Human-Machine Teaming for Intelligent Demand Planning

By Ye Ma
Advisor: Dr. Maria Jesus Saenz Gil De Gomez

Topic Areas: Forecasting, Supply Chain Planning, Artificial Intelligence

Summary: Today collaboration is switching from just among humans to between humans and machines. This study empirically analyzed the effects on forecast accuracy and inventory level of applying different human-machine teaming decision-making structures in demand adjustment processes. The research found that hybrid human-machine teaming models with adequate human intervention provided the optimal performance of forecast and inventory improvement, especially for short-term forecast accuracy of low-turnover products.

Before coming to MIT, Ye Ma worked as an e-commerce and supply chain IT manager for P&G, Greater China. Ye holds a Bachelor of Science degree with a major in Software and Computing Systems from the University of Melbourne, Australia.

KEY INSIGHTS

1. The human-machine teaming capabilities improve the demand forecast accuracy and inventory level significantly.

2. Hybrid human-machine decision-making structure with moderate human revision have the optimal performance compared to AI delegation or human overrides groups.

3. Hybrid models have much better forecast accuracy in low-turnover products than high-turnover ones, which is good for long-tail product handling.

Introduction

Demand planning is the first master planning task that defines the operation plan, which is a crucial part of supply chain management, where human knowledge particularly matters. Demand forecasting is also strongly connected with inventory management due to its impact on the replenishment schedules, production arrangements, and delivery plans. However, it has been characterized by heavily manual work and ineffective information system handling. Nowadays, the complex demand-planning process includes not only the overall information integration and interpersonal communications, but also the IT systems interaction and technologies.

Machine and artificial intelligence (AI) are reshaping the way we work, do business, and collaborate. Collaboration is switching from just among humans to between humans and machines. Mundane and repetitive tasks will be done by machines automatically, while humans can develop insights and make wise decisions supported by data streaming from intelligent machines. In supply chain management, the human-machine teaming capabilities could contribute to a more accurate demand forecast result, so that the company could further improve customer-specific order service levels, inventory efficiency, turnover rates, and even customer sales.

Human-machine teaming models could determine whether varied AI system capabilities and implementation would be successful or not. However, there is little investigation on the AI-Human teaming decision-making structures in the supply chain application, specifically in demand signal selection and adjustment, to formulate effective demand forecasting processes. On the other hand, there is plenty of research about the AI and machine learning models developed for forecast accuracy improvement. After a system-embedded AI engine is implemented, the organizational decision-making structures engaged with humans and AI algorithms would influence an organization's performance significantly. This requires that professional managers clearly understand the strengths and weaknesses of human-AI teaming capabilities and features.

This study focuses on examining the different human-machine teaming decision-making structure and their effects on demand forecast accuracy and inventory level. This thesis seeks to answer the following questions:

•If and how could different human-machine teaming decision-making structures improve demand forecast accuracy and inventory level?

•Which of the structures would provide an optimal approach for demand forecasting and inventory level: Full AI delegation, or hybrid (AI-to-Human) models with different levels of human intervention?

Data and Methodology

The data used by this research is provided by a large-scale fast-moving consumer goods company, and all the data comes from their **intelligent demand adjustment (IDA) system**. The data contains demand forecasting-related data from both the supplier internally and their key e-commerce customer externally. Those datasets include demand forecast data, actual order shipment data, and system auto-generated accuracy results related to the demand planning. Each entry in the record represents a line item of a demand forecast SKU during a specific week for a specific distribution center. The total datasets contain over 6 million entries, but this research only focuses on the quarterly performance report with SKU-level analysis.

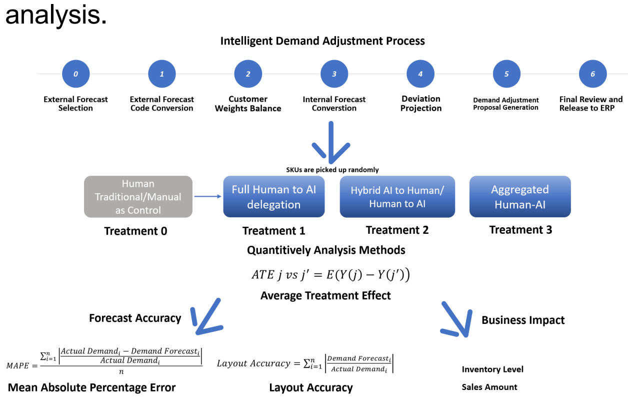

Figure 1. Experiment Analysis Structure

This research's empirical objective is to determine the causal effect of different human-AI teaming decision-making structures on forecast accuracy and their business impact. Figure 1 shows the overall experiment structure, beginning with the demand planning human-machine interaction process, followed by the treatments and control groups, then described the quantitative analysis, and finally the outcome variable measurements. All the human-machine interactions in the demand adjustment processes are through the IDA system. As shown in Figure 1, there are 7 steps of the human-machine decision and interaction points: 0: external forecast selection; 1: external forecast code conversion; 2: customer weights balance; 3: internal forecast conversion; 4: deviation projection; 5: demand adjustment and proposal generation; and 6: final review and release to ERP.

The treatments are for randomly selected SKUs under different human-AI decision-making demand forecast models, in which humans and machines interacts in the IDA in different models. These are labeled as follows: **traditional manual process group** as a control group **as Treatment 0**; **full machine delegation group as Treatment 1**; and AI-Human group **Hybrid with moderate human revision as Treatment 2**, and **Hybrid with all steps human overrides as Treatment 3**, respectively (Figure 1). All treatments are in binary form.

The traditional manual process group is treated as the baseline control group (Treatment 0); demand planners would not consider the external customer signals and would only take the average from the internal national-level forecast as the customer demand. The full machine delegation group means that the whole demand adjustment process is automatically done by AI (Treatment 1), and humans are not involved in the process. In Treatment 2, demand planners would engage in the main planning adjustment steps 2 and 4, with adequate human intervention. In the Treatment 3, humans would judgmentally adjust in all the demand adjustment processes, as overrides.

The outcomes include two parts: **forecast accuracy**—mean absolute error percentage (MAPE) as the short-term forecast accuracy measurement, and **layout accuracy**, which is defined by the demand planner, to measure the long-term demand forecast accuracy; and **business impacts** – customer inventory amount and sales volume to end consumers.

We use an augmented inverse propensity weight (AIPW) estimator to find the mean of potential outcomes for every treatment level with potential outcome distribution, which is the average treatment effects between treatment and control groups. In AIPW, we specify a regression treatment to calculate the estimated generalized propensity score, and then specify a regression outcome model for the conditional mean outcomes of every treatment level. To fulfill the conditional independence assumption, many pretreatment variables are selected to control the potential influence of SKU-specific demand forecast features. Based on the product information, this research selected the product segmentation, turnover rate, price, and product category as the pretreatment variables. The turnover rate of the pretreatment variable is also selected as the moderator variable to further study the influence of product turnover for human-machine teaming capabilities. We use the Stata multivalued effect function to calculate the ATE results.

Results and Discussion

First, the average treatment effects among treatment groups (Treatment 1, 2, 3) compared with control group (Treatment 0) are shown in Figure 2. Our findings show that adopting all human-machine teaming decision-making structures in the demand-planning process significantly improves both the forecast accuracy and the inventory level, compared with the traditional manual control group. All treatment groups significantly improve their long-term forecast accuracy error (53-56%), short-term forecast accuracy error (MAPE 50-64%), and business results (inventory level 47-70%). This finding answers our first research question: According to the results of this experiment, human-machine decision-making structures do improve both the demand forecast accuracy and the inventory level.

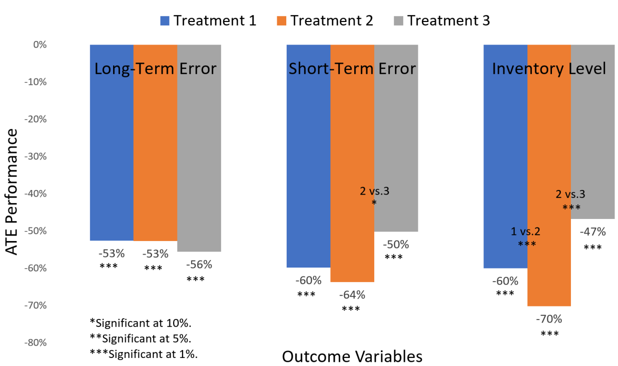

Figure 2. The overall ATEs comparison among treatments

Second, Figure 2 also shows that average treatment effects (ATEs) of inventory level are varied in different human-machine teaming decision-making structures, which depend on the level of human intervention and which planning steps humans get involved in. There is an obvious pattern that adequate human intervention (Treatment 2) in the process would improve the inventory level compared to that of the full AI delegation group (Treatment 1). In Treatment 2, demand planners would only adjust the AI-provided results according to their expertise and updated promotion information or warehouse information, in two main steps: 2: customer weights balance; 4: deviation projection between internal, and external customer forecast. If there are too many human overrides in all demand adjustment processes (Treatment 3), it would drastically reduce the human-machine teaming advantages, leading to an worse improvement in the forecast error (47%), compared with Treatment 2 (70%) and Treatment 1 (60%).

Third, the findings illustrate that overall (according to Figure 2), the optimal human-machine teaming decision-making structure for different outcome variables varies, but Treatment 2 performs very well for all outcome variables, which answers our second research question.

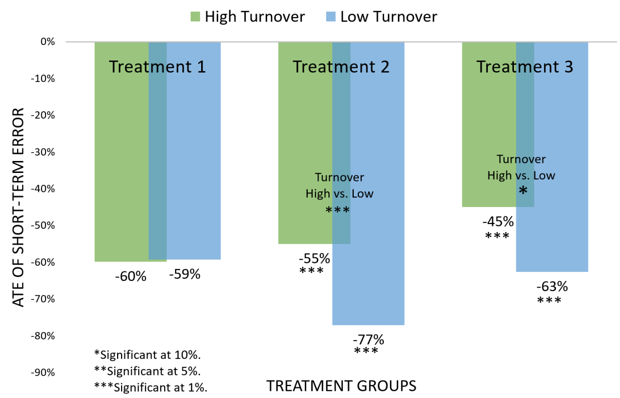

Figure 3. The ATEs of short-term forecast by turnover

Finally, according to Figure 3, the moderating analysis on the impacts of product turnover shows the human-machine teaming decision-making structures work better to improve the short-term forecast for low-turnover products, resulting in a 22% improvement when compared low-turnover product and high-turnover products accuracy.

Conclusions

This research shows that the human-machine teaming decision-making structures implemented in demand planning processes improve the organizational performance of both demand-forecast accuracy and inventory level by least 47%. Overall, the Hybrid AI-Human with adequate human intervention model is the optimal decision-making structure between human and machine, which improves the short-term forecast accuracy by 53%, long-term forecast accuracy by 64%, and inventory level by 70%. The Hybrid AI-Human model with intervention in all steps performed worse than the other models. Additionally, the AI-enabled decision-making structures work better for low-turnover products than for high-turnover ones.