Proceedings of the 2024 Winter Simulation Conference
H. Lam, E. Azar, D. Batur, S. Gao, W. Xie, S. R. Hunter, and M. D. Rossetti, eds.

CAUSAL-BASED RACK LAYOUT OPTIMIZATION IN RETAIL: INCORPORATING AGENT-BASED MODELING AND CAUSAL DISCOVERY

Shuang Chang[1], Shohei Yamane[1], and Koji Maruhashi[1]

[1]Fujitsu Research, 4-1-1 Kamikodanaka, Nakahara-ku, Kawasaki, JAPAN

ABSTRACT

Rack layout design and optimization is a main research problem in the realm of retail management. To optimize the rack layouts considering customers' preferences and routing behaviors against different layouts, agent-based models (ABM) have been developed and applied. However, applying conventional analysis methods to analyze the model may not be sufficient to proactively propose and evaluate explainable layout patterns that optimize pre-defined metrics. In this work, we extend a causal-based ABM analysis method to enable a causal understanding of ABM models across multiple scenarios and incorporate it to a real-world data calibrated model that simulates the customers' in-store traffic for rack layout optimization. By elucidating the causal relations and changes among customers' preferences, movements, and layout patterns, we demonstrated that the incorporated model and analysis method can enable causal explanation empowered rack layout optimization for improving the customer experience and store revenue, compared with conventional ABM analysis methods.

1 INTRODUCTION

Rack layout design and optimization is a main research problem in the realm of retail management, since prior studies have confirmed that layout design plays a pivotal role in shaping the perception and movements of customers, influences the exposure of products, and consequently has a critical impact on the revenue of retailers. Several metrics have been identified to quantify the impact of rack layout design, such as the time spent at shops to complete the purchase, customers' satisfaction level, exposure of products, and so on (Lu and Seo 2015). How to efficiently and effectively optimize the rack layout against those measures is therefore a critical yet challenging problem.

Various approaches have been proposed and developed to tackle layout optimization problems as how to arrange product categories on racks. In operational research, optimization methods have been widely deployed to quantify the effect of different layouts on the target metrics and to find the optimal allocation of facilities or products (Drira et al. 2007). In contrast, microscopic system modeling methods, such as agent-based modeling (ABM), can account for the heterogeneous customers' perceptions and movements against different layouts (DeAngelis and Diaz 2019). Such modeling methods are useful for inductively identifying optimized rack layouts considering customers' characteristics, yet the identified layouts often lack explanation of how and why they can optimize the target metrics. Therefore, extra efforts are required to proactively propose meaningful layout patterns, which is important for communicating with domain experts.

To strengthen the explanation power of ABM models, a variety of analysis methods which aim at understanding the model behaviors have been proposed (Lee et al. 2015). Apart from the traditional quantitative and qualitative analysis methods, there has been a trend of applying causal discovery methods to elucidate the causal paths among model components and based on which to enable a causal explanation of the model behaviors. However, regarding the rack layout optimization in retail management, it is not straightforward to apply the causal analysis for generating meaningful patterns that can optimize the layouts.

An iterative discovery process considering the customers' movements against different layouts and the causal changes across different layouts are necessary to enable the proactive proposal of optimized patterns.

To this end, incorporating ABM models and causal-based analysis methods is vital to support the rack layout optimization in terms of enabling the proposal and evaluation of explainable product arrangement patterns. To our best knowledge, such integration and application in retail management is still lacking, which forms the motivation of this work.

1.1 Related Works

In operational research, how to arrange the product categories on racks is formalized as a facility layout problem (FLP). Various methods have been developed to quantify the effects of different layouts on target metrics and based on which to find the optimal allocation of facilities (Drira et al. 2007). For example, optimization methods have been widely adopted to support the layout decision-making, aiming at maximizing a variety of metrics such as the exposure of products (Mowrey et al. 2018), stores' revenue (Yapicioglu and Smith 2012), impulse purchase (Flamand et al. 2016), and so on. This stream of research mainly focuses on optimizing the aggregated metrics, yet the customers' dynamic movements are rarely considered.

In contrast, agent-based modeling (ABM), a micro-level system modeling method, is gaining popularity. It can enhance the decision-making of retail stores with a refined understanding of the impact of various layout design by considering the preferences and interactions of heterogeneous customers. For example, the customers' intentions and routing behaviors against different spatial rack layouts can be accounted for arranging the products in racks (Schenk et al. 2007), and the flow and density of customers' in-store traffics resulted by different layouts can be modeled and explored using the empirical data collected by in-store technologies (Pantano et al. 2021). However, applying agent-based modeling alone may not be sufficient to effectively inform explainable layout patterns that can optimize target objectives. Special analysis methods of ABM therefore should be developed and incorporated to better interpret the models for rack layout design and optimization.

Several streams of ABM analysis methods have been proposed to understand the model behaviors and to strengthen the explanation power of ABM models. Machine learning techniques have been increasingly applied to evaluate the impact of input parameters on the model output (Lee et al. 2015). For example, decision trees and Bayesian networks have been applied to process both quantitative and qualitative data for deriving behavioral rules of agents (Drchal et al. 2019; Abdulkareem et al. 2019). From the perspective of micro-behaviors of agents, analysis methods such as micro-dynamic analysis using clustering methods have been proposed to examine the model behaviors under a single scenario or across different scenarios (Yamane et al. 2018; Yamada et al. 2020).

Moreover, to enable a causal explanation of the model behaviors, analysis methods leveraging causal discovery (Pearl 2009) which aim at discovering the causal relations among model components have been proposed. Similar to randomized experiments but without ethical constraints, ABM can overcome the *fundamental problem of causal inference* (Holland 1986) by simulating counter-factual outcomes and investigating the causal effect of policy interventions. A methodological framework incorporating causal discovery methods was proposed to provide a causal explanation of the simulated emergent phenomenon on the basis of the model inputs (Janssen et al. 2019). Pattern mining algorithms and causal discovery methods were integrated to inform policies targeting at specific group to support group-specific policy making and evaluation (Chang et al. 2022). A systematic causal-based analysis method for ABM was developed to identify the causal paths among model components and on the basis of which to inform indirect-control policies (Chang et al. 2023). However, it is not trivial to extend and customize the methods in retail management to generate and optimize meaningful product arrangement on racks.

Limitations. While the aforementioned causal-based analysis methods have proven effective in supporting decision-making, identifying the causal changes across multiple scenarios is still not considered and handled yet. This is particularly useful for enhancing the layout optimization by explaining the causal changes of customers' actions across different rack arrangement. Furthermore, it is not straightforward to integrate such causal analysis methods into the agent-based model for deriving explainable patterns that may optimize the layouts. Extra efforts are thus necessary to incorporate the model and analysis methods for generating causal insights to support the layout optimization.

Contribution. To this end, we aim to enable a causal-based optimization of product category arrangement on racks by innovatively incorporating ABM and a causal-based analysis method. This integration is applied to derive and explain layout patterns that optimize the target metrics on the basis of elucidating the causal relations and changes among customers' preferences, actions and rack layouts.

We first review and extend a causal-based analysis method (Chang et al. 2023) to multiple scenarios for discovering the causal structure changes in Section 2. We introduce an agent-based model that simulates the customers' in-store traffic against various layouts in a virtual store in Section 3. The model is calibrated and validated using real data collected from a real supermarket. We then incorporate the revised analysis method to the model in an iterative way to demonstrate its applicability in rack layout optimization considering the customers' preferences in Section 4. We conclude with a discussion of future works in Section 5.

2 METHOD

We briefly review a causal analysis method of ABM (Chang et al. 2023), which outperforms the conventional ABM analysis methods in policy-making, and extend it to multiple scenarios. This analysis method is designed to systematically discover the causal paths among model components for guiding the policy formulation, and to evaluate the proposed policies at macro-level against aggregated metrics and at micro-level from a causal perspective. It consists of three stages, namely **problem identification**, **policy formulation** and **policy evaluation**, to facilitate the policy-making on the basis of a causal understanding of the ABM model, as illustrated in Figure 1 . More details can be referred in the original paper.

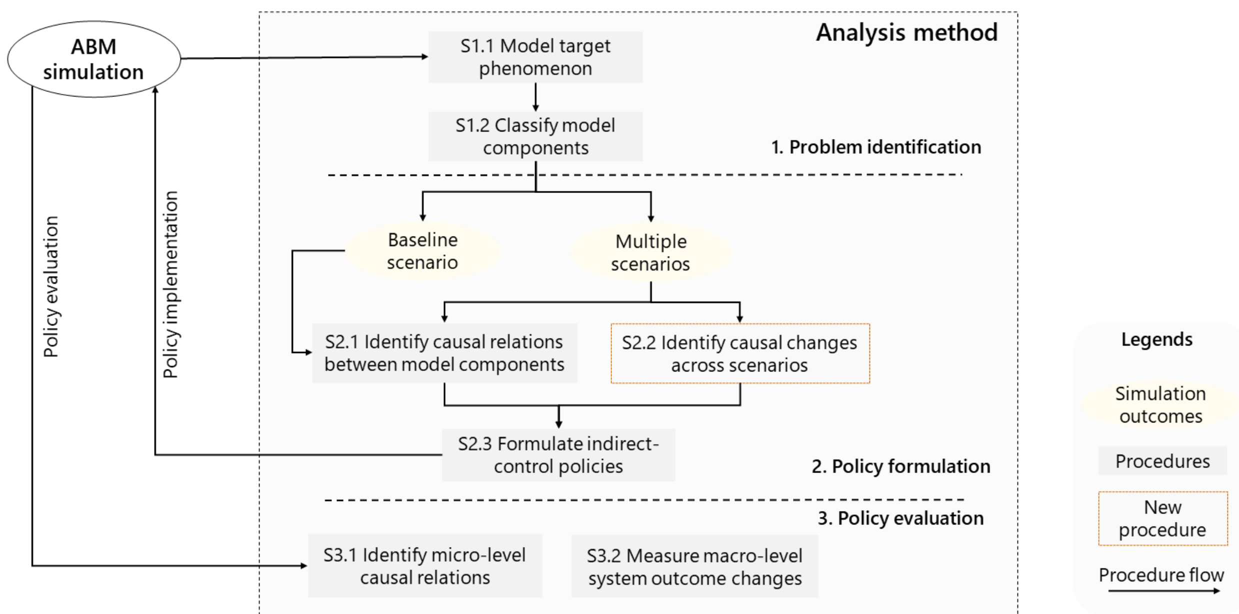


Figure 1: An extended iterative policy-making process leveraging causal discovery methods for ABM.

2.1 Problem Identification

In ABM, target social phenomena are explained and explored by modeling and simulating the micro-level dynamics. Agents with certain goals are modeled to observe the environment and make decisions on the basis of these perceptions and their internal states (DeAngelis and Diaz 2019). Policy alternatives directly or indirectly influence agents' behaviors through their reasoning processes.

2.1.1 Model Target Phenomenon (S1.1)

The definition of a target phenomenon observed from the simulation output is provided as the analysis target. The method aims to discover the causal relations between model components including agents' goals, actions, environmental factors, scenario settings and the target phenomenon for understanding the target phenomenon from a causal perspective.

2.1.2 Classify Model Components (S1.2)

The ABM model is simulated under different scenarios and the logs including agents' characteristics and behaviors through the simulation are recorded. The simulation output variables are then categorized into four model components: *Goals* as agents' planned tasks, *Actions* as agents' behaviors that realize the goals, *Environmental factors* as the environment perceived by agents and *Scenarios* as different scenario settings.

2.2 Policy Formulation

To formulate policies for improving the target phenomenon, the method provides a qualitative explanation of the model behaviors by discovering the causal relations among *Goals*, *Actions*, *Environmental factors*, and *Scenarios*. Multiple causal discovery methods are applied to generate and represent the causal relations using causal graphs.

In addition to identifying the causal relations among model components (S2.1), we extend this stage to multiple scenarios to examine the causal changes across different scenarios (S2.2).

2.2.1 Identify Causal Relations between Model Components (S2.1)

Causal discovery methods have been proven useful in discovering causal relations from observational data (Shen et al. 2020). This analysis method deploys three representative causal discovery methods, the constraint-based method Fast Causal Inference (FCI) (Spirtes et al. 2000), the score-based method Fast Greedy Equivalences Search (FGES) (Ramsey et al. 2017), and Greedy Fast Causal Inference (GFCI) (Ogarrio et al. 2016) that combines the two aforementioned methods, to provide more informative results. They are the most classic or well-performed method from each type, have different assumptions on the input datasets and handle confounders differently. TETRAD, a software package that implements a set of causal discovery tools (http://www.ccd.pitt.edu), is used for implementing the causal discovery methods.

Generate causal graphs. For each causal discovery method, the input records are bootstrapped for 200 times using the *Majority* setting in TETRAD, which will return the sparest set of edges based on the probability of occurrences of each edge. We plot the edge in the final graph if its average probability of occurrence by the three causal discovery methods is higher than a pre-defined threshold. For simplicity, we treat non-directed edges as bi-directed edges without considering the influence of confounders.

2.2.2 Identify Causal Changes across Scenarios (S2.2)

For simulation outputs from multiple scenarios, we conduct the causal analysis on each scenario respectively and additionally examine the causal structure changes by comparing the edge frequencies across scenarios. This step will generate some interesting insights on whether the effect of certain causes on the target phenomenon may vary across scenarios, and under what circumstances the effect may emerge or vanish. The derived knowledge is expected to inspire further discussion and investigation among stakeholders.

2.2.3 Formulate Indirect-control Policies (S2.3)

On the basis of these causal understandings, indirect-control policies, which do not intervene with the analysis target directly but indirectly intervene with the agent actions that lead to the analysis target through environmental factors, can be proposed. This process is iterative in the sense that variables can be iteratively added to or removed from the analysis on the basis of insights yielded from each analysis cycle.

2.3 Policy Evaluation

The model is then re-simulated with the proposed policies to generate new simulation outcomes. Causal analysis steps are conducted on these new records with the policies termed as scenario variables.

The policies are evaluated at both the macro-level against aggregated metrics (S3.2), and at the micro-level by examining the causal relations between policies and other model components (S3.1) to demonstrate how the proposed policy may affect the analysis target through model components. It explicitly provides the causal evidence of policy impact, and is expected to facilitate the communication among policy makers, modelers and other stakeholders who have different domain backgrounds.

3 APPLICATION

We incorporate the above-explained method to an agent-based model in an iterative way to optimize the rack layouts. The model simulates the in-store traffic of customers considering their preferences and actions against different rack layouts. It follows a well-acknowledged movement model in large facilities (Ohori et al. 2016) and is further calibrated and validated using the real-world data collected from a supermarket. The model components are briefly reviewed as follows.

3.1 Agent-based Model

The rack layout of a retail store is modeled as a network. As illustrated in Figure 2 , racks of a variety of product categories, entrances, exits and cashiers are represented by nodes, whilst routes connecting them are represented by edges. The customers' in-store movements are modeled as acquiring a list of racks to visit, entering the retail store from an entrance, walking through the racks to complete the list, and leaving the store from an exit once they have visited all the planned racks.

3.1.1 Characteristics of Agents and Environment

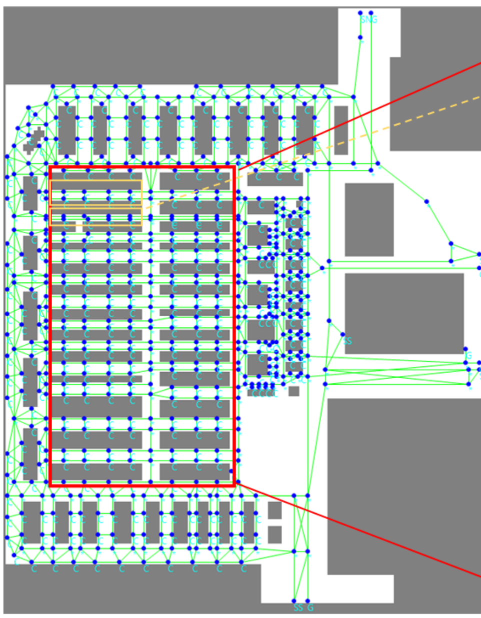
We define the rack layout network in terms of Spots and Routes, and Customers as agents.

Spots are defined by a set of spots $S = \{s_1, s_2, \ldots, s_n\}$ representing racks of products, entrances, exits and cashiers. Each s_i is defined by $\{ID,\ type\}$, where *ID* represents the unique identifier of this spot, and *type* represents whether it is an entrance, an exit, a cashier, or a rack. Each rack can place only one type of products $category_i \in Category$. In this particular case, we assume there are 10 product categories $Category = \{alcohol,\ seasoning,\ cookII,\ snack,\ quickF,\ insNoodle,\ housekeeping,\ sChoice,\ frozenF,\ drink\}$ representing the alcohol products, seasoning products, cooking ingredients, snacks, quick food, instant noodles, housekeeping utilities, self-choice products, frozen food, and drinks.

Routes are defined by a set of edges $E = \{e_1, e_2, \ldots, e_n\}$ connecting spots. Each e_i is defined by $\{(n,m),\ weight\}$, where $(n,m) \in S \times S$ represents the connected spots, and *weight* represents the weight of this edge, which is defined as the distance in the real store and will be used to find the shortest route.

Customers are defined by a set of customers $C = \{c_1, c_2, \ldots, c_n\}$. Each c_i is defined by $\{cSet,\ rackList,\ pList,\ distW,\ distWP,\ entrance,\ exit\}$, where *cSet* represents the product categories that the customer plans to purchase, *rackList* represents the list of ordered racks that the customer is going to visit, which is defined at the beginning of each simulation run, *pList* represents the list of categories which the customer has purchased and is initially empty, *distW* and *distWP* are the total walking distances with and without purchasing after this customer has visited all the racks listed in *rackList*, and *entrance* and $exit \in S$ represents the starting and ending spot as entrance and exit. Here walking distances without purchasing is defined as the longest route between any two neighbored racks in *rackList*.

Routing behaviors. Customers enter the virtual retail store from an *entrance* at simulation time t. They visit the racks following the order defined in *rackList* and choose the shortest route, which is calculated based on the *weight* of each edge. Once they reach the planned rack, they will purchase the product according to certain probability, and add the product category to *pList*. The visited rack (or spot) will then be removed from the *rackList* and customers will proceed to the next destination repeating the same procedures as long as *rackList* is not empty. Once the list becomes empty, customers will leave the retail store through an *exit*. We assume that there are no interactions among the customers and thus no congestion situation will occur.



Alcohol	Alcohol
Drink	Alcohol
Drink	Alcohol
Housekeeping utility	Personal selection
Housekeeping utility	Seasoning
Housekeeping utility	Cup noodle
Housekeeping utility	Snack
Quick food	Snack
Quick food	Snack
Cooking ingredient	Cooking ingredient
Frozen food	Frozen food
Frozen food	Frozen food

Figure 2: Store layout represented as a network (left); We define the racks facing each other along the same aisle and placed with the same product category as one rack in the simulation (as demonstrated in yellow), and focus on the 24 racks along 12 aisles circled in red; An example of the rack layout placed with different product categories is demonstrated on the right.

3.2 Calibration and Validation

The baseline scenario of the model is calibrated and validated using real-world data collected from a supermarket. Assumptions are made when real-world data is not available.

- The baseline rack layout roughly follows the layout of a real supermarket. We focus on the main area with 24 racks in 12 rows, and each rack is placed with one product category. Frozen food is always allocated by the southern end of the racks.
- *cSet* is calibrated using POS data which contains the purchasing records of customers. The visiting probability of each category is estimated from camera data.
- *entrance* of each customer is determined by a probability estimated from camera data. The probability of entering from the northern entrance is round 60%, leaves the rest 40% of customers entering from the southern entrance.
- *exit* is determined randomly.

- *rackList* is defined by mapping the product categories to the corresponding racks once the layout is determined, and the order is determined on the basis of which entrance they have chosen. If a customer enters the store from the northern entrance, then they will start from the rack placed in the north.

We apply the data collected from portable self-scanning devices used by 2521 customers in several days for validating the baseline outcomes. Customers use such kind of devices to scan the bar-codes of products and the order of purchases are thus recorded. The simulated routes from 1500 agents in terms of the order of visited racks *pList* at each round are compared with the real records for validation.

Table 1: Variables recorded for causal analysis.

Variable	Meaning	Value
Customer records for the 1^{st} round analysis		
$category_i$	Whether the agent has visited this category or not $category_i \in Category$	0: No; 1: Yes
entry	The *ID* of the entrance spot	$\mathbb{N}$
exit	The *ID* of the exit spot	$\mathbb{N}$
passedRacks	The number of racks customers have passed by	$\mathbb{N}$
distWP	Total walking distance without purchasing	$\mathbb{R}$
distW	Total walking distance	$\mathbb{R}$
Layout records for the 2^{nd} round analysis		
$categoryAvg_i$	Average visited rate of each category $categoryAvg_i \in Category$	[0, 1]
width	The spanned width of popular racks	$\mathbb{N}$
position	The starting position of the popular rack from south	$\mathbb{N}$
wFrozen	Whether the category is placed next to frozen foods	0: No; 1: Yes
devDistWP	Deviation from the average walking distance without purchasing	$\mathbb{R}$
devDistW	Deviation from the average walking distance	$\mathbb{R}$

4 ANALYSIS FOR LAYOUT OPTIMIZATION

The model was simulated under around 1640 randomly generated rack layouts, and for each layout around 1500 agents were recorded. Programming language FORTRAN is used to enable a faster implementation. Each customer's preferences in terms of *cSet* will remain the same, but the routes in terms of *rackList*, *entry*, and *exit* will be different across different layouts.

By iteratively applying the analysis method on the simulation records, we aim to identity special patterns that may optimize the analysis target as minimizing customers' walking distance without purchasing, taking into account of the preferences and routing behaviors of heterogeneous customers. Walking distance without purchasing is a major metric to measure the shopping experience of customers in retail management—the smaller the value, the better shopping experience customers may have. On the other hand, maximizing customers' walking distance is another major metric for maximizing the store revenue, since impulse purchasing is assumed to be positively related to the walking distance by retail experts. Due to page limitations, we only focus on the walking distance without purchasing in this paper.

4.1 First Round Analysis

We first analyze the customer records across selected layouts to offer a causal explanation on what factors may lead to a short walking distance without purchasing and whether those factors may vary across different layouts. Variables of each customer recorded for this round of analysis are explained in Table 1 (upper table).

4.1.1 Problem Identification

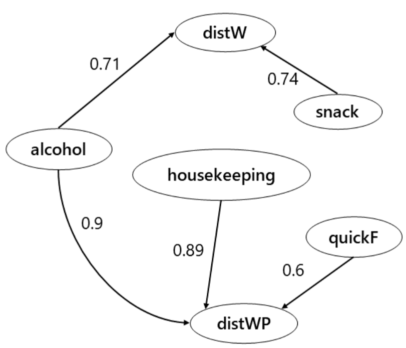
S1.1 Model target phenomenon and **S1.2 Classify model components.** We define *distWP*—the total walking distance without purchasing—as the analysis target. We further categorize the customer variables in Table 1 as follows: *category$_i$* as the customer's *Goals*, *passedRack*, *entry*, *exit*, and *distW* as *Actions*, and different layouts as *Environmental factors* and *Scenarios*.

4.1.2 Policy Formulation

We apply the causal discovery methods on the aforementioned variables with prior knowledge that may help determine the direction of causal relations. We assume that the target of analysis can not be the cause of any other variables and whether a customer has visited a category or not is not resulted by any other variables.

S2.1 Identify causal relations between model components and **S2.2 Identify causal changes across scenarios.** We plot the causal structures and compare them across different layouts in Figs. 3 and 4. For demonstrative purposes, we select two representative layouts, one (Figure 3) is with a short average walking distance without purchasing of all customers, whilst the other (Figure 4) is the opposite case. We then plot the merged graph under these two layouts. The causal graphs show that whether the customers have visited certain categories are the most frequently identified causes of the number of passed racks, the walking distance and the walking distance without purchasing. In addition, the identified product categories are different across layouts.

To generate more informative results, we further analyze the layouts in terms of the categories ranked by the average visited rate and their position (Figs. 3 and 4, middle and right). The major difference between these two graphs is that for the layout with better customer experiences, only the most popular categories are identified as the main causes. If we investigate into the details, it appears that most of the categories in this layout are allocated in a compact way. However, we don't know in what way the popular categories and their allocations may have an impact on the analysis target.



Causal relations

Category	Visited rate
Alcohol	0.48
Housekeeping	0.48
Snack	0.33
Drink	0.33
Quick food	0.28
Frozen food	0.27
Personal selection	0.18
Seasoning	0.15
Cooking ingredient	0.12
Cup noodle	0.10

Ranked categories

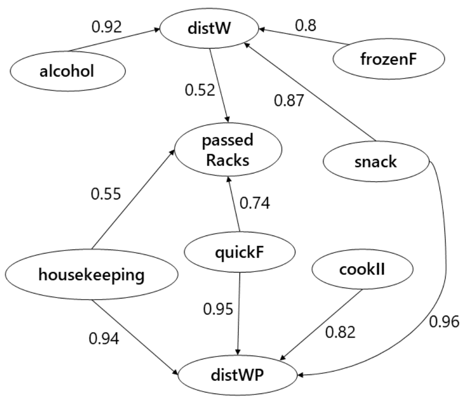
Personal selection	Seasoning
Cooking ingredient	Cooking ingredient
Cup noodle	Snack
Quick food	Snack
Quick food	Snack
Alcohol	Alcohol
Alcohol	Alcohol
Housekeeping	Housekeeping
Housekeeping	Housekeeping
Drink	Drink
Frozen food	Frozen food
Frozen food	Frozen food

Rack Layout

Figure 3: Left: Causal relations with the average probability of occurrence displayed on each edge when the value > 0.5. X $\longrightarrow$ Y indicates that X might be a cause of Y; Middle: Ordered categories by the average visited rate; Right: Rack layout with a relatively good customer experience; popular categories highlighted in red.

S2.3 Formulate indirect-control policies. On the basis of the aforementioned analysis, we realize that the arrangement of popular categories, for example, should they be placed in the middle of the racks or at sides, and whether certain categories should be placed together or not, may have an impact on the customer experience.

To quantify the impact of special arrangement, we define some special layout patterns as new variables for each layout record, namely the starting rack of the popular items from south as *position*, how many



Causal relations

Category	Visited rate
Alcohol	0.55
Drink	0.37
Housekeeping	0.34
Frozen food	0.3
Quick food	0.3
Snack	0.28
Cooking ingredient	0.24
Personal selection	0.16
Cup noodle	0.09
Seasoning	0.08

Ranked categories

Alcohol	Alcohol
Drink	Alcohol
Drink	Alcohol
Housekeeping	Personal selection
Housekeeping	Seasoning
Housekeeping	Cup noodle
Housekeeping	Snack
Quick food	Snack
Quick food	Snack
Cooking ingredient	Cooking ingredient
Frozen food	Frozen food
Frozen food	Frozen food

Rack layout

Figure 4: Left: Causal relations with the average probability of occurrence displayed on each edge when the value > 0.5. Middle: Ordered categories by the average visited rate; Right: Rack layout with a relatively bad customer experience; popular categories highlighted in red.

racks the popular categories have spanned as *width*, and whether the identified unpopular categories are placed next to the frozen food or not as *wFrozen*. These layout pattern variables will be used for the next round analysis to examine their causal impact on the analysis target.

4.2 Second Round Analysis

We then repeat the analysis steps from S1.1 to S2.3 on the layout records. Variables of each layout recorded for this round of analysis are explained in Table 1 (lower table).

4.2.1 Problem Identification

S1.1 Model target phenomenon and **S1.2 Classify model components.** We set *devDistWP*—the deviation from the average walking distance without purchasing of all customers—as the analysis target. We categorize the average visited rate of each category $categoryAvg_i$ as *Goals*, the deviation from the average walking distance *devDistW* as *Actions*, and *position*, *width*, and *wFrozen* as *Environmental factors*.

4.2.2 Policy Formulation

S2.1 Identify causal relations between model components. We first conduct this step on the records without the variable *wFrozen*. The causal relations in Figure 5 (left) show that the starting position of the popular categories is the root cause of the analysis target. The position is also frequently identified by two of the causal discovery methods. Interestingly, the category drink is frequently identified as a cause of the position and requires further investigation. *wFrozen* becomes the root cause if we include this variable in the analysis, as in Figure 5 (right).

To further investigate whether the impact is positive or negative, we estimate the causal effect by assuming the causal relation is linear and follows a Structure Equation Model (SEM) structure. The function *Estimator* provided in TETRAD is used to estimate the effect. The results show that the causal effects of *position* and *wFrozen* on *devDistWP* are negative whilst those of *width* and *devDistW* are positive. It indicates that the compact arrangement of popular items can shorten the walking distance without purchasing. On the other hand, retail managers may want to extend the walking distance of customers, since it will prompt impulse purchasing through walking around the store.

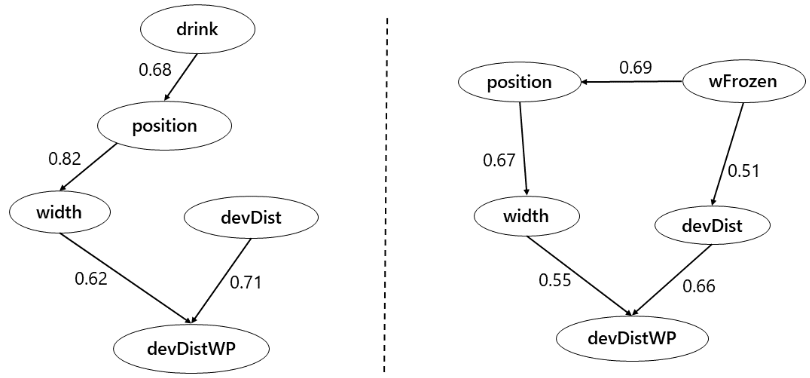


Figure 5: Causal relations with the average probability of occurrence displayed on each edge when the value > 0.5. Left: without *wFrozen*; Right: with *wFrozen* considered.

S2.3 Formulate indirect-control policies. As a result, we propose optimized layouts as placing the popular items in a compact way and some unpopular items with the frozen food, which may prompt positive customer shopping experience. In addition, placing the popular products in a compact way on the southern side of the store may potentially increase the store revenue since it extends the walking distance and may prompt more impulse purchases. This step can also be treated as **S3.1 Micro-level evaluation**, which evaluates the proposed patterns from a causal perspective.

4.2.3 Policy Evaluation

S3.2 Macro-level evaluation. We plot a heat diagram in Figure 6 which measures the layouts against the walking distance and walking distance without purchasing across all possible combinations of the starting position and width of the most popular categories. It aims to triangulate whether the proposed pattern—compact layouts in the middle or in the south—can optimize the customers' experience.

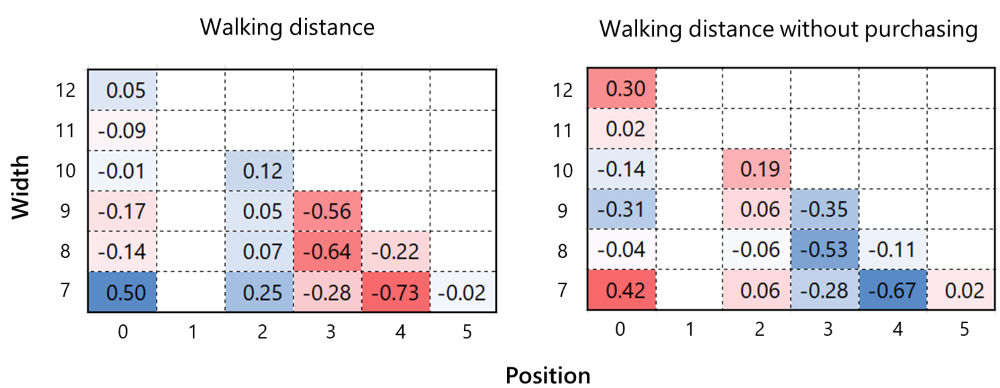


Figure 6: Evaluation of the proposed special patterns defined by *position* (x-axis) and *width* (y-axis). Cell value is the average value of *devDistW* (left) and *devDistWP* (right) of all layouts that fall into the corresponding pattern. Since the frozen food is always placed at the southern racks in two rows, the cell value for *position* $= 1$ is not available and thus blank.

It reveals that layouts with a compact arrangement (*width* ≤ 9) of popular items in the center (*position* > 2), as cells in blue of Figure 6 (right), prompt a better customer experience in terms of a short walking distance without purchasing. Layouts with a compact arrangement (*width* ≤ 9) of popular items in the south (*position* ≤ 2), as cells in blue of Figure 6 (left), have the potential to prompt impulse purchasing since the walking distance is extended, and at the same time can guarantee an acceptable customer experience. These results evaluated from the macro-level perspective are consistent with the insights derived from the causal relations, and thus demonstrate the applicability of the proposed method in optimizing rack layouts with explainable patterns.

4.2.4 Comparison with Conventional Analysis Method

Compared with the conventional simulation analysis methods, which may identify the optimized layouts inductively from randomly generated layouts, we demonstrated that the proposed method can efficiently support the decision-making of retail management by explicitly and proactively proposing special patterns that optimize the layouts against the target metrics. In addition, the proposed optimal layouts can be explained with explicit causal evidences, which is expected to greatly improve the communication among stakeholders of different backgrounds.

5 CONCLUSION AND FUTURE WORKS

Rack layout optimization is an important yet challenging task for retail management especially when it comes to consider the dynamics between customers and layouts. It is necessary to efficiently identify and effectively explain special patterns for layout optimization yet difficult by conventional modeling or analysis methods. In this work, we extended a causal-based analysis method of ABM to handle multiple scenarios, calibrated an agent-based model using a variety of real-world data that simulates the customers' routing behaviors against different rack layouts in a virtual retail store, and innovatively incorporated them in an iterative way for optimizing rack layouts with explainable patterns.

Our findings show that the proposed method can successfully derive explainable layout patterns of arranging the products that prompt positive customers' experiences and potentially maximize the store revenue by iteratively discovering the causal relations and changes among customers' preferences, actions and analysis targets across different layouts. The incorporated model and analysis method is abstract and general enough to capture the essential components of retail stores. We therefore expect this causal explanation empowered layout optimization can be customized and applied in real applications to support the decision-making in retail and to greatly enhance the communication among stakeholders.

There are several future directions to strengthen the model and the analysis method. First, the model can be further improved, calibrated and validated with empirical data collected by more advanced in-store technologies, such as the data on customers' purchasing behaviors extracted from in-store cameras. Second, with an increased number of variables, the global structure of the causal relations may become complicated and harder to verify. Domain knowledge or advanced causal discovery methods are necessary to provide more reliable explanations. Finally, optimization methods can be integrated in the analysis process to further validate the causal-enabled optimized patterns.

REFERENCES

Abdulkareem, S. A., Y. T. Mustafa, E.-W. Augustijn, and T. Filatova. 2019. "Bayesian Networks for Spatial Learning: A Workflow on Using Limited Survey Data for Intelligent Learning in Spatial Agent-Based Models". *Geoinformatica* 23(2):243–268.

Chang, S., T. Asai, Y. Koyanagi, K. Uemura, K. Maruhashi and K. Ohori. 2022. "Incorporating AI Methods in Micro-dynamic Analysis to Support Group-Specific Policy-Making". In *PRIMA 2022: Principles and Practice of Multi-Agent Systems*, edited by R. Aydoğan, N. Criado, J. Lang, V. Sanchez-Anguix, and M. Serramia, 122–138. Cham: Springer International Publishing.

Chang, S., T. Kato, Y. Koyanagi, K. Uemura and K. Maruhashi. 2023. "An Iterative Analysis Method using Causal Discovery Algorithms to Enhance ABM as a Policy Tool". In *2023 Winter Simulation Conference (WSC)*, edited by C. G. Corlu, S. R. Hunter, H. Lam, B. S. Onggo, J. Shortle, and B. Biller, 1–12. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

DeAngelis, D. L. and S. G. Diaz. 2019. "Decision-Making in Agent-Based Modeling: A Current Review and Future Prospectus". *Frontiers in Ecology and Evolution* 6 https://doi.org/10.3389/fevo.2018.00237.

Drchal, J., M. Čertický, and M. Jakob. 2019. "Data-driven Activity Scheduler for Agent-based Mobility Models". *Transportation Research Part C: Emerging Technologies* 98:370–390.

Drira, A., H. Pierreval, and S. Hajri-Gabouj. 2007. "Facility Layout Problems: A Survey". *Annual Reviews in Control* 31(2):255–267.

Flamand, T., A. Ghoniem, and B. Maddah. 2016. "Promoting Impulse Buying by Allocating Retail Shelf Space to Grouped Product Categories". *Journal of the Operational Research Society* 67:953–969.

Holland, P. W. 1986. "Statistics and Causal Inference". *Journal of the American Statistical Association* 81(396):945–960.

Janssen, S., A. Sharpanskykh, R. Curran, and K. Langendoen. 2019. "Using Causal Discovery to Analyze Emergence in Agent-based Models". *Simulation Modelling Practice and Theory* 96:101940.

Lee, J. S., T. Filatova, A. Ligmann-Zielinska, B. H. Mahmooei, F. Stonedahl, I. Lorscheid, , , *et al*. 2015. "The Complexities of Agent-Based Modeling Output Analysis". *Journal of Artificial Societies and Social Simulation* 18(4):4.

Lu, Y. and H.-B. Seo. 2015. "Developing Visibility Analysis for a Retail Store: A Pilot Study in a Bookstore". *Environment and Planning B: Planning and Design* 42(1):95–109.

Mowrey, C. H., P. J. Parikh, and K. R. Gue. 2018. "A Model to Optimize Rack Layout in a Retail Store". *European Journal of Operational Research* 271(3):1100–1112.

Ogarrio, J. M., P. Spirtes, and J. Ramsey. 2016. "A Hybrid Causal Search Algorithm for Latent Variable Models". In *Proceedings of the Eighth International Conference on Probabilistic Graphical Models*, edited by A. Antonucci, G. Corani, and C. P. Campos, Volume 52 of *Proceedings of Machine Learning Research*, 368–379. Lugano, Switzerland: PMLR.

Ohori, K., S. Yamane, H. Anai, S. Utsumi and S. Takahashi. 2016. "An Agent-based Analysis of the Effectiveness of Signage Systems in a Large-scale Facility". In *Proceeding of Social Simulation Conference 2016*. September 19$^{\text{th}}$-23$^{\text{th}}$, Rome, Italy, 1-12.

Pantano, E., G. Pizzi, E. Bilotta, and P. Pantano. 2021. "Enhancing Store Layout Decision with Agent-based Simulations of Consumers' Density". *Expert Systems with Applications* 182:115231.

Pearl, J. 2009. *Causality*. Cambridge: Cambridge University Press https://doi.org/10.1017/CBO9780511803161.

Ramsey, J., M. Glymour, R. Sanchez-Romero, and C. Glymour. 2017. "A Million Variables and More: the Fast Greedy Equivalence Search Algorithm for Learning High-dimensional Graphical Causal Models, with an Application to Functional Magnetic Resonance Images". *International Journal of Data Science and Analytics* 3:121–129.

Schenk, T. A., G. Löffler, and J. Rauh. 2007. "Agent-based Simulation of Consumer Behavior in Grocery Shopping on a Regional Level". *Journal of Business Research* 60(8):894–903.

Shen, X., S. Ma, P. Vemuri, G. Simon and Alzheimer's Disease Neuroimaging Initiative. 2020. "Challenges and Opportunities with Causal Discovery Algorithms: Application to Alzheimer's Pathophysiology". *Scientific Reports* 10(1):2975.

Spirtes, P., C. Glymour, R. Scheines, S. Kauffman, V. Aimale and F. Wimberly. 2000. "Constructing Bayesian Network Models of Gene Expression Networks from Microarray Data". In *Proceedings of the Atlantic Symposium on Computational Biology*. March 15$^{\text{th}}$-17$^{\text{th}}$, Durham, USA, 1-5. https://doi.org/10.1184/R1/6491291.v1.

Yamada, H., S. Yamane, K. Ohori, T. Kato and S. Takahashi. 2020. "A Method for Micro-Dynamics Analysis Based on Causal Structure of Agent-Based Simulation". In *2020 Winter Simulation Conference (WSC)*, edited by K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, 313–324. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Yamane, S., H. Yamada, K. Ohori, H. Anai, S. Sakai, K. Takahashi, *et al*. 2018. "Systematic Analysis of Micro Dynamics in Agent Based Simulation". In *2018 Winter Simulation Conference (WSC)*, edited by M. Rabe, A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 4214–4215. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Yapicioglu, H. and A. E. Smith. 2012. "Retail Space Design Considering Revenue and Adjacencies using a Racetrack Aisle Network". *IIE Transactions* 44(6):446–458.

AUTHOR BIOGRAPHIES

SHUANG CHANG is a Researcher in the Fujitsu Research. She earned her Ph.D. in System Sciences from Tokyo Institute of Technology. Her research interests include long-term care systems and public service design mainly by an agent-based simulation approach. Her email address is chang.shuang@fujitsu.com.

SHOHEI YAMANE is a Researcher in the Fujitsu Research. He earned his Ph.D in Information Science from Kyoto University. His email address is yamane.shohei@fujitsu.com.

KOJI MARUHASHI is a Project Manager in Fujitsu Research. He earned the Ph.D in Engineering from Tsukuba University. His email address is maruhashi.koji@fujitsu.com.