



A PLM-based data analytics approach for improving product development lead time in an engineer-to-order manufacturing firm

Kai Sun*, Yunpeng Li, Utpal Roy

Department of Aerospace and Mechanical Engineering, Syracuse University, Syracuse, 13244, US

Email: kasun@syr.edu

ABSTRACT

A critical challenge to Engineer-To-Order (ETO) manufacturing firms is the long lead time in product development, due to the nature of highly customized products produced in low volume. The emergent Data-Driven Design (D³) approach is a possible solution to alleviate this tension by utilizing rich product-related data throughout the product's lifecycle and building reusable data analytics models, in order for more accurate lead time estimate in a continuously updated way. In this paper, we propose a product lifecycle management (PLM) based framework to enable Prescriptive Analytics capability in ETO companies to improve their lead time estimate. The PLM has been extended to support data analytics model lifecycle management by incorporating a formal data analytics process model, CRISP-DM. It also implements a model-based approach for optimization model development, management, and execution. The proposed PLM-based Predictive Analytics framework has been applied in a local ETO manufacturing company for testing and validation.

Keywords: CRISP-DM, Engineer-to-order (ETO), Model-based Optimization, Prescriptive Analytics, Product Lifecycle Management (PLM).

1. INTRODUCTION

Engineer-To-Order (ETO) companies design and produce products that are generally highly customized to meet individual customer requirements but produced in low volume [1]. High levels of customization lead to increased costs, higher risks, and long lead times. Many ETO companies have recognized these difficulties, and are trying to increase design standardization based upon modular design principles, termed Mass Customization (MC) [2]. MC simplifies the manufacturing process, thus lowering costs of the manufactured products. It also facilitates automation of some internal processes, and encourages more customer involvement, hence speeds the delivery times. In many cases, however, this approach has been proven challenging. ETO products are often hard to standardize; product configuration involves extensive knowledge base design. Further, the predefined product solution space limits the options for customers, which should have been originally a competitive advantage for an ETO company.

The customer delivery performance, a key competitive factor in ETO markets, is based on lead time estimates. The lead time is usually estimated without sufficient information on available capacity as there are always several quotations awaiting responses from potential customers. Detailed specifications that determine work contents and durations are also uncertain at this early stage. So, it needs to be reevaluated after the order being accepted. Improving the delivery performance relies on two factors: reducing the lead time and increasing the reliability of lead time estimates. Lead time reduction can be achieved by shortening the duration of individual processes and by encouraging parallel activities of the processes through concurrent engineering. Improving the estimate accuracy, on the other side, requires decision makers and the relevant decision support systems to take advantage of the data collected from all product-, production-, and servicerelated activities and turn them into meaningful information.

Product Lifecycle Management (PLM) with advanced data analytics technologies provides the necessary concepts and tools to facilitate these two aspects [3]. PLM includes a shared platform for creating, managing, and disseminating productrelated information across the extended enterprise [4]. It also embeds capabilities to access, use, and maintain product definition information, as well as the business processes related to all lifecycle activities. Data analytics is intended to answer the questions of "what has happened" (descriptive analytics), "what could happen" (predictive analytics), and "what should we do" (prescriptive analytics), using statistical and machine learning techniques. Data analytics provides the techniques needed to increase the degree of data-driven decision making.

One of our earlier studies demonstrated the benefit of extending traditional PLM functions to support data analytics model development which is similar to a physical product development process [3]. A standard analytics model representation language, PMML (Predictive Model Markup Language), and a standard process model, CRISP-DM (CRoss-Industry Standard Process for Data Mining), are integrated into a PLM system, so that metadata and lifecycle information of analytics models can be fully managed. We focused on Predictive Analytics in that study. In this paper, we report an ongoing effort to enable Prescriptive Analytics in PLM in order to provide manufacturing firms data-driven decision-making capabilities. We apply the proposed methodology to the engineering department of a local ETO company to improve its product delivery performance by improving its project resource (particularly, the engineers) planning efficiency.

2. CHALLENGES

2.1 Prescriptive analytics in manufacturing

Prescriptive analytics has been referred to as the final frontier of analytic capabilities not only to predict future outcomes, but also to make recommendations based on those outcomes. Prescriptive analytics typically involves decision optimization techniques, such as mathematical and constraint programming [5].

There are two critical challenges for conducting data analytics project in a manufacturing firm. The first challenge comes from the data variety when considering the relevant data related to a product's full lifecycle. The product-, production-, and service-related data are available in various manufacturing information systems (PLM, MES, and ERP) [6], even the data may reside in external supply chain partners' systems. The second challenge is the inability of big data processing and fusion due to limitations of IT resources in a manufacturing firm [7]. Few manufacturing experts are familiar with the modern big data analytics techniques which naturally ask for interdisciplinary skills. The information from the upstream (sales and project planning) and the downstream (production and quality assurance) of the engineering processes may not accessible by the engineering department.

In addition, the current manufacturing practice is that every analytics task is often implemented from scratch, following a task-focused approach due to the trial-and-error nature. The sharing and combination of analysis resulting from isolated methods and terminologies on local datasets become significantly limited [8]. This leads to high-cost and longduration development, and results in models and algorithms that are difficult to modify, extend, and reuse [9]. There is an additional challenge particularly to ETO companies. The product produced by ETO companies varies significantly, so do their customer demands and available capability (engineers, plants, equipment, etc.). This variability makes it difficult to prescribe best practice(s) for optimal market responses [1].

The intention of prescriptive analytics is to make algorithms self-adaptive as guided by minimal rules for given data sets. The algorithms will be programmed in such a way that they can take over and adapt based on changes in established parameters, instead of humans controlling them. With algorithms optimizing automatically, their ability to predict the future becomes better with time. This implies the decision optimization modeling and execution shall be embedded in the overall product development process and continuously incorporate the newly available data.

2.2 Prescriptive analytics in PLM

Our earlier study shows the feasibility of uniform modeling of physical products and analytics models [3]. Though this previous work primarily focused on predictive models, the idea can be generalized for other types of analytics models (e.g. descriptive models and prescriptive models). Indeed, a descriptive model can be immediately seen as a predictive model that makes perfect predictions [10]. The CRISP-DM process model is generic enough to provide the lifecycle contexts to all kinds of analytics models. The task and activity details naturally vary from project to project, and we will discuss (in section 3) how it could be utilized for prescriptive analytics modeling.

However, for the prescriptive model meta data, extra efforts are needed. The PMML employed in [3] supports popular predictive models such as classification, regression, clustering, and association rules. While it can capture the information of data transformation, model structure, and model performance, it does not support the goal and constraints presented in an optimization model. We need a structured language to encode prescriptive models.

Brodsky et al. [5] proposed to use JSON-based structured language to describe an optimization model. It combines the features from several optimization modelling languages: AMPL (A Modeling Language for Mathematical Programming), GAMS (General Algebraic Modeling System), and OPL (Optimization Programming Language). Microsoft provides a model-based optimization package which is open source and termed MSF (Microsoft Solver Foundation) [11]. It uses a proprietary optimization modeling language, OML. These languages define optimization-specific objects: parameters, decisions, constraints, and goals. These schemas and the relevant optimization model execution engine need to be appropriately integrated into our PLM-based architecture.

3. PLM-BASED PRESCRIPTIVE ANALYTICS



Figure 1. An itegrated concept development process model

The PLM-based prescriptive analytics framework is shown in Figure 1. It consists of three tiers: 1) a PLM platform with core product lifecycle management functions and built-in computational methods for the execution of analytics models; 2) a formal data analytics process model to guide the datadriven model formulation, evaluation and deployment; and 3) an application tier for applying prescriptive analytics techniques to engineering problems, for instance, engineer task assignment in a product development process.

3.1 PLM platform tier

In [3], we employed an object-oriented, web-based PLM system as part of a service-oriented architecture (SOA), in order to provide a unified development environment for physical components and predictive analytics models. In addition, we extend the PLM core functions (metadata and relationship management, workflow management, version control and change management, item structure and configuration management, etc.) to include two additional functions: 1) Visualization - visualizating of data, models, and their relationships; and 2) Analytics model execution - providing engines to interpret analytics models and apply them to new data.

The PLM system is, essentially, a data repository. As such, it stores all the available instance data and instance models. It also stores all the necessary lifecycle information such as states and revisions of the data and the models. The PLM system uses the concepts *Item* and *Relationship* to abstract arbitrary objects and connections between objects. This *Item-Relationship-Item* architecture has been proven to be suitable in capturing an analytics model and its associated information. The PLM system is able to return information regarding any physical component model, dataset, analytics models, and any of their compositions, corresponding to different levels of queries requested. It also provides computational engines to retrieve and execute the models stored in the PLM database.

3.2 Data analytics process model tier

The CRISP-DM is the leading reference model for data analytics projects. It provides the necessary lifecycle context for analytics model development [8]. CRISP-DM defines six phases to complete a data analytics project and each phase further defines several key generic tasks and outputs. The six phases are: business understanding, data understanding, data preparation, model building, model evaluation, and model deployment. The tasks and outputs of each stage are described in detail by Shearer [12]. Prescriptive analytics has its own characteristics compared to other types of analytics. For instance, the goal and constraints to the decision making must be identified; and an optimization model usually consists of other types of models (e.g., a descriptive model for a particular variable with certain distribution). Finally, the optimization results can be encoded as decision trees or business rules so that they can be deployed in the production environment. Hence, it requires appropriate adaptation of the CRISP-DM model.

3.2.1 Business understanding and data understanding

Data may come from various sources (e.g. patents, bill-ofmaterials, engineering drawings, simulation, machinery, sales, and ERP data, etc.) that are related to products, processes, and the organization. The relationship among these data need to be recognized before constructing the decision models. The DSM (Design Structure Matrix) method is capable of representing both static and temporal information as directed graphs. A variety of analytical techniques (sequencing, clustering, banding, and tearing) are available to analyze information presented in DSMs [13], [14]. Accordingly, DMM (Domain Mapping Matrix) is to connect DSMs representing different domains in order to model and visualize interactions across domains [15]. The DSM/DMM analyses are extensively used in the *Business Understanding* and *Data Understanding* phases (Figure 1) for identifying the relationship between the various information domains. Each analysis involves the following steps: 1) decomposing the system into elements (e.g. decompose the resources in the engineering department into engineering managers, project managers, quality engineers, drafters, document and manual specialists, etc.); 2) understanding and documenting the interactions between the elements (i.e., the collaboration of the engineers); and 3) analyzing potential reintegration of the elements via clustering (integration analysis) [13], [16]. Qualitative and quantitative measures can be employed when filling the matrices.

3.2.2 Data preparation and modeling

With variables and their relationships identified in the previous two phases, the data can be aggregated and descriptive/predictive models can be formulated to capture the system characteristics. Then, the optimization model is formulated to capture the system goal and constraints. As proposed in our earlier study [17], a model-driven approach is employed to formulate a data-driven optimization model that begins with a set of data concerning the domain of interest. The relevant variables, attributes, and rules that govern the model are identified through data analytics. Scenarios are dynamically classified and new elements of the system are continuously incorporated. The simulation and optimization models simulate and evaluate the performance of different scenarios; the optimization engine is used for clustering, feature selection, scenario classification, and the what-if analyses. This data-driven simulation and optimization model provides a practical approach to discover the system and evolve the system. With a PLM platform, the data comes from the PLM database; the optimization model, the intermediate models used by the optimization model, and the results of the optimization process can be stored back to the PLM database.



Figure 2. A data-driven simulation/optimization model (adpated from [17])

3.2.3 Evaluation and deployment

The evaluation phase is a quality assurance step to the datadriven optimization model, which can be seen as a data product produced by the data analytics process. The previous steps to create the model need to be thoroughly reviewed to ensure that the model achieves the business objectives.

The final phase is to deploy the optimization model into the production environment. The usual way is to extract a decision tree or a set of business rules to encode the optimal results according to classification of scenarios. This turn to deploy a set of predictive models, which we have discussed in the earlier study. Another way is to directly deploy the optimization model so that it can be periodically run on production data, to dynamically update the decisions.

The PLM platform consists of computational engines so that the optimization model can be executed as demanded. The change of the optimization model can be versioned, and so does the optimization results. Hence, the model selection and deployment can be done by comparing different versions of the results using a Champion-Challenge model that is usually employed in business analytics community [18].

3.3 Application tier

The top application tier provides users the interface to formulate the problem, configure the data processing and modelling strategies, and visualize the results. For a resource planning problem concerned by an engineering manager (or project manager in product development), the prescribed decisions are assigned tasks to individual engineers, given project task requirements, the availability of each engineer and his/her capability to do a particular job, plus other necessary information. The essential inputs to initialize the problem include the project scope, goal statements, critical constraints and parameters, and the specific performance measures.

The application is able to send/receive information from the PLM system at different lifecycle phases. The optimization model can be created based on modification to an early model stored in PLM or can be totally new. The application can be embedded in the PLM platform or runs independently. This is because the analytics models can either be (1) executed inside the PLM system using PLM built-in execution engines and methods, or (2) retrieved and utilized from an external application through web services or PLM's application programming interfaces (API) [3].

4. CASE STUDY

In this section, we present a real case study that applies the proposed framework in this paper to a local company.

4.1 Problem description

This local manufacturing firm produces industrial liquid filtration systems and waste disposal equipments. The company employs an Engineer-To-Order (ETO) business model to develop a customized product system for each individual customer as a project. Each project has a long lead time, and there are usually tens of ongoing projects. The company also offers its customers a preventative maintenance program to insure long, efficient operational life of the equipment performance in service. The challenge for the engineering manager is how to improve the development process so that the productivity of each engineer is maximized and the lead time of each project can be minimized, while maintaining the product/service quality. The company is using information systems to manage the sales order, project budget estimation, project planning, and customer information management. However, the engineering team's activities are not properly synchronized with these systems, i.e. the product data management is an isolated silo from other business systems.

4.2 Data collection and data analysis

We collected the on-site data in the company, and interviewed the engineers, engineering manager and business owners. The documents and information collected include: project quality management plan, engineering meeting agenda, engineer weekly time sheets, project plans with project memo, drawings and bill of materials. The data is cleaned, transformed, and stored in the PLM system.

Advanced data analytics and visualization techniques are applied to the data in order to reveal the patterns and bottlenecks of the as-is situation. For instance, DSM/DMM techniques are employed to investigate the dependencies among project tasks, products, and engineer resources. We develop several measures to quantify the engineer skills, project task priorities, and product module importance weights as the qualitative and quantitative measures for the visualization. The results (shown in Figure. 3) illustrate that the working loads are heavily concentrated on several engineers, and the existing product modules are not evenly reused.

The as-is decision making process in the company as interviewed is recorded in a DMN (Decision Model and Notation) model (see Figure. 4). The primary data sources include historical project data, engineer weekly time-sheets and meeting review, and new project requirements. Three analytics models - the engineer capacity determination, engineer availability determination, and the project task estimation - could be established based upon these data. The final task assignment decision is then based on the three analytics models. It is also constrained by several corporate policies: the project quality management plan guide and the task classification code.



Figure 3. Product-project-resource visualization in a sankey diagram

4.3 Optimization modeling



Figure 4. A project-task-assignment decision making model in DMN



Figure 5. A network max-flow formulation of the problem



Figure 6. XML Data file for optimization input

The tasks in a project are usually precedence-constrained (or time-bounded) and resource-constrained. Hence, it turns into a resource planning problem that can be reduced to a network max-flow problem (or equally, min-cut problem) with virtual source and target nodes [19]. And a network max-flow problem can be solved using Linear Programming algorithm such as Simplex. The identified variables, constraints, and the utility function to minimize the overall project costs, can be encoded using an optimization modelling language. We use the Microsoft OML in this study.

The data from the PLM system first needs to be translated into an XML (Extensible Markup Language) format to feed to the optimization engine as model inputs. This input XML specifies the activities (Figure. 6). Each task has an earlieststart-time and latest-end-time. The tasks can be scheduled in between those times. Each task specifies an amount of time working on the task. Each task has an active-time (the amount of time dedicated to a task) and a passive-time (the time that the task can be run automatically). Other tasks can be scheduled while passive activities are occurring. The user specifies the maximum number of times a task can be interrupted (set as 0 in Figure. 6). Thus, the active time can be broken up as long as the total time doing that task is achieved. The goal is to end the day as early as possible with all the activities completed. So that the total working time of all the engineers can be minimized.

The model is then encoded as OML that can be parsed by an optimization engine (we use Microsoft Solver Foundation in the project) and executed based on the given data.

4.4 Model evaluation and deployment

We compare the result generated by the optimization model to the schedules (historical working hours in the engineer timesheets) created by human beings of the company to validate the model.

The model can be deployed in the PLM system and it is executed as demanded. The associated data can be retrieved based on the relationship (e.g. the Application-Product-Task-Deliverable relationship). The relationship between the involved elements can be visualized in real-time and be updated dynamically. The final application is embedded in the PLM system (Figure. 7).



Figure 7. The final application in the PLM system

5. CONCLUSION

In this paper, we focus on a critical challenge, long lead time in product development to ETO manufacturing firms. To address this problem, we propose a PLM-based framework to enable Prescriptive Analytics capability in ETO companies to reduce the lead time and improve the lead time commitment. Also, the PLM system is extended to support a model-based approach for optimization model development, management, and execution. The proposed PLM-based Predictive Analytics framework has been applied in a local manufacturing company for testing and validation. And we have proved that the emergent data-driven design approach is a possible solution to alleviate ETO companies' pressures on more accurate lead time estimate in a continuously updated manner, by utilizing rich data collected from the product's lifecycle and building reusable data analytics models.

Using the data analytics process model, the data is transformed and stored in an open source PLM platform. With the domain-specific information collected through the application tier, the optimization problem has been formulated and the data analytics plan has been executed and deployed. With the product and application, project and task, engineer data stored in the PLM database, the built-in optimization engine (Microsoft Solver Foundation) and data analytics engine compute the optimal engineering task schedule and offer visualization of the results (as shown in Figure. 3). The application (Figure. 1) can be internal or external to the PLM platform. And finally, the data analytics model can help in improving the delivery performance.

The current work has its limitations. For example, lack of the measurements of engineer capabilities, and insufficient estimation of different kind of tasks' importance. The optimization results can be improved with more accurate estimation of these factors, but efforts are needed from both the research group and the industrial partners. For the future work, we wish to enable the data analytics processes to become part of the day-to-day business and its environment. The efforts on development and utilization of prescriptive analytics models as described in this paper, along with the description of predictive analytics models in the PLM platform, will invariably improve the future data analytics application.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support of the CASE Center at Syracuse University, Syracuse COE AM-EC Internship program and the collaboration of a local company in carrying out this work.

REFERENCES

- Hicks C., McGovern T., Earl C.F. (2000). Supply chain management: a strategic issue in engineer to order manufacturing, *International Journal Production Economics*, Vol. 65, No. 2, pp. 179-190. DOI: 10.1016/S0925-5273(99)00026-2
- [2] Haug A., Ladeby K., Edwards K. (2009). From engineerto-order to mass customization, *Management Research News*, Vol. 32, No. 7, pp. 633–644. DOI: <u>10.1108/01409170910965233</u>
- [3] Li Y., Roy U., Shin S., Lee Y.T. (2015). A "smart component" data model in PLM, 2015 IEEE International Conference on Big Data, pp. 1388-1397. DOI: <u>10.1109/BigData.2015.7363899</u>
- [4] Corallo A., Latino M.E., Lazoi M., Lettera S., Marra M., Verardi S. (2013). Defining product lifecycle management: a journey across features, definitions, and concepts, *International Scholarly Research Notices Industrial Engineering*, Vol. 2013. DOI: <u>10.1155/2013/170812</u>
- [5] Brodsky A., Shao G., Krishnamoorthy M., Narayanan A., Menascé D., Ak R. (2015). Analysis and optimization in smart manufacturing based on a reusable knowledge base for process performance models, 2015 IEEE International Conference on Big Data, Santa Clara, CA, USA, pp. 1418-1427. DOI: 10.1109/BigData.2015.7363902
- [6] Roy U., Zhu B., Li Y., Zhang H., Yaman O. (2014). Mining big data in manufacturing: requirement analysis, tools and techniques, ASME 2014 International Mechanical Engineering Congress and Exposition, Montreal, pp. V011T14A047. DOI: 10.1115/IMECE2014-38822
- [7] Brisley A. Oracle, Agile engineering collaboration overview, from <u>http://tcmoaug.communities.oaug.org/multisites/tcmoaug/ media/Documents/Presentations/2007/2007-12-11/ABMPSL AppsUnplugged Engineering Collaboratio <u>n.pdf</u>, accessed on 21 Jan. 2017.</u>
- [8] Groger C., Niedermann F., Schwarz H., Mitschang B. (2012). Supporting manufacturing design by analytics,

2012 IEEE 16th International Conference on Computer Supported Cooperative Work in Design, Wuhan, China, pp. 793-799. DOI: <u>10.1109/CSCWD.2012.6221911</u>

- [9] Elliott T. (2017). Business analytics and digital business, gartner BI: analytics moves to the core, from <u>http://timoelliott.com/blog/2013/02/gartnerbi-emea-</u>2013-part-1-analytics-moves-to-the-core.html, accessed on 21 Jan. 2017.
- [10] Dhar V. (2013). Data science and prediction, Communications of the Association for Computing Machinery, Vol. 56, No. 12, pp. 64-73. DOI: <u>10.1145/2500499</u>
- [11] Gearhart J.L., Adair K.L., Detry R.J., Durfee J.D., Jones K.A., Martin N. (2013). Comparison of open-source linear programming solvers, Sandia National Laboratories, Albuquerque, New Mexico, USA, Sandia Report SAND2013-8847.
- [12] Shearer C. (2000). The CRISP-DM model: the new blueprint for data mining, *Journal of Data Warehousing*, Vol. 5, No. 4, pp. 13-22.
- [13] Sharman D., Yassine A. (2004). Characterizing complex product architectures, *Systems Engineering*, Vol. 7, No. 1, pp. 35–60. DOI: <u>10.1002/sys.10056</u>
- [14] Eppinger S.D., Browning T.R. (2012). *Design Structure Matrix Methods and Applications*, The MIT Press, Cambridge, MA, USA.
- [15] Danilovic M., Browning T.R. (2007). Managing complex product development projects with design structure matrices and domain mapping matrices, *International Journal of Project Management*, Vol. 25, No. 3, pp. 300-314. DOI: <u>10.1016/j.ijproman.2006.11.003</u>
- [16] Pimmler T.U., Eppinger S.D. (1994). Integration analysis of product decompositions, ASME Design Theory and Methodology Conference 1994, Minneapolis, MN, USA, pp. 343-351. DOI: <u>10.1.1.207.8509</u>
- [17] Li Y., Roy U. (2015). Challenges in developing a computational platform to integrate data analytics with simulation-based optimization, ASME 2015 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Boston, MA, USA, pp. V01BT02A035. DOI: 10.1115/DETC2015-46410
- [18] Keyworth M. (2008). Identifying the optimum collections strategy through the use of champion challenger analytical tools, Experian, Bristol, UK, White Paper.
- [19] Möhring R.H., Schulz A.S., Stork F., Uetz M. (2003). Solving project scheduling problems by minimum cut computations, *Journal of Management Science*, Vol. 49, No. 3, pp. 330-350. DOI: <u>10.1145/1839379.1839419</u>