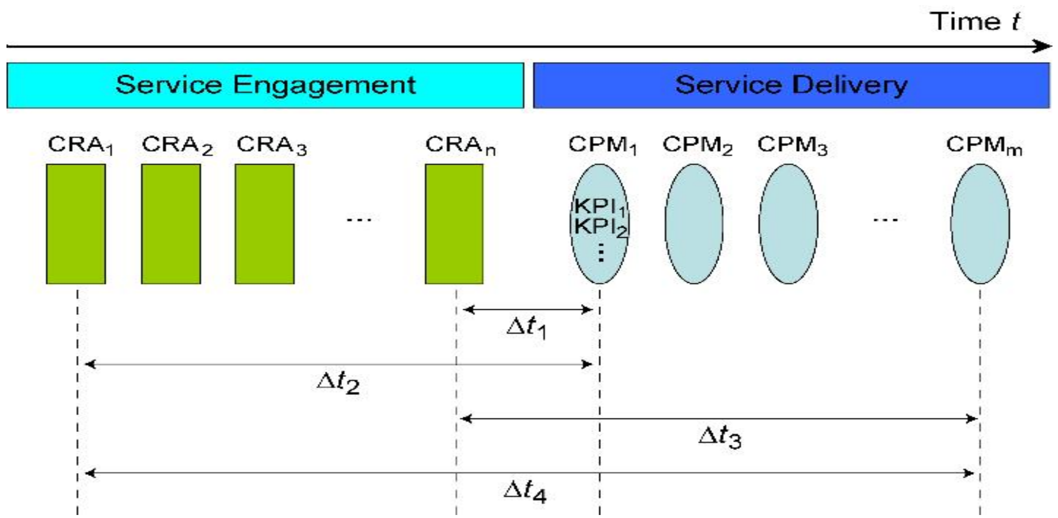
- Manufacturing
  - Our methodology could be used to select the best representation of a manufacturing project in terms of its risk assessments and performance measurements to train a model to predict whether a new similar project is going to be successful
    - E.g., Car manufacturing
    - **Several assessments** during manufacturing → project risks
    - **Several performance tests** before delivery + through actual usage → project performance outcomes
    - For a given new car manufacturing project similar to some historical ones in terms of project features and risks, need to predict whether it will be successful in terms of a performance metric (e.g., no engine problems)
    - To train the predictive model, need to understand which {assessment x performance} pairs best represent a historical car manufacturing project
    - Our methodology on Slide 6 can be used to obtain the optimal data set selection (optimal pairing)

- Natural Resources Management
  - Our methodology could be applied to optimize drilling/mining conditions in natural resources management. In this case, the CRAs represent recovery-related risks (e.g. operational, environmental, etc.) while the CPMs represent key performance indicators such as resource recovery rates and return on investment which are monitored on a continuous basis.

New Approach: Determine Data Set Selection through Data Optimization

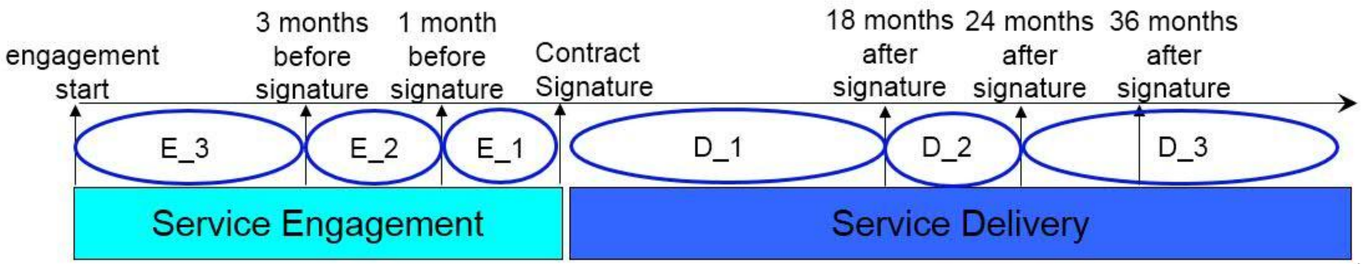
- Develop a systematic approach for optimal data selection to maximize prediction accuracy
  - Choose a parameter for data set selection
    - Time delay: first vs. last
    - Quality: best vs. worst
  - Determine if selected parameter (*time delay)* has any significance in selecting training data
    - For example, if a given historical contract has same CRA repeated several times, understand if using first one vs. last one has any effect on accuracy of models trained with such CRAs.
    - Create all data combinations
    - Train / test with a predictive model
  - If selected parameter (time delay) does have significance, select the *optimal data set combination (optimal time window)* in the data set
  - Train the predictive model using this data set to maximize prediction accuracy

For Historical Contract X



Optimal Data Set Selection

- Further divide and conquer



Engagement and Delivery time windows in the SO lifecycle

- Generate training samples by taking a combination of {Engagement X Delivery} windows.
  - Yields 9 *Time Window Combinations (TWC)* and associated data sets to train the model with
  - Matches N^{th} Engagement window with the M^{th} Delivery window, and so on.

- Once these data sets are generated, we determine the optimal {Engagement X Delivery} TWC by evaluating the "informativeness" of each TWC

- Developed a method based on statistical two-sample tests (see next page).
  - For each of the training data sets belonging to the 9 TWCs, separate the historical contracts into two groups depending on the directionality (positive or negative) of their Gross Profit Variance.
  - Evaluate the difference between probability distributions of the historical contracts' CRA questions.
    - To quantitatively measure the distributional distance, we use the single-variable Kolmogorov–Smirnov (KS) statistics averaged over the CRA questions.
    - The bigger the averaged KS statistic is, the more informative the TWC is.
  - If there is no significant difference between the positive and negative Gross Profit Variance groups, we conclude the TWC is not informative.