



GOMA: Supporting Big Data Analytics with a Goal-Oriented Approach

DOI:

[10.1109/BigDataCongress.2016.26](https://doi.org/10.1109/BigDataCongress.2016.26)

Document Version

Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

Citation for published version (APA):

Supakkul, S., Zhao, L., & Chung, L. (2016). GOMA: Supporting Big Data Analytics with a Goal-Oriented Approach. In *Proceedings - 2016 IEEE International Congress on Big Data, BigData Congress 2016* IEEE. <https://doi.org/10.1109/BigDataCongress.2016.26>

Published in:

Proceedings - 2016 IEEE International Congress on Big Data, BigData Congress 2016

Citing this paper

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.



GOMA: Supporting Big Data Analytics with a Goal-Oriented Approach

Sam Supakkul
Travel Network Architecture
Sabre, USA
sam.supakkul@sabre.com

Liping Zhao
School of Computer Science
University of Manchester, UK
liping.zhao@manchester.ac.uk

Lawrence Chung
Dept. of Computer Science
Univ. of Texas at Dallas, USA
chung@utdallas.edu

Abstract—The real value of Big Data lies in its hidden insights, but the current focus of the Big Data community is on the technologies for mining insights from massive data, rather than the data itself. The biggest challenge facing industries is not how to identify the right data, but instead, it is how to use insights obtained from Big Data to improve the business. To address this challenge, we propose GOMA, a goal-oriented modeling approach to Big Data analytics. Powered by Big Data insights, GOMA uses a goal-oriented approach to capture business goals, reason about business situations, and guide decision-making processes. GOMA provides a systematic approach for integrating two types of the resulting insight from data analytics to goal-oriented reasoning and decision-making processes: descriptive insights are the ones that describe the current state (e.g., the current customer retention rate) and predictive insights are the ones that predict likely future phenomena by inference from the data (e.g., customers who are likely to defect). To aid in the description and illustration of the GOMA approach, a retail banking churning scenario is used as a running example throughout this paper.

Keywords—Goal-oriented modeling; Big Data analytics; business intelligence, business insights; business goals.

I. INTRODUCTION

The real value of Big Data lies in its hidden insights that can be turned into business value to help business decision-making [1]. However, the Big Data community has so far mainly focused on the technologies for mining insights, not on the technologies for helping businesses use the insights. As a result, current analytic tools do not pay enough attention on assessing business goals or mitigating business problems through the insights derived from the data. This problem is recognized in a global industrial survey of nearly 3000 executives, managers and analysts across more than 30 industries and 100 countries [1]. The survey shows that the biggest challenge facing industries is not how to identify the right data, but instead, it is how to use insights obtained from Big Data to improve the business. The survey concludes that industrial leaders need better analytic tools that can provide easier understanding of insights obtained from Big Data and that can facilitate better utilization of insights at all levels of the business. Since the primary measure of success of a software system is the degree to which it meets its user requirements, suffice it to say that current Big Data analytics have failed to live up to the expectations of business. Consequently, time and resources

spent on some data analytics may be wasteful, as the results from current analytic tools may not be useful for the business problem at hand.

To address this challenge, we propose GOMA, a goal-oriented modeling approach [2] to Big Data analytics. Powered by Big Data insights, GOMA uses a goal-oriented approach to capture business goals, reason about business situations, and guide decision-making processes. GOMA provides a systematic approach for integrating two types of the resulting insight from data analytics to goal-oriented reasoning and decision-making processes: descriptive insights are the ones that describe the current state (e.g., the current customer retention rate) and predictive insights are the ones that predict likely future phenomena by inference from the data (e.g., customers who are likely to defect) [3]. To aid in the description and illustration of the GOMA approach, a retail banking churning scenario is used as a running example throughout this paper.

This paper is organized as follows. As a motivation to the GOMA approach, Section II shows the wide applications of goal-oriented approaches. Section III presents the banking churning scenario as a running example for this paper. Section IV then describes and illustrates the GOMA approach. Section V discusses some lessons we have learned from developing and using GOMA. Finally, Section VI concludes the paper with a summary of contributions and future work.

II. MOTIVATION: EVIDENCE OF USING GOAL-ORIENTATION AND BENEFITS THEREIN

In this section, we provide the evidence of using the goal-oriented approach in different areas relevant to our work and describe the benefits from using these approaches.

A. Goal-Oriented Approach for Requirements Engineering

Goal orientation has a long-standing tradition in requirements engineering (RE) [2] and has been recognized to be the most successful and effective approach to performing RE tasks, such as requirements elicitation, analysis, specification, and validation [4][5]. A goal-oriented approach captures stakeholders needs as goals and identifies obstacles that may impede goal achievement; through goal-obstacle analysis, business analysts can make a rationale decision as to which of the goals should be selected to guide business

operations and why the choice is made. The selected goals can help determine business requirements on intended software systems [6][7][8].

The use of stakeholders' goals to provide focus and rationale that drive engineering efforts have motivated and convinced us to pursue a goal-oriented approach to data analytics. In our approach, business goals are refined and mapped onto low-level data analytics requirements, including the desirable insights, data sets and queries.

B. Goal-Oriented Approach for Banking Churn Analysis

In [9], goals represent interested sequences of transactions during data mining, for instance, a sequence of account withdrawal and balance inquiry. This data-level goal-oriented pattern-matching algorithm uses predefined goals or interested items to match transaction patterns. The matched patterns are then analyzed to predict the potential causes for customers to defect.

This application shows that a data-level goal-oriented approach can constrain the search space and direct the search to the specific transaction items. In doing so, this approach has increased the efficiency and accuracy of the search.

Our approach uses business-level goals to constrain the descriptive and predictive data mining and analytics.

C. Goal-Oriented Approach for Healthcare

Goal-oriented approaches can benefit patient care and medical treatment. In [10], a goal-oriented approach is proposed for better patient care. The approach focuses on a patient's individual health goals within or across a variety of aspects and uses these goals to make healthcare decisions, assess outcomes, and measure success. For instance, instead of aiming for the traditional medicine-oriented outcome of disease free survival, one may aim for a goal-oriented outcome of survival until grandchild's wedding.

In another goal-oriented treatment strategy [11], goals provide the decision boundaries for achieving acceptable long-term results in patients with severe illness. Different treatment options (alternative goals) are explored to decide on which one is the best.

Similarly, we adopt the goal-oriented approach to decision-making in business to deal with business goals that are oftentimes conflicting.

D. Goal-Oriented Approach for Web Search

A number of goal-oriented search engines [12][13] approaches use natural language processing to parse a user's search goal and to translate this goal into an effective query to the web.

Google search engine is evolving towards goal-orientation, as we can now use the keywords such as "define" and "translate" to ask Google questions. These examples serve to demonstrate the future direction of web search. Since Big Data analytics are search-based, we believe that

the goal-oriented approach to Big Data is not only highly relevant, but also essential.

III. THE RUNNING EXAMPLE

To illustrate our goal-oriented approach, and also as a worked example, this paper uses and augments a real-world data analytics case study on how a retail bank used data analytics to deal with a checking account customers churning problem [14]. The bank hired a company, specializing in data mining, to help address the churning problem by using insights from detailed transaction data in a newly installed powerful data warehouse. The company hypothesized potential reasons the customers were canceling their accounts, and validated them with descriptive insight mined using a data classification technique. The bank used predictive insights to identify customers at risk of leaving so that incentives can be offered to entice the customers to stay. The consulting company assumed that business goals and problems were well understood and focused mainly only on data analytics aspect. This paper augments the case study to reverse engineer as an empirical study to show how business goals and problems can be systematically analyzed and used in conjunction with data analytics in solving the customer churning problem, which in turn create value by achieving the business goals.

IV. METHOD

When applying GOMA to Big Data analytics, there are four major process steps to be walked through:

- S1: Capture business goals
- S2: Understand business problems
- S3: Identify business actions
- S4: Take business actions

There is a significant amount of iteration over the steps to allow continuous and incremental refinement on the model of each step. The GOMA approach provides a notation technique for representing the GOMA models. The models serve as a cognitive probe for the Big Data analyst to reason about an organization's goals and problems, and make decisions for taking appropriate actions.

In the remaining section, we describe and illustrate the GOMA approach.

S1: Capture Business Goals

The aim of this step is to provide an understanding of the current business goals and their achievement through data analytics. This step involves the following four tasks:

- S1.1: Capture business goals and achievement criteria
- S1.2: Define supporting data sets
- S1.3: Collect and mine data
- S1.4: Assess goal achievement

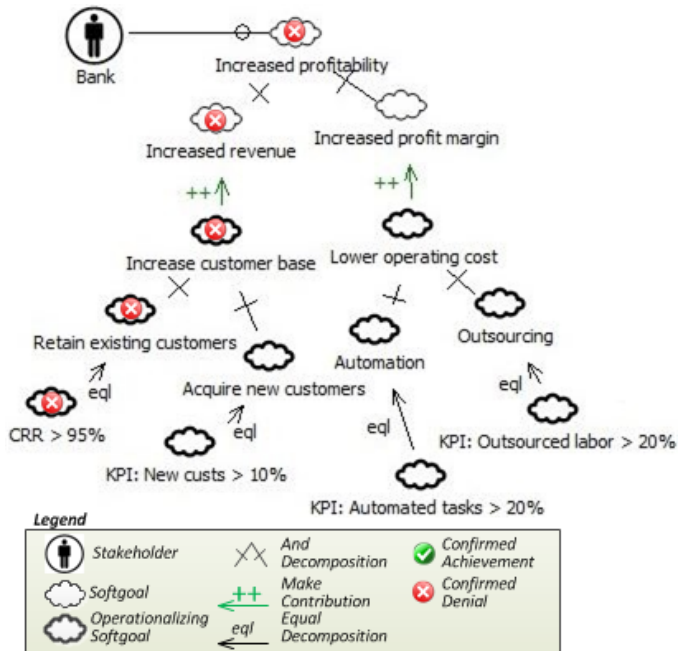


Figure 1. An example of a goal model created by step S1.1 and assessed by step S1.4 and S4.3.

S1.1: Capture Business Goals and Achievement Criteria: In this step, the highest corporate level business goals are captured. Each goal is then refined using AND or OR decomposition to more specific sub-goals until they are operationalizable. Each operationalizable goal is then used to explore operational means, which can be further refined by AND or OR decomposition until each can be validated by a single insight.

Using the bank churning example, Fig. 1 shows a portion of the Softgoal Interdependency Graph (SIG) [8] that represents the relevant business goals. At the highest level, it depicts a corporate level goal of “Increased profitability”, which can be further refined using an AND decomposition to “Increased revenue” and “Increased profit margin” sub-goals, where the former is to be operationalized by “Increase customer base” strategic level goal. “Increase customer base” is then further AND-decomposed to more specific operationalizing goals of “Retain existing customers” and “Acquire new customers”, where the former’s achievement is defined by “CRR > 95%” goal, which represents the desirable Customer Retention Rate (CRR) Key Performance Index (KPI) goal.

S1.2: Define Supporting Data Sets: To assess the business goals are currently sufficiently achieved, we identify the desirable insights obtainable from the data analytics of internal and external data to help judge the achievement of lowest level goals, then use goal-based reasoning to determine whether higher level goals are achieved.

To make it easier for business analysts to specify at high-

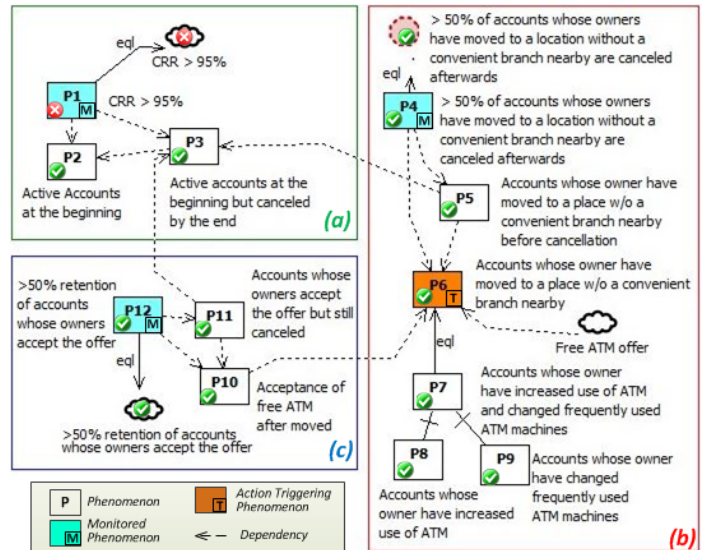


Figure 2. A portion of the Phenomenon Model for the bank churning example. (a) phenomena defined by step S1.2 and used by S1.3, S1.4 and S4.3 for assessing business goals, (b) phenomena defined by step S2.2 and used by S2.3, S2.4, S4.2 and S4.3 for validating business problems and triggering business actions, (c) phenomena defined by step S3.2 and used by S4.2 and S4.3 for evaluating the effectiveness of business actions.

level for data engineers and data scientists the desirable insights and the corresponding data and data analytics, we propose Phenomenon Modeling to define the desirable real-world phenomena and their inter-dependency. Each Phenomenon represents an insight observable through data and data analytics. Accordingly, each Phenomenon includes the following specifications:

- Input - specifications for the required data elements.
- Selection - a function for selecting data elements from Input. If none is specified, all Input data elements are selected by default.
- Output - specifications and/or functions for producing the resulting data elements from the selected Input data elements.
- Validation - a predicate to determine whether this Phenomenon is observed to be valid through the data and data analytics.

Using Fig. 2 as an example, the desirable insight for assessing “CRR > 95%” KPI goal is represented by P1, which is the ratio between the number of active accounts at the beginning of a reporting period and the number of the same accounts that remain at the end of the period [15]. As the result, P1 depends on phenomena P2 and P3, which represent the active accounts at the beginning and the same active accounts that are canceled by the end of the reporting period respectively.

Phenomena P1, P2 and P3 may be specified as follows, where “P_n” represents a data set produced by a Phenomenon and “{...}” represents a set of the data elements.

Phenomenon P1

Description:
Customer retention is higher than > 95%.

Input:
P2, P3

Output:
CRR(P3,P2)

Validation:
CRR(P3,P2) > 0.95

Function:
CRR({canceled},{baseline}):
1 - count(canceled)/count(baseline)

Phenomenon P2

Description:
Active accounts at the beginning of a reporting period.

Input:
{ Account.id as acct_id,
interval(Account.open_date,Account.close_date) as A },
interval(ReportingPeriod.begin_date,
ReportingPeriod.end_date) as R

Output:
{ acct_id, A }

Selection:
not (R older A) and not (R after A)

Validation:
count(Output) > 0

Phenomenon P3

Description:
Active accounts at the beginning of a reporting period
that are canceled by the end of the period.

Input:
P2 as A,
interval(ReportingPeriod.begin_date,
ReportingPeriod.end_date) as R

Output:
A

Selection:
R survives A or R tail-to-tail A

Validation:
count(Input) > 0

Fig. 3 shows possible accounts selected by the Selection criteria in P2 and P3 respectively. Note in Fig. 2 and P3’s specification that P3’s Input depends on the Output from P2, which results in the need for P3 having to deal with a smaller data set pre-selected and pre-processed by P2. Therefore, P3’s Selection criterion does not need to repeat the Selection criterion specified in P2. In effect, the data analytics performed for P2 and P3 are chained in a pipelining fashion for reusability of data sets to achieve higher performance, which is important when dealing with a large volume of data at the Big Data scale.

Fig. 3 depicts the temporal relationship expression “not (R older A) and not (R after A)” specified in Phenomenon P2 for selecting accounts that are active at the beginning of a reporting period (e.g, 1Jan15-31Dec15), as well as the temporal relationship expression “R survives A or R tail-to-tail A” specified in Phenomenon P3 for selecting the resulting accounts from P2 that are canceled by the end of the reporting period.

“older, after, survives and tail-to-tail” are example of 13 and 11 temporal relationship expressions defined by James Allen [16] and Christian Freksa [17] respectively. We adopt these temporal relationships to more intuitively represent temporal relationships than the typical low-level technology-

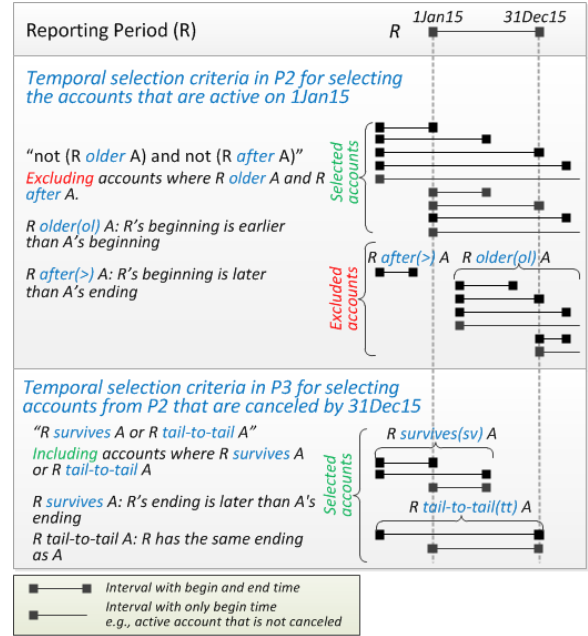


Figure 3. Temporal selection criteria in P2 and P3, and the resulting selected accounts

dependent queries such as “acct.opendate is not null and acct.opendate < reportperiod.begin” SQL query.

SI.3: Collect and Mine Data: In this step, data is collected, cleansed, prepared and mined as specified by the Phenomenon specifications, such as P1, P2, and P3, to obtain the desirable insights. The Validation predicates are used to validate or invalidate the corresponding Phenomena, which in turn confirm whether the lowest level goals are achieved.

Referring to the banking example, if Phenomenon P1’s Validation predicate returns false to represent that the CRR is not higher than 95%, P1 would be invalidated, as depicted by a circled cross mark in Fig. 2, which in turn would confirm that the “CRR > 95%” goal is denied as depicted by a circled cross mark as well.

SI.4: Assess Goal Achievement: Once it is confirmed that the lowest level goals have been achieved or not achieved, the achievement status is evaluated and propagated up the goal hierarchy based on their inter-dependency and the Label Propagation Procedure that defines goal achievement evaluation rules [8]. The evaluation and label propagation are repeated upward the goal graph to the highest level corporate goals to confirm and reflect how well the organization goals are achieved based on the insights obtained from the data.

Using Fig. 1 as an example, the lowest level goal of “CRR > 95%” is confirmed to be denied by the data analytics modeled in Fig. 2. Through the eql relationship, the parent goal (“Retain existing customers”) is also denied. The goal denial is propagated to “Increase customer base” goal since one of the sub-goals from the AND-decomposition is denied. By repeating the achievement evaluation from other peer

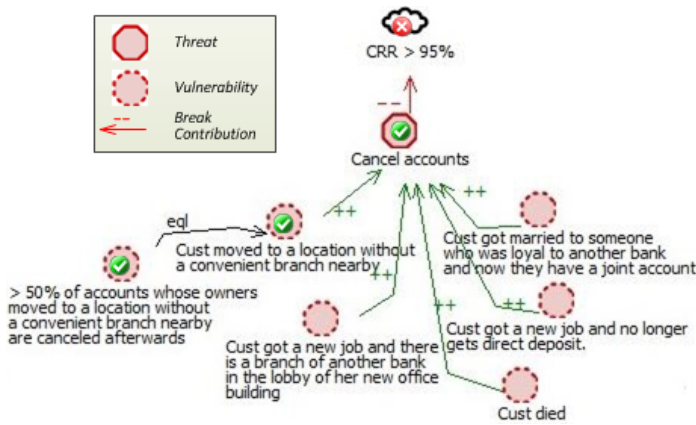


Figure 4. An example of a problem model defined by step S2.1 and validated by S2.4.

and parent goals, the highest goal (“Increase profitability”) is considered denied even without other goals being confirmed since the label propagation for AND-decomposition requires only one sub-goal to be denied to consider that the corresponding parent goal is also denied.

S2: Understand Business Problems

We need to understand the business problems that are obstacles to achieving business goals so that appropriate business actions can be implemented to mitigate them. To understand business problems, we follow the following tasks:

- S2.1: Identify potential problems and root causes
- S2.2: Define validation criteria and supporting data sets
- S2.3: Collect and mine data
- S2.4: Validate problems

S2.1: Identify Potential Problems and Root Causes: In this step, we identify potential problems that are obstacles to achieving each lowest level business goal. Similar to goal refinement, each problem may be further refined by AND or OR decomposition to more specific problems. The refinement may repeat until problems are specific enough that specific threats and contributing vulnerabilities can be identified until each lowest level threat or vulnerability is specific enough that a mitigating business action can be identified [18].

Using Fig. 4 as an example of Problem Interdependency Graph (PIG) [18], “CRR > 95%” could be denied by “Cancel accounts” problem, which could be caused by a number of reasons, including “Cust moved to a place without a convenient branch nearby” and “Cust got a new job and there is branch of another bank in the lobby of her new office building”, among other possible causes, as shown in the diagram.

S2.2: Define Validation Criteria and Supporting Data Sets: The identified potential causes are only hypotheses

that need to be validated by data and their analytics. In this step, we identify the criteria and supporting data sets needed to perform data analytics to validate the root causes.

Using the banking example, we identify “> 50% of accounts whose owners moved to a location without a convenient branch nearby are canceled afterwards” as a validation criterion for “Cust moved to a place without a convenient branch nearby”, as shown in Fig. 4.

Next, we identify Phenomena P4-P9 to represent the desirable insights and data set, as shown in Fig. 4, where P4 represents KPI “> 50% of accounts whose owners moved to a location without a convenient branch nearby are canceled afterwards”. P4 depends on P5 and P6 that represent “Accounts whose owner have moved to a place without a convenient branch nearby before cancellation” and “Accounts whose owner have moved to a place without a convenient branch nearby” respectively. In the same fashion as the specification for P1, P4 is specified to validate “count(P5)/count(P6) > 0.50”.

However, many customers may not notify the bank that they have moved to a location without a convenient branch nearby. It is therefore oftentimes necessary to derive some phenomena like this from other phenomena. In this case, the bank hypothesized that customers who have moved to a location without a convenient branch nearby tend to increase the use of ATM where the locations of frequently used ATM locations have also changed, the Phenomena captured as P8 and P9 respectively, as shown in Fig. 2. P8 and P9 are related through an AND-decomposition to P7 that represents the conjunction of the two Phenomena. Once P7 is validated by the data analytics, it is propagated to validate P6 through the *eql* relationship that represents the phenomenon derivation. Due to space limitation, only the specification for P8 is given below as an example.

Phenomenon P8

Description:
Accounts whose owner have increased the use of ATM during a reporting period.

Input:
{ ATM_Transaction.acct_id as acct_id,
ATM_Transaction.trans_date as trans_date,
ATM_Transaction.branch_id as branch_id,
interval(ReportingPeriod.begin_date,
ReportingPeriod.end_date) as R }

Output:
FirstDayATMUsageIncreases(Input)

Validation:
count(Output) > 0

Function:
FirstDayATMUsageIncreases({acct_id, trans_date,
branch_id, period}) returns {acct_id, trans_date}

Description: A regression function that analyzes ATM transactions usage patterns to detect a list of accounts and the first day whose owners are likely to have increased the use of ATM.

S2.1: Collect and Mine Data: In this step, the data and analytics specified for the Phenomena are collected, cleansed and mined to obtain the desirable specified insights.

Using Fig. 2 as an example, P4-P9 are performed in

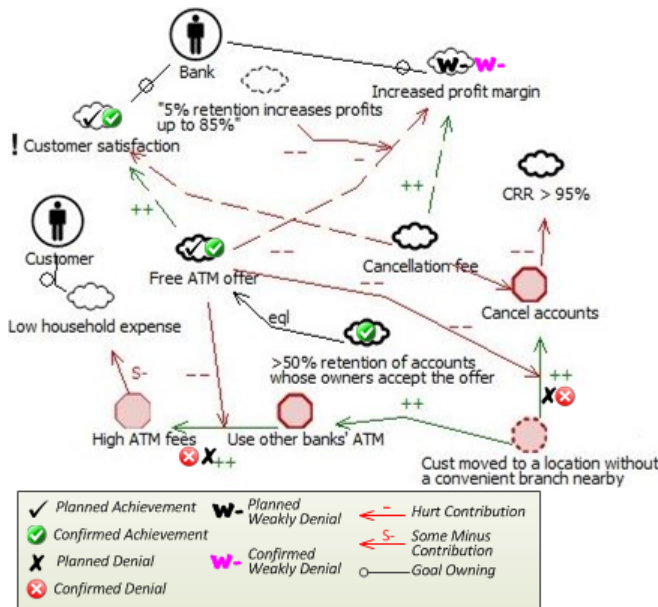


Figure 5. Problems from the customer’s perspective identified by step S2.1. Alternative solutions (“Free ATM offer” and “Cancellation fee”) identified by S3.1 and selected by the trade-off analysis in S3.2, carried out in S4.2, and assessed for effectiveness in S4.3.

the order of P8, P9, P7, P6, P5, then P4 based on the AND-decomposition and inter-Phenomena dependencies. If P4 is validated to confirm that more than 50% of accounts whose owners who have moved also canceled the accounts afterwards, as depicted by a green circled check mark, it also validates the hypothesis “>50% of accounts whose owners have moved to a location without a convenient branch nearby are canceled afterwards” problem that is associated with P4 via an *eql* relationship .

S2.2: Validate Problems: Once the lowest level problems or root causes have been confirmed to be validated or invalidated, the results are evaluated and propagated upward the problem graph (PIG) using the same Label Propagation Procedure [8] to validate the problems identified to be the direct obstacles to the goal achievement.

Using Fig. 4 as an example, the lowest level problem of “> 50% of accounts whose...” is confirmed by the data analytics, the result is propagated upward to confirm the “Cust moved to a location...” problem, which is in turn propagated to validate “Cancel accounts” problem, which is then in turn confirm the denial of “CRR > 95%” goal as the result.

S3: Identify Business Actions

This step is carried by the following tasks:

- S3.1: Identify alternative business actions as solutions to the validated problems
- S3.2: Select business actions and define success criteria
- S3.3: Define monitored and triggering points

S3.1: Identify Alternative Business Actions as Solutions to the Validated Problems: To mitigate the validated problems, we identify alternative options to mitigate the problems and/or their root causes. Using Fig. 5 as an example, two options, including providing “Free ATM offer” and imposing “Cancellation fee” are identified as alternative solutions, where the former option is for mitigating “Use other banks’ ATM” problem that leads to “High ATM fee” problem, while the latter option is used to discourage customers from canceling accounts in the first place. At the same time, we also capture the correlations of the contributions the two options make towards other goals. For example, “Free ATM offer” has a positive correlation (Make) towards “Customer satisfaction” and a negative correlation (Break) towards “Increased profit margin” goals respectively. On the contrary, imposing “Cancellation fee” has a negative correlation (Break) towards “Customer satisfaction” and a positive contribution (Make) towards “Increased profit margin” goal.

S3.2: Select Business Actions and Define Success Criteria: While multiple options can have synergistic positive contributions towards multiple goals, it is oftentimes the case that their contributions towards the goals can be conflicting as described earlier. For this reason, a trade-off analysis is performed based on the priority of all relevant goals to select options that provide the best compromises among the available options [8][19]. Once desirable business actions are selected, we identify operational goals to assess the effectiveness of the actions using data analytics.

In Fig. 5, “Free ATM offer” is selected over “Cancellation fee” as a better business action even though it hurts the profit margin goal because study shows that reducing defection by just 5% can increase profits by 25% to 85% [20]. To evaluate the expected impacts to the business goals, we explicitly mark “Free ATM offer” with *Planned Achievement* label depicted by a black and white check mark. Using the label Propagation Procedure [8], the *Planned Achievement* is propagated to indicate that “Customer satisfaction” and “Increased profit margin” goals are expected to be achieved and weakly denied respectively, as depicted by the corresponding black and white check mark and “W-” label.

To determine the action’s effectiveness, we define the goal of having “>50% retention of accounts whose owners accept the offer” as an operational goal. Since the operational goal for effectiveness evaluation is defined after the trade-off analysis, it does not have a *Planned Achievement* label as it is not needed for the trade-off analysis.

S3.3: Define Monitored and Triggering Points: An effective way to mitigating problems is by prevention or deterrence, which means the corresponding business actions must be carried out before the problems are materialized. This requires the need to plan for business environment monitoring and the conditions to trigger the pre-defined business actions. Using our approach to take advantage of data analytics, we identify the Phenomena that would

produce the insights for monitoring and triggering business actions, as well as measuring their effectiveness.

Using Fig. 2 as an example, we identify P1, P4 and P12 as monitored Phenomena that represent the relevant KPIs, and P6 as a triggering point that is depended upon by “Free ATM offer” business action against the accounts produced by P6 as Output. The monitored and triggering point Phenomena along with other dependent Phenomena define the data sets that are needed to obtain the required insights and KPIs.

S4: Take Business Actions

Once the business actions have been rolled out, we carry out the following tasks to take the business actions:

- S4.1: Periodically collect and mine data
- S4.2: Take Business Actions at the Triggering Points
- S4.3: Assess business actions’ effectiveness

S4.1: Periodically Collect and Mine Data: To monitor and take business actions at the right time, the pre-defined monitoring and trigger points in the Phenomena model are periodically validated, which requires the corresponding data sets be collected and mined to perform the associated data analytics periodically, for instance at a regular interval determined to be desirable and practical for the organization.

Using Fig. 2 as an example, we periodically collect and perform data analytics for Phenomena P1, P4, P6 and P12 as defined in step S3.3.

S4.2: Take Business Actions at the Triggering Points: During the monitoring interval, we take pre-defined business actions against the data sets produced by the action triggering Phenomena. For example, on a weekly basis, a list of accounts produced by P6 that represent the accounts whose owners are likely to defect after moving to a location without a convenient branch nearby. The owners of those accounts are then contacted with the free ATM offer for a period of time to entice them to stay.

S4.3: Assess Business Actions’ Effectiveness: As an ongoing basis, we monitor all of the pre-defined KPIs to assess the effectiveness of the business actions in mitigating the business problems. If the actions are effective as planned, the monitored KPIs should also reflect how well the corresponding denied business goals are being improved and ultimately whether they become achieved.

Continuing with our example, the customers who have accepted the free ATM offer will be collected as Input to P11 that represents accounts whose owners have moved and accepted the free ATM offer.

When the validations of the Phenomena are propagated to the goals in the goal model, we would be able to track the actual goal achievements against planned. In Fig. 5, the actual achievement or denial would be depicted by colored labels, e.g., the color check and cross marks, that can be depicted side-by-side with the planned achievements and denials depicted by black and white check and cross marks.

V. DISCUSSION

GOMA meets several of the recommendations identified by the survey [1] to address some of challenges of turning Big Data into business value. To have insights that are easy for business users to understand, GOMA recommends presenting insights in the form of KPIs, the practice that is well known in the the business community, as well as other forms such as lists of business objects such as customers. To have insights that are closely linked to business strategy, GOMA links analytics to corporate and strategic goals via the operational KPI goals. To embed insights into organizational processes where actions can be taken at the right time, GOMA uses Phenomenon Model to represent desirable insights, where some are identified as triggering points that are monitored to trigger business actions at the right time. The ultimate value to the business is the ability to clearly and systematically model and use data analytics at the Big Data scale (e.g., via the Phenomena pipelining) to mitigate business problems and in turn achieve business goals.

We have learned some important lessons from applying GOMA to the banking example during the 4-step process. In the business goals capturing step (S1), we found that goal-oriented modeling made systematic identification of important and useful KPIs much easier.

Existing literature on KPIs (e.g., [21][15]) provides good motivation and process description of using KPIs. Yet, it lacks detailed guidelines on how to systematically identify KPIs in the context of business goals and problem resolutions, and incorporating them into the business process with data analytics. This paper fills this literature gap.

In the business problem understanding (S2) and business actions identification steps (S3), we found that goal and problem modeling from the perspectives of multiple relevant stakeholders can help lead to more suitable business actions. For example, if we had tried to more strictly understand the problems and identify actions only from the bank’s perspective, imposing cancellation fee would have appeared to be a desirable business action to take since it could mitigate the churning problem and increase revenue. However, by trying to understand the account cancellation action, which is a problem from the bank’s perspective, that it is in fact a solution from the customer’s perspective to reduce the household expenses. From that perspective, it was clear that imposing cancellation fee would not meet the customer’s goal of reducing household expenses and the bank’s goal of keeping the customer satisfied. This more comprehensive multiple perspectives led to a better goal-oriented trade-off analysis that led to a more suitable business action of offering free ATM fee option.

In the business action taking step (S4), we found that it is important to identify triggering points and associate them with data analytics and business actions. This seems to provide a better focus and directions for data analytics.

On the contrary, many existing data analytics projects tend to focus on by-chance undirected knowledge discovery efforts where some analytical efforts may not produce any insights or some resulting insights, while interesting, may not be directly useful or actionable to the business.

The goal and problem models in Fig. 1, 4 and 5 were created using RE-Tools [22]. In addition to supporting the visual modeling of goals and problems, the tool also provides two reasoning features: automated goal achievement evaluation and automated trade-off analysis. The first automation automates GOMA steps S1.4, S2.4 and S4.3 based on rules defined in the Label Propagation Procedure [8]. The second automation automates GOMA step S3.2. that recommends the most desirable alternatives that have the highest score from a weight-based selection function taking into account all positive and negative correlations [19].

The tool is being extended to support the Phenomenon Modeling (GOMA steps S1.2, S2.2 and S3.3), with a plan to generate data analytics specification (e.g., PMML [23]), and other useful artifacts, such as ready-to-use code modules for a Hadoop Map-Reduce environment that are chained to form the pipelines according to the inter-Phenomena dependency defined in the Phenomenon model (e.g, Fig. 2).

VI. CONCLUSION

This paper has proposed GOMA, a goal-oriented approach to Big Data modeling and analytics. GOMA has a 4-step process, including 1) capturing business goals and achievement, 2) hypothesizing and validating business problems, 3) exploring and identifying business actions and success criteria, and 4) taking actions and measuring results. GOMA provides a novel approach to confirming and re-confirming goal achievement using data analytics. More specifically, the Phenomenon Modeling is a new modeling technique that bridges between business goal modeling and data analytics. All in all, GOMA addresses the problem of turning insights into business value shared by organizations at all data analytics maturity levels [1]. Much work remains in refining this approach, including more case studies to further validate the benefits, as well as tool support for modeling, specification and code generation. Notwithstanding these issues, we feel that this paper has given one initial basis for linking a number of important concepts for Big Data to be used beneficially.

REFERENCES

- [1] S. LaValle, E. Lesser, R. Shockley, M. S. Hopkins, and N. Kruschwitz, "Big data, analytics and the path from insights to value," *MIT sloan management review*, vol. 21, 2013.
- [2] K. Yue, "What does it mean to say that a specification is complete?" in *Proc. IWSSD-4, Fourth International Workshop on Software Specification and Design*, 1987.
- [3] J. Han, M. Kamber, and J. Pei, *Data mining: concepts and techniques*. Elsevier, 2011.
- [4] J. Mylopoulos, L. Chung, and E. Yu, "From object-oriented to goal-oriented requirements analysis," *Communications of the ACM*, vol. 42, no. 1, pp. 31–37, 1999.
- [5] A. van Lamsweerde, "Goal-oriented requirements engineering: A guided tour," in *Proc. 5th Intl. Symp. Requirements Engineering*, 2001, pp. 249–262.
- [6] —, *Requirements Engineering: From System Goals to UML Models to Software Specifications*. Wiley, 2009.
- [7] E. Yu and J. Mylopoulos, "Understanding why in software process modelling, analysis, and design," in *Proc. 16th Int'l Conf. on Software Engineering*, 1994, pp. 159–168.
- [8] L. Chung, B. A. Nixon, E. Yu, and J. Mylopoulos, *Non-Functional Requirements in Software Engineering*. Kluwer Academic Publishers, 2000.
- [9] D.-A. Chiang, Y.-F. Wang, S.-L. Lee, and C.-J. Lin, "Goal-oriented sequential pattern for network banking churn analysis," *Expert Systems with Applications*, vol. 25, no. 3, pp. 293–302, 2003.
- [10] D. B. Reuben and M. E. Tinetti, "Goal-oriented patient carean alternative health outcomes paradigm," *New England Journal of Medicine*, vol. 366, no. 9, pp. 777–779, 2012.
- [11] M. Hoepfer, I. Markevych, E. Spiekerkoetter, T. Welte, and J. Niedermeyer, "Goal-oriented treatment and combination therapy for pulmonary arterial hypertension," *European Respiratory Journal*, vol. 26, no. 5, pp. 858–863, 2005.
- [12] H. Liu, H. Lieberman, and T. Selker, "Goose: a goal-oriented search engine with commonsense," in *Adaptive Hypermedia and Adaptive Web-Based Systems*. Springer, 2002.
- [13] A. Faaborg and H. Lieberman, "A goal-oriented web browser," in *Proc. of the SIGCHI conference on Human Factors in computing systems*. ACM, 2006, pp. 751–760.
- [14] M. J. Berry and G. Linoff, *Data mining techniques: for marketing, sales, and customer support*. John Wiley & Sons, Inc., 1997.
- [15] B. Marr, *Key Performance Indicators (KPI): The 75 measures every manager needs to know (Financial Times Series)*. FT Press, 2012.
- [16] J. F. Allen, "Maintaining knowledge about temporal intervals," *Communications of the ACM*, vol. 26, no. 11, pp. 832–843, 1983.
- [17] C. Freksa, "Temporal reasoning based on semi-intervals," *Artificial intelligence*, vol. 54, no. 1, pp. 199–227, 1992.
- [18] S. Supakkul and L. Chung, "Extending problem frames to deal with stakeholder problems: An agent- and goal-oriented approach," in *Proc. of the 2009 ACM Symposium on Applied Computing (SAC '09)*, 2009, pp. 389–394.
- [19] S. Supakkul, T. Hill, L. Chung, T. Tun, and J. Leite, "An NFR pattern approach to dealing with NFRs," in *18th IEEE Intl. Requirements Engineering Conf.*, 2010, pp. 179–188.
- [20] F. P. Reichheld and W. E. Sasser, "Zero defection: Quality comes to services," *Harvard business review*, vol. 68, no. 5, pp. 105–111, 1990.
- [21] D. Parmenter, *Key performance indicators: developing, implementing, and using winning KPIs*. John Wiley & Sons, 2015.
- [22] S. Supakkul and L. Chung, "The RE-Tools: A Multi-notational Requirements Modeling Toolkit," in *Requirements Engineering Conference (RE), 2012 20th IEEE International*, Sept 2012, pp. 333–334.
- [23] R. Pechter, "What's PMML and What's New in PMML 4.0?" *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 19–25, 2009.