

テーマ番号	EP075	
プロジェクト テーマ	和文 英文	深層学習に基づく建築受注を決定づける顧客属性の発見に関する研究 Research on the Discovery of Customer Attributes to Determine Building Orders Based on Deep Learning
プロジェクト メンバー		指導教員 長田 茂美 教授
	4EP2-06 石田 康太郎 (Kotaro Ishida) 4EP1-26 谷 尚弥 (Naoya Tani)	

Abstract 本研究では、過去の住宅イベントでの来場者が回答したアンケートデータと、その来場者が建築受注のどの段階までに至ったかのデータセットを活用し、深層学習を用いることで、新たな住宅イベントの来場者のアンケートデータから、その来場者が建築受注のどの段階まで至るかを推測するシステムを提案する。また、学習済みモデルを使用して、その出力に効く入力パターンを求め、解析することで、建築受注を決定づける顧客属性の発見を実現する。

Keywords knowledge discovery, deep learning, multilayer perceptron, backpropagation, iterative inversion, sensitivity analysis.

1. まえがき

一般に、建築における請負契約は、いくつかの段階を経て締結に至るが、締結に至るか否かを大きく左右する要因として、顧客属性が挙げられる。そのため、多くの建築会社では、顧客属性と請負契約との関係性を把握するために、住宅イベントを開催し、来場者にアンケートに回答してもらうことにより、この関係性の分析に努めている。

本研究では、顧客属性と請負契約との関係性を効率的に把握するために、過去の住宅イベントでの来場者が回答したアンケートデータと、その来場者が建築受注のどの段階までに至ったかのデータセットを活用し、深層学習を用いることで、新たな住宅イベントへの来場者が回答したアンケート結果から、その来場者が建築受注のどの段階まで至るのかを分類・推測するシステムを提案する。また、この学習済みモデルを使用して、その出力に効く入力パターンを求め、解析することで、建築受注を決定づける顧客属性の発見を実現できるシステムについても提案する。

2. システム概要

2.1 建築受注段階の推測システム

図 1 に、住宅イベントへの来場者が回答したアンケートデータから、建築受注のどの段階までに至るのかを推測するシステム、すなわち、アンケートデータを建築受注のいくつかの段階に分類するシステムの概要を示す。本システムでは、7 層からなる多層パーセプトロン (MLP, Multilayer perceptron) を 2 段に配置し、1 段目の MLP で、アンケートデータを chance(受注の見込みのある案件) および hopeless(受注の見込みのない案件) の 2 種類のカテゴリに大分類し、次に、1 段目の MLP で chance に分類されたアンケートデータを、2 段目の MLP で chase(営業活動の継続段階), meeting(初回打合せ段階), design(設計契約締結段階), estimate(見積作成段階), contract(成約), lost(営業活動中に逸失) の 6 種類のカテゴリに詳細分類するという構成を探っている。なお、MLP への入力は、アンケートデータを数値化した 218 次元の 0, 1 パターン、出力は各カテゴリに対応した 6 次元の 0, 1 パターンである。

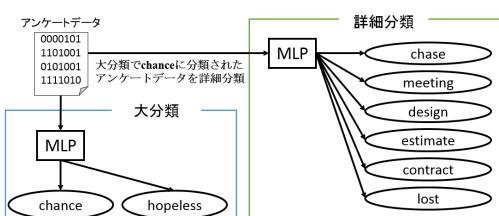


図 1 本システムの概要
Fig.1 Overview of the system.

2.2 顧客属性の発見システム

請負契約の締結に至る顧客属性を発見するために、上述した建築受注段階の推測システムの学習済みモデルに iterative inversion 法^[1]を適用し、「出力に効く入力データ項目」を求める、解析する。これにより、建築受注を決定づける重要な顧客属性を発見する。

3. システム評価

3.1 建築受注段階の推測システム

本システムの有用性を確認するために、建築受注の段階を表すカテゴリが既知であるアンケートデータを数値化したデータ 579 件（訓練データ：387 件、テストデータセット：192 件）を使用し、chance の 6 種類と hopeless の 1 種類の計 7 種類の詳細分類カテゴリに分類する評価実験 I を行った。評価実験 I では、1 段目の MLP にアンケートデータと大分類カテゴリとの関係を、2 段目の MLP にはアンケートデータと chase, meeting, design, estimate, contract, lost の詳細分類カテゴリとの関係を学習させた後、テストデータセットを用いて、7 種類の詳細分類カテゴリへの分類を試みた。表 1 に、訓練データセットとそのアンケートデータの件数を、表 2 に、テストデータに対する分類結果を示す。

表 1 カテゴリと訓練データセット
Table 1 Category name and training dataset.

大分類カテゴリ	詳細分類カテゴリ	訓練データ数
chance	chase	153 件
	meeting	4 件
	design	1 件
	estimate	1 件
	contract	7 件
	lost	52 件
hopeless	hopeless	169 件

表 2 評価実験 I の結果
Table 2 Result of evaluation experiment I.

カテゴリ名	正解率	カテゴリ名	正解率
chase	53/76	contract	0/3
meeting	0/2	lost	6/26
design	0/1	hopeless	65/83
estimate	0/1		

評価実験 I の結果、1) 建築会社から提供いただいた訓練データ数が少なく、十分な分類精度の実現が難しいこと、また、2) 訓練データ数が少ないカテゴリほど分類精度も低下することがわかる。この原因を分析してみると、1) chase と meeting の特徴を捉えきれず、meeting のデータを chase のカテゴリに誤分類する、2) design, estimate, contract

の特徴を捉えきれず、これらのデータが *chase* と *hopeless* 以外のカテゴリに誤分類される傾向が見られた。

この誤分類の問題を解決するため、表 3 に示すように、詳細分類カテゴリの再定義を行った。初回打合わせ段階を表す *meeting* は、営業活動継続段階を表す *chase* とあまり意味合いが変わらないという理由から、*chase* のカテゴリにまとめた。また、設計契約段階を表す *design* と見積作成段階を表す *estimate* は、成約を表す *contract* の特徴を多く有することから *contract* のカテゴリにまとめ、*agreement*(設計、見積作成、成約など何らかの契約に至る段階)とした。

このように再定義した後のカテゴリに対して、本システムの有用性を確認するために、表 1 に示すカテゴリが既知のデータセットを使用して、*chance* の 3 種類と、*hopeless* の計 4 種類のカテゴリに分類する評価実験 II を行った。

評価実験 II では、1 段目の MLP にアンケートデータと大分類カテゴリとの関係を、2 段目の MLP にはアンケートデータと *chase*, *agreement*, *lost* の詳細分類カテゴリとの関係を学習させた後、テストデータセットを用いて、再定義した詳細分類カテゴリへの分類を試みた。表 1 に、訓練データセットとそのアンケートデータの件数を、表 4 に、テストデータに対する分類結果を示す。

表 3 再定義した詳細分類カテゴリ

Table 3 Redefinition category.

大分類カテゴリ	詳細分類カテゴリ	元の詳細分類カテゴリ
chance	<i>chase</i>	<i>chase</i>
		<i>meeting</i>
		<i>design</i>
	<i>agreement</i>	<i>estimate</i>
		<i>contract</i>
		<i>lost</i>
hopeless	hopeless	hopeless

表 4 評価実験 II の結果

Table 4 Result of evaluation experiment II

カテゴリ名	正解率	カテゴリ名	正解率
<i>chase</i>	65/78	<i>lost</i>	9/26
<i>agreement</i>	1/5	<i>hopeless</i>	61/83

評価実験 II の結果、*design*, *estimate*, *contract* を *agreement* のカテゴリとしてまとめたことで、提案システムが設計、見積作成、成約などの契約に至る段階としての特徴をうまく捉えることができ、分類精度の向上が確認できた。しかし、依然として訓練データ数の少ない *agreement*, *lost* の分類精度は、著しく低いことがわかる。結論として、各々の詳細分類カテゴリの特徴を正確に捉えられるように訓練データセットを収集できれば、これらの分類精度の向上は十分に期待できるものと考える。

3.2 顧客属性の発見システム

MLP への入力は、アンケートデータのチェックの付いた回答項目を 1、チェックの付いていない回答項目を 0 とした 218 次元の 2 値パターンである。ここでは、評価実験 II で用いた MLP の学習済モデルに iterative inversion 法を適用することにより、*agreement* を表す MLP の出力パターンを出力する(*agreement* に至るであろう)MLP の入力パターンを求める。iterative inversion 法による MLP の入力パターンの更新が収束した時点での入力値を見ると、“1 に近ければ近いほど、その回答項目にチェックを付ける方が *agreement* に効く”と解釈でき、“0 に近ければ近いほど、その回答項目にはチェックを付けない方が *agreement* に効く”と解釈できる。同様に、他のカテゴリについても、そのカテゴリに効く回答項目を発見できる。図 2 に、*agreement* に対して、iterative inversion 法により求めた回答項目ごとの入力値を示す。*agreement* に効く入力パターン

を求める際には、選択式の單一回答の設問において、iterative inversion 法の適用により、複数回答となるのを防ぐため、入力値を設問ごとに最大値を 1、その他を 0 とする処理を加えている。図 3 に、この処理を加えた結果を示す。この入力パターンが *agreement* に至る理想の回答であると解釈できる。また、図 4 に、iterative inversion 法で求めた勾配値 (MLP の学習済モデルの出力パターンと *agreement* を表す出力パターンとの誤差の入力変数に関する微分値) を用いて、設問ごとの勾配の平均値を求めた結果を示す。この勾配の値は、設問単位の誤差への影響度(感度)を表す指標であり、その絶対値が大きいほど、MLP の学習済モデルに *agreement* を出力させるために大きな影響力をもつ設問であると考えることができる。

評価実験では、訓練データ数が少なく、重要な顧客属性の発見には至っていないが、1) 年齢を問う設問は、*agreement* に影響力をもたない、2) *agreement* に至った多くの顧客が同じ回答を選択している設問でも、*agreement* への影響力をもたない、などといった結果が得られた。2) については、*agreement* 以外のカテゴリに属する顧客の多くが同じ回答を選択しているなどの傾向が見られており、訓練データセットの拡充が図れれば、提案システムは、重要な顧客属性の発見に有効にものとなろう。

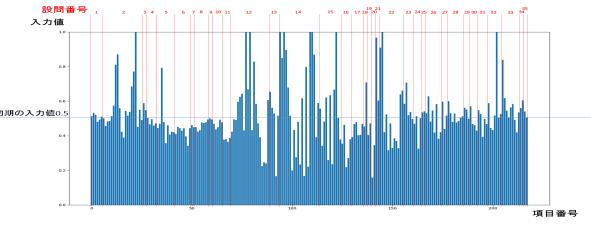


図 2 回答項目ごとの入力値

Fig.2 Input value for each reply.

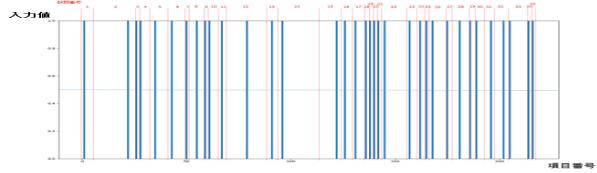


図 3 設問ごとの入力値の最大値

Fig.3 Maximum value for each questions.

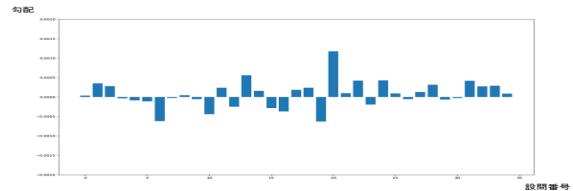


図 4 設問ごとの勾配の平均値

Fig.4 Average value of gradients for each question.

4. むすび

本研究では、深層学習を用いて、住宅イベントでの来場者アンケートから、その回答者が最終的に建築受注のどの段階まで至るかを推測するシステムと、建築受注に有効に作用する顧客属性を実現するシステムを提案した。今後も引き続き、評価・改良を継続し、実用的なシステムへと精度向上を図っていく予定である。

最後に、住宅イベントの来場者アンケートデータをご提供いただいた(株)丸屋建設殿に感謝いたします。

参考文献

- [1] 長田茂美, 増本大器, 山川宏, 木本隆, “ニューラルネットによる感覚運動融合の階層型モデル,”日本ロボット学会誌, Vol. 12, No .5, pp. 685-694, 1994.