Decoding Policy Interactions in System Dynamics: A Balanced Tree Clustering Approach

Yuta Sumiya[*1] Naoyuki Echizenya[*1] Tomoyuki Taguma[*1] Kyosuke Toda[*1]
Takahisa Ando[*1] Yoshiaki Oida[*1] Yutaka Takahashi[*2]

[*1] Fujitsu ltd. [*2] Senshu University

This study aims to provide a method for interpreting System Dynamics (SD) simulations using balanced tree clustering, focusing on analyzing nonlinear interactions and feedback loops in complex systems. While SD is a powerful tool for exploring dynamic behaviors, understanding variable impacts and identifying critical thresholds across numerous scenarios remains challenging. To address this, we show a new method to extract thresholds for policy variables by applying balanced tree clustering to the bass model. The results reveal clear splits in clusters at specific policy values, offering a structured approach to interpreting SD scenarios. This method helps decision-makers identify critical factors and derive actionable guidelines for achieving desired outcomes.

1. Introduction

System Dynamics (SD) is an approach for modeling and simulating complex systems. It helps in understanding their dynamic behavior, generating simulation results through the interaction of policy parameters and internal model structures. Identifying the relationship between the range of policy parameter values and simulation outcomes leads to a deeper understanding of the model's essence, enabling effective interventions in real-world systems. However, due to the nature of SD models, interpreting the impact of specific system elements becomes difficult because policy parameters interact nonlinearly within the model.

To address this challenge, it is effective to interpret the model's behavior post hoc based on scenarios of policy parameters and simulation results. Decision tree clustering is one method that allows for the post-hoc interpretation of SD models from their simulation outcomes. This approach clusters scenarios based on policy parameter values, constructing a tree structure related to these parameters. Interpreting this tree structure allows us to identify how differences in scenarios arise due to the influence of specific policy parameters, pinpointing watersheds in policy values. However, interpreting diverse policy combinations along these watersheds requires generating a larger number of scenario patterns. Since policy parameters are set within specific ranges, we can generate a wide variety of scenario patterns by setting numerous random values. In this study, we aim to establish a new interpretation method that combines Monte Carlo simulation and decision tree clustering for SD models, enabling interpretation through a tree structure while considering the characteristics of the scenarios.

2. A Balanced Tree Clustering

SD models, multiple policy parameters are input to generate time-series data. This time-series data represents a

Contact: Yuta Sumiya, Fujitsu Limited, Japan, Tel: +81-80-7250-3561, Email: sumiya-yuta@fujitsu.com

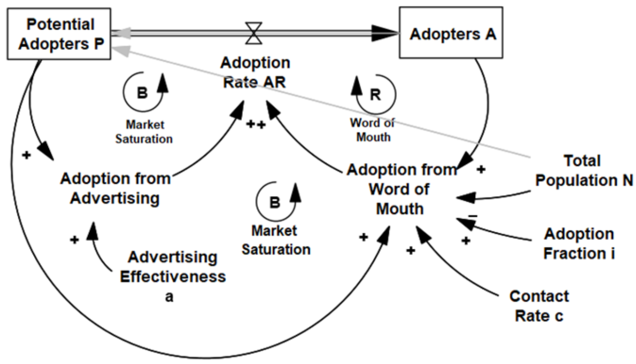
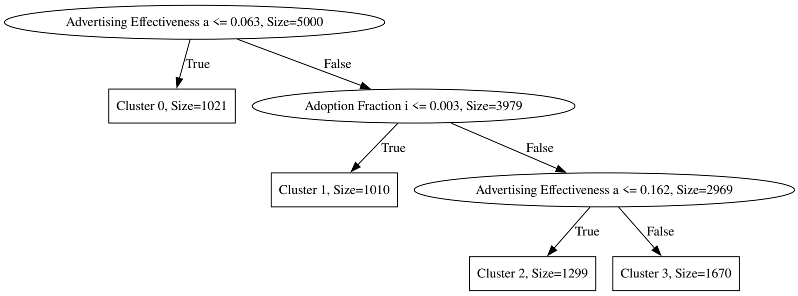
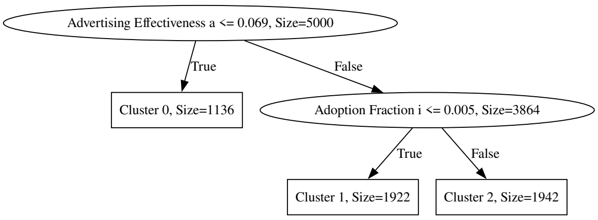


Figure 1: The adoption process in the Bass model. The Adoption Rate (AR) is influenced by Advertising and Adoption from Word of Mouth, facilitating the transition from Potential Adopters (P) to Adopters (A). The effect of advertising is represented by Advertising Effectiveness (a), while the effect of word-of-mouth is determined by a combination of the Contact Rate (c) and the Adoption Fraction (i). These factors also contribute to the progression of adoption, being influenced by Market Saturation.

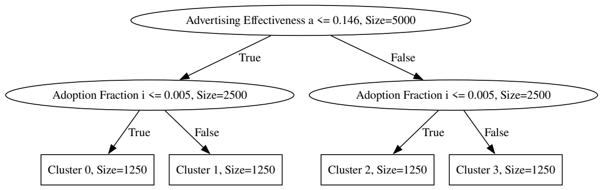
single scenario, which helps in analyzing the impact of different policy combinations on the system. However, simply executing simulations with SD only allows for the evaluation of scenarios under specific policy combinations, making it difficult to assess the complex interplay between policies. In this regard, clustering techniques, which classify and interpret data into potential groups, can be effective. Utilizing clustering when interpreting multiple scenarios helps clarify the characteristics exhibited by each group. Yet, clustering solely based on time-series data, while able to capture data fluctuations over the time axis, cannot specifically determine which policies are directly related to the improvement or decline of particular economic indicators. Conversely, clustering that considers only policy parameters can analyze their inherent characteristics and interactions, but it fails to leverage the time-series features of the scenarios,



(a) Decision tree ($\alpha = 0.1$).



(b) Decision tree ($\alpha = 0.3$).



(c) Decision tree ($\alpha = 0.5$).

Figure 2: Decision trees obtained by applying clustering to the Bass model. The three decision trees were derived by varying the adjustment coefficient. The minimum number of scenarios contained in each cluster is 1000.

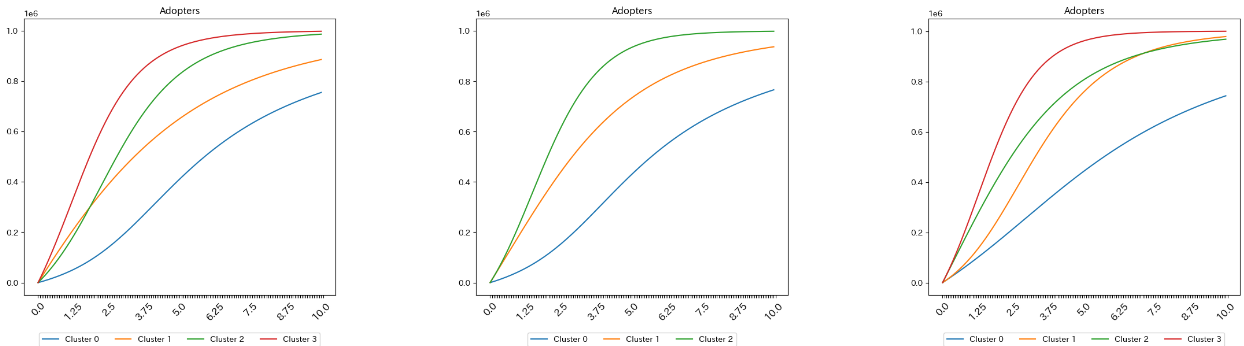


Figure 3: The time-series progression of adoption rates for each cluster is shown. From left to right, these correspond to adjustment coefficients $\alpha = 0.10, 0.30$, and 0.50. The horizontal axis represents time-series steps, and the vertical axis represents the adoption rate. The time-series range is from 0 to 10 in increments of 0.0625.

thus preventing a proper evaluation of changes based on dynamic behavior.

SD models requires applying clustering techniques that can consider both policy parameters and time-series data. Decision tree-based methods [1, 2] allow for cluster evaluation using time-series dependent features and cluster division using time-series independent features. Applying this approach to SD enables interpretation that simultaneously considers policy parameter values and the resulting time-series scenarios. In this study, we evaluate the quality of scenarios along the decision tree branches, using combinations of randomly set policy parameters and scenarios. Furthermore, we propose a novel interpretation method where we add a new balance constraint to enhance the interpretability of the tree, allowing it to be interpreted as a balanced tree.

In balanced tree clustering, clusters are divided using the following evaluation function:

$$S(\mathcal{C}_1, \mathcal{C}_2) = f(\mathcal{C}_1, \mathcal{C}_2) + \alpha g(\mathcal{C}_1, \mathcal{C}_2), \tag{1}$$

Here, $\mathcal{C}_1$ and $\mathcal{C}_2$ are the two divided cluster sets, $f(\bullet, \bullet)$ is a function that calculates the surrogate silhouette coefficient for cluster division [2], and $g(\bullet, \bullet)$ is the following function representing the ratio of the number of data points in the left and right clusters, serving as a balance constraint:

$$g(\mathcal{C}_1, \mathcal{C}_2) = -\frac{||\mathcal{C}_1| - |\mathcal{C}_2||}{\max(|\mathcal{C}_1|, |\mathcal{C}_2|)}. \tag{2}$$

Additionally, α is an adjustment coefficient. We perform an exhaustive search for the policy parameters and thresholds that maximize $S(\bullet, \bullet)$.

3. Experiments and Discussion

We adopt the Bass model, shown in Figure 1, as the SD model. The Bass model predicts how the Adoption Rate (A) of a product increases over time, driven by two policy parameters: Advertising Effectiveness (a) and Adoption Fraction (i). In our experiment, we use balanced tree clustering to interpret how these two policy parameters influence the overall market adoption rate behavior.

The clustering results are presented in Figure 2. From the decision tree structures, it is evident that as the adjustment coefficient increases, the depths of the left and right subtrees become more uniform. Referring to the nodes, Advertising Effectiveness is consistently selected at the root node in all cases, with Adoption Fraction tending to be prioritized in the subsequent layer. No changes are observed in the policy parameters at different depths of the tree structure; instead, the tree's overall shape changes according to the watersheds of policy values. Figure 3 illustrates the time-series progression for each cluster. Clusters are differentiated by their adoption rate's speed of rise to convergence, ranging from fast to slow, clearly reflecting the characteristics of each cluster. In all tree structures, clusters located on the right-hand side of a node show faster rises and quicker convergence in their graphs. This indicates that the Adoption Rate increases when both Advertising Effectiveness and Adoption Fraction exceed the thresholds specified in the nodes. Furthermore, the more uniform the depths of the left and right subtrees are (as seen in the rightmost tree in Figure 2), the easier the interpretation becomes.

4. Conclusion and Future Work

We proposed an interpretation methodology that utilizes balanced tree clustering to identify effective policy combinations from SD model scenarios. This approach categorized scenarios according to a tree structure defined by policy parameters, which served to delineate the watersheds associated with individual policy values. As for future prospects, we intend to validate the applicability of this clustering method to more intricate models possessing a larger set of policy parameters compared to the Bass model.

References

[1] M. Moshkovitz *et al.*, "Explainable k-means and k-medians clustering," *Proc. Int. Conf. Mach. Learn. (ICML)*, pp. 7055–7065, 2020.

[2] M. Higashi *et al.*, "Decision Tree Clustering for Time Series Data: An Approach for Enhanced Interpretability and Efficiency," *Proc. Pacific Rim Int. Conf. Artif. Intell.*, pp. 457–468, 2023.