

Web-Based プロダクションシステムにおける自習ログの分析

Log Analysis of Self Study on Web-based Production System

森田純哉*1 三輪和久*2 中池竜一*3 寺井仁*4
Junya Morita Kazuhisa Miwa Ryuichi Nakaike Hitoshi Terai

*1北陸先端科学技術大学院大学 Japan Advanced Institute of Science and Technology
*2名古屋大学 Nagoya University
*2京都大学 Kyoto University
*2東京電機大学 Tokyo Denki University

The present study investigates learning process in cognitive science introductory course. The participants who took the course engaged three sequential tasks of cognitive modeling mainly at home. They used the web-based production system architecture that implements supporting features for development of cognitive models. We analyzed the requests that were sent to our server during the period of the course (71 days). As a result, we confirmed (1) increasing of the number of the requests from the first task to the last task, (2) changes in the strategy of cognitive model developments through the three tasks, and (3) significant correlations between the strategy of cognitive model developments and subjective evaluations on the results of the final task.

1. はじめに

計算機による認知機能のモデリングは、認知科学の進展を促す重要な研究アプローチである。特に、人間の知識を明示的で実行可能なルールとして表現するプロダクションシステムは、現在に至るまで数々の研究において用いられてきた。たとえば、近年の認知科学研究において、盛んに用いられる Soar (state-of-the-art computational theory of the mind [Newell 94]) や ACT-R (Adaptive Control of Thought-Rational [Anderson 04]) は、アーキテクチャの根幹にプロダクションシステムを据えるものである。

しかしながら、近年の日本において、認知モデルを用いた研究は、それほど多くない [三輪 09a]。認知モデルに関する研究が沈滞している理由の 1 つは、認知モデルの構成に専門的なスキルが要求されることにあると考えられる。認知モデルの利点は、人間の認知の構造とプロセスを、厳密で完全な表現として実装することにある。しかし、その厳密さゆえに、認知モデルは、人間にとって理解が困難なものにもなりうる [Fum 07]。この困難さを克服するために、認知モデルのユーザビリティを高める研究が盛んに行われている。たとえば、Ritter は、認知モデルの直感的な記述言語や認知モデルの再利用を容易にする枠組みの構築を目指している [Ritter 09]。

認知モデル研究の促進に向け、著者らのグループは、上記を補完する別のアプローチを模索してきた。そのアプローチは、認知モデル教育の充実である。認知科学の初学者に対して、認知モデルの面白さを伝えつつ、そのスキルの向上を促すのである。このようなアプローチは、認知モデルのユーザビリティを高めるアプローチに比べ、ボトムアップ的、あるいは現場志向的である。そして、著者らは、このアプローチの具体的な実践として、教育用の認知モデル開発環境を構築し [中池 09]、それを利用する授業デザインを考案してきた [三輪 09b]。

本研究は、著者らの提案するアプローチに、データに基づく裏付けを与えようとする。つまり、授業参加者の振る舞いの分析により、開発された教育環境を用いた授業実践を評価する。具体的には、(1) 設定された授業デザインが有効に働いたか、(2) 学習者による認知モデリングのプロセスに変化が生じたか、

(3) 学習者のモデリングのプロセスは、最終的なパフォーマンスに反映されたかといった項目を取り上げる。これらの項目の検討を通し、これまでの実践の成果を確かめ、今後検討すべき問題を導く。

2. どこプロ

著者らが開発してきた認知モデル教育用の環境は、どこでもプロダクションシステム (通称: どこプロ) と呼ばれる [中池 09]。どこプロの特徴は、Web アプリケーションとしての実装にある。この特徴により、学習者は、インターネットを利用可能な環境にいれば、どこでも自習に取り組むことができる。また、この特徴ゆえ、どこプロの操作に関わる学習者の全ての振る舞いは、サーバへのリクエストとして記録されることになる。サーバに記録されたログは、認知モデルの学習における実情を示し、その後の授業デザインの改善に役立てられるだろう。本研究では、実際に、どこプロを利用した授業において記録されたログを分析する。以下では、そこでの分析と関連したどこプロの機能を示す。

モデルの構築 どこプロによる認知モデルの構築 (ワーキングメモリの定義、ルールの追加・編集・削除) は、Web ブラウザ上のフォームを介してなされる (図 1)。通常のプログラミングに用いられるエディタは必要ない。たとえば、図 1 の右側領域は、モデルが保持するルールを表示する。ユーザは、この中の任意の節をクリックし、ルールの編集を開始する。なお、どこプロ上で記述されたモデルを外部エディタへエクスポートすること、エディタで構築されたモデルをインポートすることも可能である。

モデルの実行 どこプロ上で構築されたモデルは、画面上のコントローラによって動作する。コントローラに配置されるボタンをクリックすることで、プロダクションシステムのサイクルが、1 ステップ進められる。その結果は、画面左上部に表示されるワーキングメモリ (WM) に反映される。

ヒントの閲覧 どこプロは、ヒントの生成により、手詰まりにいたる学習者をサポートする。ヒントは、コントローラ中央のボタンをクリックすることで、画面左下部のメッセー



図 1: どこプロの概観 [中池 09].

ジョウィンドウに表示される。ヒントは、その時点において発火可能なルールの一覧、各ルールについて構成された変数束縛の一覧を示す。これらの一覧を参照することで、学習者は、自身のモデルが稼働しない理由を発見できる。

3. 対象授業

情報科学関連大学院において実施された授業を対象とした [三輪 09b]。本節では、対象とした授業のコンセプト、実施された授業のスケジュールを示す。

3.1 授業デザインのコンセプト

授業デザインのコンセプトは、社会的相互作用の喚起である。認知モデリングのスキル獲得には、長時間にわたる自習が要求される。多くの学習者は、認知モデルの魅力を体験する前に、孤独な自習の段階で興味を失ってしまう。この問題をふまえて、どこプロは、認知モデルの自習を積極的にサポートすることを狙ってデザインされた。この授業は、どこプロによる自習のサポートを効果的なものとするために、学習者間での社会的相互作用を導入した。具体的には、演習の各段階において、各学習者が構築したモデル同士が競い合う場を設けた。そのことにより、学習への興味を持続させ、学習者の知識を明示的なものとするを試みた。

3.2 授業スケジュール

授業参加者は大学院修士課程 14 名であった。多くの学生は、プログラミング経験を有したが、プロダクションシステムのモデリングに関する経験はなかった。参加者間の社会的相互作用は、以下 6 つのフェーズにおいて、段階的に導入された。

1. *Introduction*: 参加者は、プロダクションシステムの概要のレクチャーを受講した。その後、[Larkin 87] を参考にした Pulley 問題が与えられ、それを解くモデルを構成した。Pulley 問題の例を図 1 に示す。

2. *Initial Task*: 参加者 1 人ひとりに、オリジナルな Pulley 問題を生成することが求められた。図 1b はその例を示す。さらに、その自分が作った問題を解くモデルを構成した。この初期段階で構成されるモデルを、Initial Model と呼ぶ。生成された問題は回収され、それらの問題からなる問題集が作成された。

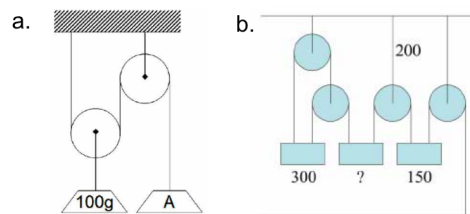


図 2: Pulley 問題の例 (a. 基礎的な問題, b. 参加者により構成された問題)。

3. *Initial Model Challenges*: 問題集が学習者にフィードバックされ、Initial Model がそれらの問題を解くコンテストがひらかれた。

4. *Revised Task*: 提示された問題集中の可能な限り数多くの問題を解くことができるように、モデルの改良が求められた。改良されたモデルを、Revised Model と呼ぶ。そして、再び、Revised Model が問題を解くコンテストが開催された。

5. *Final Task*: 問題を自由に設定し、それを解決するモデルを構成することが求められた。この段階で構成されるモデルを Final Model と呼ぶ。参加者が設定した問題は、ルービックキューブ、宣教師と人食い問題、囚人問題などであった。このうち、ルービックキューブを扱った参加者は、170 のルールを実装し、1000 ステップで 6 面をそろえるモデルを構築した。

6. *Presentation of Final Models*: Final Model のプレゼンテーションをする発表会を実施した。発表会では、全参加者が、プレゼンテーションによって示された全てのモデル (自分自身のモデルも含む) を評定した。評定の項目は、独創性、エンターテインメント性、技術性、学術性の 4 つとした (5 段階評定)。

4. 分析と結果

授業期間 (71 日間) における、どこプロのログを分析した。分析の対象は、最終課題を提出した 13 名の参加者である。分析の焦点は、(1) 導入された社会的相互作用が、参加者の動機付けに及ぼす影響、(2) 認知モデルを構築するプロセスの、学習の進行による変化、(3) 操作ログと構築されたモデルのパフォーマンスの関係を検討することであった。

4.1 社会的相互作用が動機付けに及ぼす効果

学習者の動機付けを測る指標として、自習の量に着目した。まず、授業の進行に伴う自習量の推移を、サーバへのリクエストの累積数から確認した (図 3)。図中の縦線は授業の実施時間、図上部の三角形は 3 つの課題の提出時期を示している (3.2 節における 2, 4, 5 に対応)。この図から、リクエスト数の累積が、課題の提出時期と対応するように伸びていること、授業時以外に、多くのリクエストがサーバへ送られたことがわかる。すなわち、対象授業において、学習者が多くの自習を行ったことが示された。

続いて、3 回の課題のそれぞれにおける学習者の動機付けを検討した。図 4 は、各課題の遂行期間におけるどこプロへのアクセス時間、リクエスト数の平均値を示す。なお、アクセス時間は、複数のセッション (操作の連続した時間) の合計により算出した。操作の連続は、リクエストの時間間隔が 60 分以内にある場合と定義した。この図から、社会的相互作用の繰り返しに伴い、どこプロへのアクセス時間、操作回数が増加した。

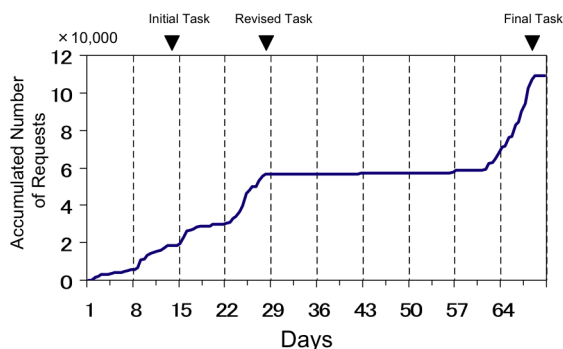


図 3: どこプロへの累積リクエスト回数．グラフ中の縦線は，授業が実施された時間帯を示す．

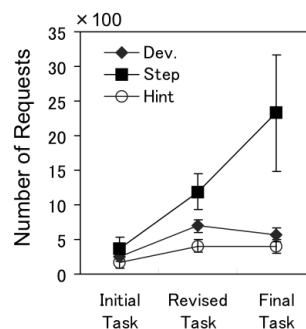


図 5: 操作タイプ別のリクエスト数．エラーバーは標準誤差を示す．

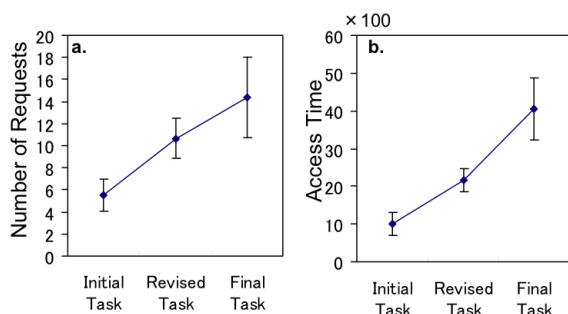


図 4: 各課題における操作量 (a. アクセス時間, b. リクエスト回数)．エラーバーは標準誤差を示す．

ことがわかる．

このことを確かめるために，アクセス時間，リクエスト数を従属変数とし，課題を独立変数とする被験者内分散分析を実施した．結果，アクセス時間とリクエスト数の両者で，課題の主効果が有意となった [アクセス時間: $F(2, 24) = 7.73, p < .01$, リクエスト数: $F(2, 24) = 2.97, p < .01$]．LSD 法による多重比較の結果，Initial Task から Final Task へかけて，アクセス時間とリクエスト数が増加したことが確かめられた．さらに，リクエスト数については，Initial Task と Revised Task における差が有意となった (いずれも $p < .05$) ．

著者らは，これら結果を，社会的相互作用の累積的效果によるものと考え．つまり，参加者は，社会的相互作用の機会を重ねるごとに，徐々に認知モデルの面白さに気づき，認知モデリングに没頭していったのだと解釈する．

4.2 認知モデリングのプロセスにおける変化

どこプロによる認知モデリングのプロセスを探るために，2節で示した3つの機能の使用状況を検討した．つまり，サーバ上のログから，モデルの構築に関わるリクエスト (Dev と表記)，構築されたモデルの実行に関わるリクエスト (Step と表記)，ヒントの閲覧に関わるリクエスト (Hint と表記) を抽出し，操作タイプを定義した．そして，学習の進行に伴い，それぞれの操作タイプの頻度がどのように増加，あるいは減少したのかを検討した．

図 5 は，3つの課題におけるリクエスト数を，操作タイプに分けて示す．学習の進行に伴うリクエスト数の増減は，3つの操作タイプで異なっているように見える．その詳細を検討するために，3 [操作タイプ (被験者内)] × 3 [課題 (被験者内)] 分散

分析を実施した．結果，操作タイプの主効果 [$F(2, 24) = 5.28, p < .01$]，課題の主効果 [$F(2, 24) = 5.08, p < .01$]，操作タイプと課題の交互作用 [$F(4, 48) = 3.20, p < .05$] が有意となった．さらに，Step と Dev における課題の単純主効果が有意となった [Step における課題の効果: $F(2, 24) = 3.93, p < .01$, Hint における課題の効果: $F(2, 24) = 1.63, n.s.$, Dev における課題の効果: $F(2, 24) = 5.53, p < .01$]．そして，多重比較 (LSD 法) により，Initial Task から Final Task へかけて Step が増加したことが示された (ともに $p < .05$) ．

これらより，3つの課題における参加者のプロセスを推測することができる．まず，Initial Task でのモデリングは，相対的にヒントを多く閲覧する試行錯誤的なものと解釈される．その後の Revised Task では，Dev と Step が増加し，Hint の閲覧は，割合として減少した．このプロセスの変化は，試行錯誤的なモデリングから，ヒントの助けを必要としないモデリングへの変化とみることが出来る．そして，最後の Final Task では，Step の頻度のみが増加し，長い問題解決プロセスを伴う複雑なモデルが，最小限のルール修正によって構築されたと判断できる．

4.3 自習ログとパフォーマンスとの関連

ここまでの分析により，社会的相互作用による累積的な動機付け，学習の進展に伴うモデリングのプロセスの変化を確かめた．続いて，これらの傾向が，Final Task におけるモデルのパフォーマンスにどのように反映されたのかを検討した．パフォーマンスに関わる指標としては，Final Task のモデルに対する評定値 (3.2 節の 6) に着目した．

まず，評定値の信頼性を，ケンドールの一一致度係数によって検討した．結果，4つの指標全てにおいて，有意な一致度が得られた [独創性: $W = .253, \chi^2(12) = 30.64, p < .01$, エンターテインメント性: $W = .484, \chi^2(12) = 58.12, p < .01$, 技術性: $W = .571, \chi^2(12) = 68.58, p < .01$, 学術性: $W = .379, \chi^2(12) = 41.00, p < .01$] ．

ただし，上記の評定値は，プレゼンテーションに基づくもので，実際に構築されたモデルのパフォーマンスを正確に反映するとは限らない．構築されたモデルには，モデルを構築した本人のみがわかる問題もあるだろう．そこで，各モデルに対して，モデルを構築した本人が与えた評定値 (自己評定) と，他者が与えた評定値 (他者評定) を区別し，両者の相関を検討した．なお，他者評定については，本人を除く評定値の平均値を指標とした．結果，技術性の項目をのぞき，両者に有意な相関を認めることができなかった [学術性: $r = -0.12, n.s.$, 技術

表 1: 操作ログと評定の相関係数.

		Dev	Step	Hint
自己評定	学術	-0.01	0.19	-0.54
	技術	-0.11	0.86**	-0.29
	エンタメ	0.41	0.22	0.04
	独創性	-0.48	0.67*	-0.40
他者評定	学術	-0.18	0.27	-0.07
	技術	-0.32	0.51	-0.31
	エンタメ	0.02	0.42	0.09
	独創性	0.14	0.20	0.08

Note: **($p < .01$), *($p < .05$).

性: $r = 0.72$, $p < .01$, エンターテインメント性: $r = 0.36$, $n.s.$, 独創性: $r = 0.04$, $n.s.$]. この結果を受け, 以下の分析では, 自己評定と他者評定を別の指標として扱った.

表 1 は, パフォーマンスに関わる 8 つの指標と, 3 つの操作タイプの頻度との相関係数を示す. 表より, Step と自己評定における技術性, Step と独創性の間に有意な相関が得られた. この結果は, 複雑な問題解決プロセスを伴うモデルを構築した参加者ほど, 自身のモデルを技術的に優れ, 独創的なものと判断したことを示す. だが, 操作タイプと他者評定の相関が有意でなかったことをふまえれば, そのモデルの価値は, 他者には伝わらなかったといえる.

5. 総合考察

本研究では, 認知モデルの学習における自習ログを分析し, 社会的相互作用の導入に伴い, (1) 自習時の操作量が増加し, (2) モデリングのプロセスに変化が生じさせたことを確かめた. そして, (3) 自習ログに示されるプロセスの変化が, 最終的なパフォーマンスに対する自己評定と相関することを確認した. 以下では, これら 3 つの結果により示唆される授業実践の成功と, 認知モデル教育において検討すべき問題を議論する.

5.1 自習ログに示される授業実践の成功

上で述べた結果の (1) は, 分析対象とした授業実践の成功を示し, 社会的相互作用の導入による動機付けの有効性を示す. また, 結果の (2) として観察されたプロセスの変化は, 熟達化に関する認知科学研究と整合するものとみることができる. 熟達化に関わる多くの研究は, 経験の蓄積を通し, 試行錯誤的な問題解決プロセスが, ストレートフォワードなプロセスへ変化することを指摘する [Anderson 04, Larkin 80]. このような過去の熟達化研究との整合は, 対象授業において生じた学習を裏付けるものといえる. さらに, 結果の (3) では, ストレートフォワードなプロセスに至った参加者が, 自身の構築したモデルを高く評価したことも示された. これらから, 本研究で対象とした授業は, 少なくとも一部の参加者には, 十分な学習を促したと結論づけることができる.

5.2 自習ログに示される学習の問題

本研究では, 授業実践における成功だけでなく, 今後改善すべき問題も示された. 第 1 の問題は, 学習の個人差である. 図 4 の結果は, Final Task における自習量の増加を確かめた. だが, この図は同時に, Final Task における分散 (エラーバー) の大きさも示している. この分散の大きさは, Final Task において, 複数名の学習者が脱落したことを示唆する. この問題をふまえれば, 今後, 認知モデルの面白さを伝え, 学習を動機付ける更なる努力が要求されるということになる.

第 2 の問題は, パフォーマンスの評価に関わるものである. 表 1 では, 自習ログと他者評定に有意な相関はみられなかった. そもそも, 他者評定は, プレゼンテーションの巧妙さや, 扱う問題の表面的な面白さに影響されがちである. そのモデルが実際に価値があるかどうかを正確に評定できるとは限らない. 実際, 授業後に提出された Final Task のモデルを, どこプロ上でテストしたところ, 他者評価が相対的に高いものであっても, 実動しないものが存在した. その一方で, 学習者自身による自己評定にも, 客観性の問題があり, 常に信頼することはできない. このような評定のジレンマを踏まえれば, 今後, 認知モデルのパフォーマンスを, より客観的に評価する指標を開発する必要がある.

最後の問題は, 認知モデル学習のより根本的な評価に関わる. 冒頭で述べたが, 認知モデルは, 人間の認知のプロセスと構造を, 厳密で完全な表現として実装することに意義がある. つまり, 認知モデル学習において, 最も重視されるべきことは, 複雑な問題を解決するモデルの構築ではなく, 人間が暗黙的に採用するストラテジーの明示化である. 本研究では, このレベルのスキル獲得に関するデータを提出していない. しかし, 著者らは, 今回の授業参加者が, 複雑な問題を解決するモデルの構築のみを目指したとと思っているわけでもない. たとえば, Final Task の問題として, ルービックキューブを扱った被験者 (3.2 節を参照) は, 自身のモデリングを振り返り, 最終レポートに次のような記述を含めた. 「自分がそろえるときにはそう複雑な判断をしているつもりはないのであるが, 厳密に記していくとなかなか複雑な条件判断を行っていることに気づいた.」この記述は, 認知モデリングにおいて重視される上記の基準に合致するものである. 今後, このようなスキル (暗黙的知識の明示化) の獲得を促す方法を体系化し, 評価していかなければならない.

謝辞

本研究の一部は科学研究費補助金基盤 (B) (課題番号: 19300089) の助成による.

参考文献

- [Anderson 04] Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., and Qin, Y.: An integrated theory of the mind, *Psychological Review*, Vol. 111, pp. 1036–1060 (2004)
- [Fum 07] Fum, D., Missier, F. D., and Stocco, A.: The cognitive modeling of human behavior: Why a model is (sometimes) better than 10,000 words, *Cognitive Systems Research*, Vol. 8, No. 3, pp. 135–142 (2007)
- [Larkin 80] Larkin, J., McDermott, J., Simon, D., and Simon, H.: Model of competence in solving physics problems, *Cognitive Science*, Vol. 4, pp. 317–345 (1980)
- [Larkin 87] Larkin, J. H. and Simon, H. A.: Why a diagram is (sometimes) worth ten thousand words, *Cognitive Science*, Vol. 11, No. 65–100 (1987)
- [三輪 09a] 三輪 和久: 仮説演繹器・認知シミュレータ・データ分析器としての認知モデル, *人工知能学会学会誌*, Vol. 24, No. 2, pp. 229–236 (2009)
- [三輪 09b] 三輪 和久, 中池 竜一, 森田 純哉, 寺井 仁: 認知モデルの実装による認知科学の入門的授業実践, *人工知能学会第 55 回先進的学習科学と工学研究会資料*, SIG-ALST-A803-14, pp. 83–88 (2009)
- [中池 09] 中池 竜一, 三輪 和久, 森田 純哉, 寺井 仁: Web-Based プロダクションシステムの開発とその評価, *人工知能学会第 55 回先進的学習科学と工学研究会資料*, SIG-ALST-A803-14, pp. 77–82 (2009)
- [Newell 94] Newell, A.: *Unified Theories of Cognition*, Harvard University Press (1994)
- [Ritter 09] Ritter, F. A.: Two Cognitive Modeling Frontiers, *人工知能学会論文誌*, Vol. 24, No. 2, pp. 241–249 (2009)