Proceedings of the 2021 Winter Simulation Conference
S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo, and M. Loper, eds.

FARMING FOR MINING: COMBINING DATA FARMING AND DATA MINING TO GAIN KNOWLEDGE IN SUPPLY CHAINS

Joachim Hunker

Department IT in Production and Logistics
TU Dortmund University
Leonhard-Euler-Straße 5
Dortmund, 44227, GERMANY

ABSTRACT

Knowledge discovery in databases (KDD) is a frequently used method in the context of supply chains (SC). The core phase of KDD is known under the term data mining. To gain knowledge, e.g., to support decisions in supply chain management (SCM), input data for KDD are necessary. Typically, such data consist of observational data, which have to be preprocessed before the data mining phase. Besides relying on observational data, simulation can be used to generate data as an input for the process of knowledge discovery. The process of using a simulation model as a data generator is called data farming. To link both data farming and KDD, a Farming-for-Mining-Framework has been developed, where the data farming process generates data as an input for the KDD process to support decisions in SCM.

1 INTRODUCTION

Due to trends such as cost pressure, decentralization, and digital transformation, SCs are nowadays complex systems in a globalized world. Mastering this complexity is an essential task for SCM. To support decisions in SCM, gaining and visualizing knowledge is one of the key factors. KDD is understood as a process model, which includes different phases, ranging from data collection to the visualization and interpretation of results (Fayyad et al. 1996). The core phase to extract implicit, previously unknown and useful patterns which represent knowledge is known under the term data mining (Witten et al. 2011). To extract useful and valid patterns, preprocessed input data suitable for the applied data mining technique are necessary. Typically, data used for KDD are observational or ”real” data, containing different flaws, for example missing values. One possible way to address this problem is the use of synthetic data. The process of using simulation to generate synthetic data is called data farming (Brandstein and Horne 1998). Data farming uses a simulaton model as a data generator in combination with a targeted design of experiments and high performance computing (Horne and Meyer 2005).

The focus of this research lies in the purposeful combination of both data farming and KDD in the context of logistical tasks in SCs. Therefore, a Farming-for-Mining-Framework has been developed, which supports SCM decision-makers in decision-making situations. This contribution emphasizes the impact of well-designed data farming experiments specifically tailored towards the application of data mining techniques.

2 A FARMING-FOR-MINING-FRAMEWORK

When using real data, a data basis for KDD contains data that are not needed or cannot be used in raw form for a specific data mining algorithm. To create a suitable data basis, data has to preprocessed according to the specific requirements of the used algorithm for data mining, such as the properties data type and data type range. Since in our approach the input data for the KDD process are synthetic, data farming should

be aligned with the requirements of the specific algorithm used for the data mining part of KDD. Figure 1 shows the proposed framework, displaying only the relevant phases for the paper.

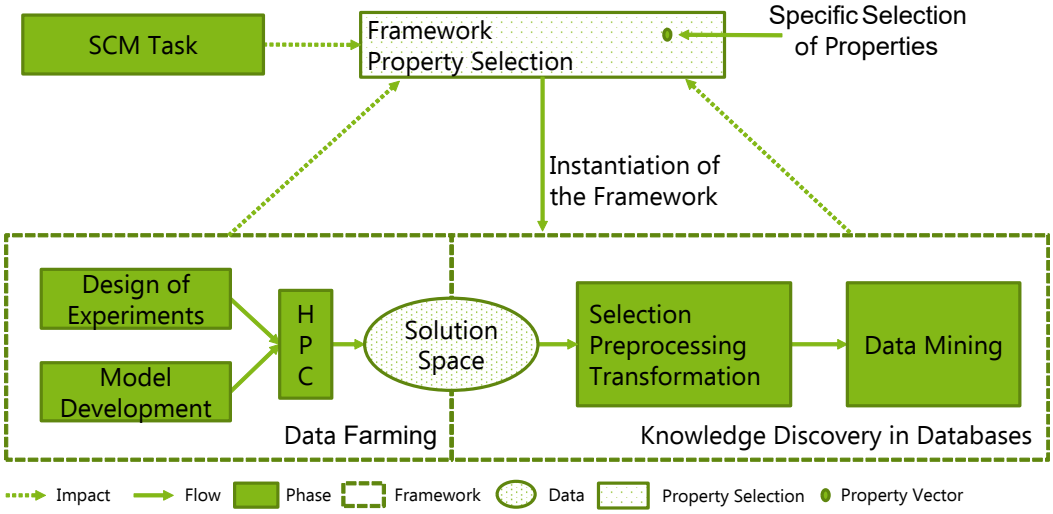

Figure 1: Relevant phases of the Farming-for-Mining-Framework.

Data farming and KDD are combined in an all-inclusive framework. It includes features from both domains, such as the design of experiments from data farming as well as the data preparation from KDD. The key element, which instantiates the framework, is the Framework-Property-Selection. This includes specific tasks or decision situations in SCM which need to be supported. Furthermore, the derivation of requirements resulting from the SCM tasks as well as from the used data mining methods are taken into account to instantiate the framework. The specific set of properties is considered while building the simulation model and the design of experiments.

The Farming-for-Mining-Framwork has been used in first experiments, where it is applied to generate and analyze a synthetical data basis to support a decision-making situation for a logistics task of SCM.

3 CONCLUSION, LIMITATIONS, AND OUTLOOK

The Farming-for-Mining-Framework for the purposeful combination of data farming and data mining takes into account the multilayered complexity requirements regarding various different properties of data mining techniques. It includes the specific requirements for a farming-for-mining-study proved beneficial in first experiments, which have been conducted. For further research, the framework needs to be extended to more complex tasks and decision-making situations of SCM. This includes the derivation of further development steps for the Farming-for-Mining-Framework as well as validating the practical relevance and credibility.

REFERENCES

Brandstein, A. G., and G. E. Horne. 1998. “Data Farming: A Meta-technique for Research in the 21st Century”. *Maneuver Warfare Science*:93–99.

Fayyad, U. M., G. Piatetsky-Shapiro, and P. Smyth. 1996. “From Data Mining to Knowledge Discovery in Databases”. *AI Magazine* 17(3):37–54.

Horne, G. E., and T. E. Meyer. 2005. “Data Farming: Discovering Surprise”. In *Proceedings of the 2005 Winter Simulation Conference (WSC)*, edited by M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joins, 1082–1087. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Witten, I. H., E. Frank, and M. A. Hall. 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. 3 ed. Morgan Kaufmann Series in Data Management Systems. Burlington: Morgan Kaufmann.