2A2-4

場面遷移ネットと強化学習モデルを用いたサービス設計のための顧客意思決定シミュレーション手法の提案

舘山 武史[*1] 川田 誠一[*2] 下村 芳樹[*1]
Takeshi Tateyama Seiichi Kawata Yoshiki Shimomura

[*1]首都大学東京システムデザイン学部
Dept. of System Design, Tokyo Metropolitan University

[*2]産業技術大学院大学
Advanced Institute of Industrial Technology

Recently, a new academic field, "service engineering" has been very actively investigated. However, there are few effective software tools to simulate and evaluate services designed based on the concept of service engineering. In the past, the authors proposed a service flow simulation method using scene transition nets(STN) which is a graphic modeling and simulation method for discrete-continuous hybrid system. However, this method cannot simulate complex service flows including customers' decision-making. Nowadays, "neuro economics" and "neuro marketing" have gotten a lot of attention as new study fields to understand customers' behaviors from a viewpoint of brain science. In these studies, it turned out that mechanism of reinforcement learning concerns behavioral selections of customers. In this paper, the authors propose to develop decision-making processes models of customers and to simulate customers' behaviors and service flows by using reinforcement learning models and STN.

1. はじめに

近年, 産業界ではサービス産業が一層重要視される傾向にあり, サービスの生産性を向上させることが重要な課題となっている. このような背景から, 工学的な視点からサービスの設計・製造の方法論を確立することを目的とした新しい学問体系であるサービス工学 [下村 05] が提案され, サービス設計を支援するサービス CAD の開発などが進められている. それに伴い, サービス工学の理論に基づいて設計されたサービスの評価を行うためのシミュレーション技法の確立が求められており, これまでに著者らは, 離散・連続混合システムのモデリング・シミュレーション手法である場面遷移ネット (scene transition nets, STN)[川田 93] を用いてサービスの流れをモデル化し, シミュレーションによって評価を行う手法を提案している [佐藤 07]. この手法は, サービスの流れを視覚的に確認することを可能にするとともに, 顧客満足度などのパラメータの時間的推移をシミュレートしサービスの評価を行うことを可能としている. しかし, 本モデリング手法ではサービスの単純な流れ, 例えば条件分岐などが存在しない一方通行のサービスの流れの表現にとどまっており, 状況に応じた顧客の行動選択の違いがサービスの流れにどのような影響を及ぼすかを分析・評価することは困難である. 顧客のふるまいはサービスの成否に多大な影響を及ぼすため, 設計したサービスが設計者が意図したとおりに流れるものであるか否かを検証するためには, 顧客のふるまいを考慮に入れることが不可欠である. そして, そのためには何らかの手法を用いて顧客の意思決定プロセスをモデル中に導入する必要がある.

近年, ニューロエコノミクスやニューロマーケティングなどの研究領域が発達しつつあり, 脳科学的アプローチなどにより消費者の心理や行動を理解するための研究が注目されている. それらの中で, 例えば茂木は消費者の行動は強化学習のメカニズムが深く関係していると述べている [茂木 06]. また, 強化学習モデルは, 生物の行動系列生成のメカニズムと強く関連付けられるという仮説が有力視されている [Schultz 97][Doya 00]. そこで本研究では, Sutton らの強化学習モデル [Sutton 98] を用いて顧客の行動選択モデルを構築し, STN を用いたサービスシミュレーション上に実装することにより, 顧客の意思決定シミュレーションの結果に基づき, 設計したサービスの分析・評価を可能とする手法を提案する.

連絡先: 舘山 武史, 首都大学東京システムデザイン学部
〒 191-0065 東京都日野市旭が丘 6-6
Tel:042-585-8472, E-mail:tateyama@sd.tmu.ac.jp

2. 場面遷移ネット (STN)

場面遷移ネット (以下 STN) は, アクタとシーンという概念に基づき、離散・連続混合システムを図式的に表現するためのモデルである. STN は離散事象システムのモデリングに用いられるペトリネットのコンセプトに基づいており, 並列的に動作する複数のサブシステムが互いに関わり合う様子を明示的かつ動的に表現することが可能である. STN はアクタ (Actor), シーン (場面, Scene), トランジション (Transition), そしてシーンとトランジションを結ぶアーク (Arc) で構成される. アクタはペトリネットのトークン (Token) に相当する概念であるが, 自らの状態変数を保持する点がトークンとの大きな違いである. アクタは後述するシーンに記述されたダイナミクスに従って自らの状態変数を動的に変化させる. また, アクタ変数がある条件 (後述する出力トランジションに記述されている条件) を満たすことにより, アクタは別のシーンに遷移する. つまり, アクタは状態変数を変化させながら, ネットワーク内を移動することになる. また, シーンはペトリネットのプレース (Place) に相当し, アクタ変数の時間的変化を表すダイナミクスが記述される. STN のトランジションは, ペトリネットのトランジションに相当する. トランジションとシーンはアークで結ばれている. 各トランジションには入力側のシーンに位置するアクタが出力側のシーンに遷移するための発火条件と, 遷移時の状態遷移則が記述されている. STN では, 上記の要素を組合せてネットワークを構築後, シミュレーションを実行するが, この際, ユーザーは各アクタの場面遷移の様子を観察することで離散事象システムの解析, またアクタ変数の時間的推移を観察することにより連続変数システムの解析を同時に行なうことが可能である. なお, GUI の利用により, 容易に STN の

シミュレーションを行えるツールとして, 著者らは既に STN GUI Simulator[Tateyama 07] を開発している.

3. STN を用いたサービスフローシミュレーション

既に述べたように, 著者らはこれまでに, STN を用いてサービスの流れのモデリングを行い, シミュレーションを行う手法を提案している [佐藤 07]. 本手法では, サービスの個々のイベントをそれぞれ 1 つのシーンにより表現する. 本手法におけるモデリングとシミュレーションの手順を以下に示す.

1. サービスのイベントの時間的推移を表すサービススクリプト [Fisk 05] と呼ばれる概念を用い, 顧客のサービスの需給過程を STN で構築する
2. 顧客とプロバイダが直接的な相互作用を行うイベントのシーンをサービスエンカウンタ [Fisk 05] として抽出する
3. 抽出したエンカウンタに対応するプロバイダのシーンを作成する
4. エンカウンタ間の遷移を実現するためのプロバイダのシーンを追加作成し, プロバイダのネットワークを構築する
5. 顧客のネットとプロバイダのネットをサービスエンカウンタのシーンを重ねることにより, 1 つのネットワークに統合する
6. シミュレーションを実行し, アクタの動きを観察することにより, サービスの流れを確認する. また, 顧客満足度の時間的推移を観察し, サービスの評価を行う

4. 提案手法：強化学習モデルを導入したサービスフローシミュレーション

4.1 概要と目的

本研究の目的は, STN でサービスの流れをモデル化し, さらに顧客の行動選択のモデルを STN に組み込み, 顧客の意思決定シミュレーションを行うことにより, 設計したサービスの詳細な分析・評価を可能とすることである. 顧客の意思決定モデルは, Sutton らの価値関数を用いた強化学習モデル [Sutton 98] を用いて構築する.

4.2 強化学習と消費者の行動選択

茂木は脳科学の視点から, 人間の欲望と行動選択には, 強化学習のメカニズムが深く関わっているということを述べている [茂木 06]. 人間がある商品やサービスを購入した結果として大脳皮質の下からドーパミンが投射され快楽が得られた場合, 脳は当該する快楽が得られる前に行っていた行動を強化する. すなわち, 快楽が得られる商品やサービスを繰り返し購入することになる. この現象は強化学習と呼ばれ, 茂木はいかに消費者の脳にドーパミン射出によるアディクション (中毒) を生じさせることができるかが, 消費を生み出す鍵であると述べている. また, Sutton の temporal difference learning(TD 学習) に代表される価値関数 (value function) を用いた強化学習モデル [Sutton 98] は, 生物の行動系列生成のメカニズムと強く関連付けられるという仮説が有力視されている. Schultz らは, TD 学習における予測した状態評価値と行動実行後の評価値の差である TD-error とドーパミンの関連性を動物実験によって示している [Schultz 97]. また, Doya は Sutton の強化学習モデルのパラメータ群とドーパミン, アセチルコリンなどの神経伝達物質との関連性の仮設を立てている [Doya 00].

これらのことから, 著者らは Sutton らの強化学習モデルを用いて顧客の行動選択モデルを構築し, 顧客の意思決定シミュレーションを行うことによって設計したサービスの分析・評価を行うことは実効性の高いサービスを開発する上で非常に有効であると考える.

4.3 強化学習モデルと学習アルゴリズム

本研究では, 以下に示すようなマルコフ決定過程 (Markov Decision Processes, MDPs) 環境における強化学習モデル [Sutton 98] を用いる. 環境内における状態集合を S, エージェントが実行可能な行動の集合を A とする. 時刻 t で観測した状態 $s \in S$ においてエージェントが行動 $a \in A$ を実行したとすると, エージェントは時刻 $t+1$ で状態遷移確率 $P^a_{ss'}$ に従って次の状態 $s' \in S$ に遷移する. なお, 状態遷移に要する時間が任意の時間 n である環境モデルを, セミマルコフ決定過程 (Semi-Markov Decision Processes, SMDPs) という. エージェントは遷移先の状態で確率的に報酬 r_{t+1} を受け取る. エージェントの目的は, 任意の時刻 t から今後得られることが予想される報酬の重み付きの和 (Return)

$$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + ... = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \qquad (1)$$

を最大化するような最適政策 (optimal policy) を求めることである. ここで $\gamma(0 < \gamma \leq 1)$ は割引率であり, 未来の報酬をどの程度重視するかを決定するパラメータである. 多くの強化学習法では, 任意の時刻 t における状態 s において行動 a を実行し, その後政策 (policy) π に従って行動する場合の Return の期待値である行動価値関数 (action value function)$Q^\pi(s,a)$ を状態と行動の組の評価関数として用いる.

$$Q^\pi(s,a) = E_\pi\{R_t | s_t = s, a_t = a\} \qquad (2)$$

政策 π の実装方法としては, 状態 s における行動 a を選択する確率 $P(s,a)$ を下記に示す式によって計算し, 確率的に行動を選択する Boltzmann 選択 [Sutton 98] がよく用いられる.

$$P(s,a) = \frac{exp(Q(s,a)/T)}{\sum_{b=1}^{n} exp(Q(s,b)/T)} \qquad (3)$$

ここで T は温度定数であり, 探索行動と報酬獲得のための行動の選択比率を決定するパラメータである. エージェントは, 環境との相互作用から行動価値関数を学習するが, 学習方法としては次式で行動価値関数を単位時間ごとに更新する Q-learning[Watkins 92] などが一般的に多く用いられる.

$$\begin{aligned} Q(s_t,a_t) \quad \leftarrow \quad & Q(s_t,a_t) + \alpha[r_{t+1} + \gamma \max_a Q(s_{t+1},a) \\ & -Q(s_t,a_t)] \end{aligned} \qquad (4)$$

ここで, $\alpha(0 < \alpha \leq 1)$ は学習率である. また, 予測した状態評価値と行動実行後の評価値の差は TD-error と呼ばれ, Q-learning などの更新式は, TD-error と学習率の積を差分として更新しているといえる.

4.4 サービスフローモデルと強化学習モデルの対応

4.4.1 顧客と強化学習エージェント

STN のアクタとして記述される顧客は, シミュレーション上では強化学習エージェントとして, 与えられたサービス内で得られる利益を最大化させることを目的として行動し, その経験から利益を予測する関数を更新する. なお, シミュレーション上ではサービス供給者 (プロバイダ) もアクタとして記述するが, 本論文では学習エージェントとしては扱わない.

4.4.2 サービスのイベントと状態

STN では, サービスの個々のイベントをそれぞれ 1 つのシーンとして表現している. 個々のシーンは, 強化学習モデルにおける個々の状態に相当する.

4.4.3 顧客の行動選択と状態遷移

時刻 t において顧客が位置しているシーンを s_t とする. このとき, シーン s_t において, 顧客は選択できる行動集合 A_{s_i} の中から 1 つの行動 $a_t \in A_{s_t}$ を後述する行動価値関数を用いて確率的に選択し, 実行する. その結果, 前述の SMDPs のモデルに従い, 顧客は時刻 $t+n$ でシーン s_t からシーン s_{t+n} に遷移する. 遷移に要する時間 n 等の遷移条件は, STN のトランジションに記述する.

4.4.4 顧客が得る利益と報酬

顧客は時刻 t において, 提供されているサービスのイベントの内容を評価する. 本提案手法では, この単位時間ごとの評価値は強化学習における報酬に相当し, そのイベントの評価が顧客にとってプラスの評価 (喜びや金銭的利益など) である場合は正の値の報酬, マイナスの評価 (精神的苦痛や金銭的損失など) である場合は負の値の報酬としてスカラー値 r_t を顧客が受け取ることとする.

4.4.5 利益の予測と行動価値関数

顧客は, 個々のシーン s において行動 a を実行した場合の報酬和の期待値を計算する関数を保持する. この関数は, 行動価値関数 $Q(s,a)$ に相当する. 顧客は, この行動価値関数をサービスの経験により更新する. 本提案手法では, 行動価値関数 $Q(s,a)$ を単位時間ごとに Q-learning のアルゴリズムを用いて更新することとする. また, 行動価値関数を用いた行動選択には, 前述の Boltzmann 選択のアルゴリズムを用いる. なお, 価値関数の更新式と行動選択の計算式は, 各シーンにそれぞれ記述する.

4.4.6 強化学習パラメータ

強化学習のパラメータとしては, 先に述べた学習率 α, 割引率 γ, 温度定数 T がある. Doya は, 上記のパラメータは生物の脳の神経伝達物質に対応しているという仮説を立てており [Doya 00], 学習率はアセチルコリン, 割引率はセロトニン, 温度定数の逆数はノルアドレナリンに対応するとしている. また, TD-error はドーパミンに対応しており, 報酬予測からの増減を表現している. TD-error の値は報酬の値で調整することが可能であることから, 上記 3 つのパラメータと各シーンにおいて顧客が受け取る報酬の値はを適切に設定することにより, 個々の顧客の性質を表現することが可能である. サービス設計者は, サービスのターゲットとなる顧客に応じてこれらのパラメータの値を適切に設定し, シミュレーションを行うことにより, 様々な顧客像を想定したサービスの分析や評価を行うことが可能となる. パラメータ値の決定法としては, 顧客アンケートによって決定する方法や, 既存のサービスをモデル化し, 実際の顧客の挙動と一致するように調整する方法などが考えられる.

4.5 シミュレーションの手順と評価方法

以下に, 本手法によりサービスのシミュレーションを行うまでの手順を示す.

1. 3 章で説明した佐藤らの手法を用いて, STN によってサービスのモデリングを行う
2. 顧客のアクタ変数として, 行動価値関数によって計算される全てのシーン-行動の組の行動価値を定義する
3. さらに, 顧客のアクタ変数として強化学習のメタパラメータ α, γ, T と各シーンでの獲得報酬を定義する
4. 顧客のアクタが遷移可能な全てのシーンに, Q-learning の行動価値関数の更新式及び Boltzmann 選択のための行動選択確率の計算式を記述する
5. 各トランジションに状態遷移の条件を記述する
6. シミュレーションを実行し, 顧客のふるまいを観察することにより, 設計したサービスの分析・評価を行う

シミュレーション結果を分析することにより, 設計者が意図したように顧客が行動しているか (途中でサービスの受給を中断したり, 競合サービスを選択してしまう顧客が多くないかなど) を評価することが可能となる. また, 本手法は STN GUI Simulator によるアニメーションによって, 視覚的にサービスの流れを確認することを前提としているため, 顧客の行動が思わしくない場合は設計したサービスのどこに問題があるか, また複数の顧客同士, または顧客とサービス提供者の相互作用により, サービスの流れを阻止する状況が発生しないかなどを容易に確認することが可能である. また, 顧客アクタの行動価値の値を計測し, この値を顧客満足度と関連付けてサービスの評価を行うことも可能である.

5. 例題

本章では, 例題を用いて本提案手法の有効性を検証する. ここでは, オンラインレンタル DVD 配送サービスにおける顧客の意思決定シミュレーションを行う. 図 1 に, STN GUI Simulator で作成した例題のネットワークを示す. 例題のサービスのモデルでは, 顧客はまず (1) レンタル業者のホームページを観覧し, (2) レンタルを希望する DVD を検索し, 在庫がある場合は (3) 注文/支払手続きを行うが, 在庫がない場合は (4) サービスを受けることを中断する. (3) の手続きが完了すると, (5)DVD が配送されるまで待ち状態となる. DVD の到着後, 顧客は (6)DVD を受け取り, 鑑賞する. 鑑賞終了後, (7) 返却手続き (DVD をポストに投函) を行い, 本サービスの一連の流れが終了する. (1)〜(7) のサービスのイベントは, モデル上ではシーンとして表現される. 本来のサービスフローシミュレーションでは, 顧客だけではなくサービスプロバイダのネットワークも構築し, 顧客とプロバイダの相互作用を検証するが, 本検証では顧客のネットワークのみの表現に簡略化している. 本検証では, サービスの競合業者 2 社 (A 社及び B 社) が存在すると仮定し, 顧客はそれぞれの業者のサービスを受けることで得られる価値を予測し, どちらの業者を選択するかの意思決定を行う. 顧客はそれぞれのサービスを繰り返し受けることにより, 価値を予測する価値関数を学習していくが, ここでは顧客は特定の個人 1 人ではなく複数の人間であるとし, 世間

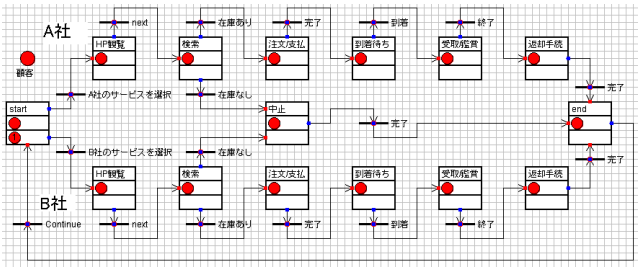

図 1: 競合する 2 社のオンライン DVD 配送サービスの STN

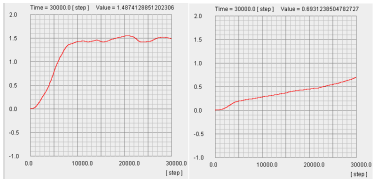

図 2: A 社を選択する行動価値 (左) と B 社を選択する行動価値 (右) の時間的推移 ($p_A = 0.9, p_B = 0.7, d_A = d_B = 1$)

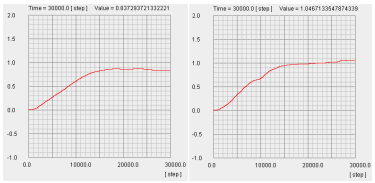

図 3: A 社を選択する行動価値 (左) と B 社を選択する行動価値 (右) の時間的推移 ($p_A = 0.9, p_B = 0.7, d_A = 3, d_B = 1$)

の風評や口コミなどで一般的な顧客の価値関数が学習されることを仮定した. A 社と B 社は基本的に同様のサービスを提供するが, 顧客がレンタルを希望する DVD の在庫がある確率 (A 社, B 社それぞれ p_A, p_B) と注文手続き完了から DVD が顧客の自宅に到着するまでの日数 (d_A, d_B[days]) は異なるとした. 顧客は各シーンにおいて, 利益または不利益を報酬として受け取る. 報酬は, 希望する DVD の在庫がある場合に+1, ない場合は-1, 到着までの待ち状態において, 1 日あたり-0.5, DVD を受け取り, 鑑賞が可能となった状態で+2 とした. 顧客の強化学習モデルのパラメータは, ここでは任意の値として $\alpha = 0.01, \gamma = 0.9, T = 0.5$ とした. 検証では, 各業者のパラメータ p_A, p_B, d_A, d_B が変化したとき, 顧客はどのような意思決定を行うかを観察した. まず, $p_A = p_B = 0.9$, $d_A = d_B = 1$ とそれぞれ等しい場合は, 顧客の温度定数 T の値によっては 1 つの業者を若干多く選択する結果になる傾向もみられたが, 平均的には 2 つの業者を等確率で選択する結果となった. 次に, 在庫がある確率を $p_A = 0.9, p_B = 0.7$ として実験を行った. 顧客がスタートのシーンから業者 A を選択する行動の価値 (Q 値)Q_A と Q_B の時間的変化の典型的な一例を図 2 に示す. この図から, 在庫確率が高い A 社を選択する報酬の予測値が高くなっていることがわかる. このときの A 社と B 社の選択比率は, 約 5.3:1 であった. 次に, 在庫確率を $p_A = 0.9, p_B = 0.7$ としたまま, DVD の到着までの日数を $d_A = 3, d_B = 1$ として実験を行った. このときの Q_A と Q_B の時間的変化の一例を図 3 に示す. この図から, 在庫確率は低いが待ち時間が短い B 社の報酬予測値が高くなっていることがわかる. このときの A 社と B 社の選択比率は, 約 1:1.3 であった. この結果は, 顧客が先に待ち時間が短い B 社を優先的に選択し, 在庫がない場合は A 社に変更する, という意思決定を行っているものと考えられる. このように, モデル化したサービスの設計値の違いにより, 異なった行動選択を行う顧客の意思決定シミュレーションを行うことが可能であることを確認した. また, 本論文では結果の詳細は省略するが, 顧客の意思決定の傾向を決定づけるパラメータ α, γ, T を変化させることにより, 異なる性格を持つ顧客のシミュレーションを行うことが可能であることを併せて確認している.

6. おわりに

本論文では, Sutton らの強化学習モデルを用いて顧客の行動選択モデルを構築し, STN を用いたサービスシミュレーション上に実装することにより, 顧客の意思決定シミュレーションを行い, 設計したサービスの分析・評価を行う手法を提案した. シミュレーション結果と実世界のマーケットとの整合を実現するためには, 適切な顧客の強化学習パラメータの同定が必要となるが, サービス工学に基づく顧客の満足度関数の同定方法 [吉光 07] やアンケートなどを導入することによりこれらの課題を今後解決する予定である.

謝辞

本研究の一部は、科学研究費基盤 B「サービス評価をするために連続数理表現を導入したサービス設計支援システム」の支援を得て実施した。

参考文献

[Doya 00] K. Doya: Metalearning, Neuromodulation, and Emotion, Affective Minds, pp.101-104, Elsevier Science B.V. (2000).

[Fisk 05] R. P. Fisk, et al.: サービスマーケティング入門, 法政大学出版局 (2005).

[川田 93] 川田誠一, 川田尚吾, 渡辺敦: 場面の概念を用いた離散連続混合システムのシミュレーションモデル, 日本機械学会論文集 C 編, Vol.59, No.563, pp.10-16 (1993).

[茂木 06] 茂木健一郎, 田中洋: 欲望解剖, 幻冬舎 (2006).

[佐藤 07] 佐藤友亮, 鈴木遼, 原辰徳, 下村芳樹, 新井民夫: サービス工学に基づくサービス CAD システムの構築 (第 36 報) -サービス・マーケティング手法と場面遷移ネットに基づくサービスフロー・シミュレーション-, 2007 年度精密工学会春季大会学術講演会講演論文集, pp.941-942 (2007).

[Schultz 97] W. Schultz, P. Dayan and P. R. Montague: A Neural Substrate of Prediction and Reward, Scinence, 275, pp.1593-1599 (1997).

[下村 05] 下村芳樹, 原辰徳, 渡辺健太郎, 坂尾知彦, 新井民生, 富山哲男: サービス工学の提案 - 第 1 報, サービス工学のためのサービスモデル化技法 -, 日本機械学会論文集 C 編, Vol.71, No.702, pp.315-322 (2005).

[Sutton 98] R. S. Sutton and A. G. Barto: Reinforcement Learning An Introduction, MIT Press (1998).

[Tateyama 07] T. Tateyama, S. Kawata and Y. Shimomura: Development of Scene Transition Net(STN) GUI Simulator for Discrete-Continuous Hybrid Systems, In Proceedings of the 7th Japan Korea Workshop on CAD/CAM -Design Engineering Workshop-, Tokyo, pp.82-87 (2007).

[Watkins 92] C. J. C. H. Watkins and P. Dayan: Technical Note: Q-learning, Machine Learning, 8, pp.55-68 (1992).

[吉光 07] 吉光陽平, 木見田康治, 原辰徳, 新井民夫, 下村芳樹: サービス工学に基づくサービス CAD システムの構築 (第 34 報) -受給者視点の属性重要度に基づくサービス評価モデル-, 2007 年度精密工学会春季大会学術講演会講演論文集, pp. 937-938, (2007).