

Instance-based learning in dynamic decision making

Gonzalez, C., Lerch, J. F., & Lebiere, C.
Cognitive Science 27 (2003) 591-635

Introduction

DDM - Dynamic decision making (Brehmer, 1990; Edwards, 1962)

- 複数の独立した現実的な時間的制約が存在
- 意思決定者の行為とは独立に変化する環境
- 意思決定者の行為の結果として変化する環境

DDM に関する過去の多くの研究(review Kershold & Raajimakers, 1997)

- 動的な環境でいかに人間がブアに振舞うか
- 時系列に従った意思決定者の学習(環境への適応)プロセスは検討されていない

これまでの著者達の研究

- 動的な環境の中での意思決定スキルの学習を検討
- 経験を積んだ意思決定者はヒューリスティックスをあまり使わない
- 時間が経つにつれて蓄積された知識を利用するようになる

本研究の目的

- 動的環境下での意思決定に適応しうる学習メカニズム・Instance Based Learning Theory IBLT の提案と計算機モデルによる理論の検証

2. 関連する理論 (学習) のレビュー

Simon の理論

学習に関与するメカニズム(Simon & Langley)

- 知識ベース：宣言的知識の蓄積
- 再認 (recognition)：オブジェクト間の弁別
- ストラテジー：適応的なプロダクションシステム
- 評価関数：問題解決における探索をコントロールする行為の評価

チャンキング理論 EPAM (Chase and Simon)

- チャンクの蓄積による学習
- エキスパートはチャンクへのリンクを短期記憶に入れる

チェス(DDM 課題の一つ)の領域

- ノービス：網羅的探索
- エキスパート：環境により注意がガイド．環境中の手がかりがチャンクを検索(再認)．小さな探索木により可能性が表現．チェスのエキスパートは自分がプレイした多くのゲームの重要な特徴を記憶

類似性，インスタンススペースの学習の理論

- 直面する状況の認識は記憶中の事例とのやり取りによってなされる (Hintman)

- 直面する状況は貯蔵された事例の類似性に基づいて分類される(Medin and Schaffer)
- 類似性は経験と文脈に左右される。類似性は比較を通して決定される(Medin, Goldston, and Gentner)
- 経験を積むにつれてインスタンスが蓄積され、アルゴリズムベースによる問題解決からインスタンスベースによる問題解決に移行(Logan)
- インスタンスの検索は類似性により媒介(Nosofsky & Palmeri)

AI 研究におけるインスタンスベースの学習

- 過去の類似事例の検索によって行われる意思決定=CBDT(Gibboa & Schmeidler)
- スキル学習が事例(状況, 解, 結果)の蓄積による = CBR(Reisbeck & Schank)
- 環境との相互作用を通して学習が進む = 強化学習

これらをまとめると...

事例の蓄積, パターン認知(再認)と選択的な探索の発達, 類似に基づく記憶検索, 記憶検索における注意の逐次的な減少, ルールベースから事例ベースへの移行

しかし, 動的な課題にこれらの理論が適用されることは稀だった

3. Instance-based learning in dynamic decision making

なぜ Instance based learning か

- Instance を知識の要素
- 動的環境下での意思決定には学習が関与

インスタンスの表現...SDU スロット (Situation Decision Utility)

- Situation...環境中の手がかり
- Decision...situation に適応しうる行動
- Utility...decision に関する評価

eg., パステルカラーのオイルペイント(商品 A)で部屋を塗った

((商品 A, oil, pastel), bought, buy)

IBLT が含む学習メカニズム

- Instance based knowledge : SDU の蓄積として表現される知識の蓄積
- Recognition based retrieval : 直面する環境と記憶中のインスタンスとの類似に基づく検索
- Adaptive strategies : ヒューリスティクスからインスタンスベースのストラテジーへ
- Necessity : 選択枝の探索をガイドするメカニズム
- Feedback updates : 一旦貯蔵された SDU の utility を変更

IBLT のステップ(Fig 1)

再認 (recognition)

- 環境にどのような選択枝があるかを認識
 - ◇ ノービスは環境中の重要な情報が分からず, ランダムに選択枝が認識
 - ◇ エキスパートは環境中の重要な情報に基づいて選択枝が整理されて認識

判断 (judgment)

- 環境中の手がかりから環境が典型的か非典型的か判断
 - ◇ 過去の記憶中にその環境と類似したものがあれば典型的

- ◇ 類似したものが無ければ非典型的
- 非典型的な場合はヒューリスティクスによる行動
 - ◇ ランダムヒューリスティクス：ランダムに行動を選択
 - ◇ タイムヒューリスティクス：制限時間の迫っている行動を先に行う
- 典型的な場合はインスタンス(SDU)により取りうる行動を判断
 - ◇ 類似した SDU の検索(CBR と同様のプロセス)
 - ◇ 活性化の程度により検索されやすい SDU が変化(Anderson & Lebiere)

選択 (choice)

- もっと選択肢を探すか、現在までに検討した選択肢の中で最も良いものを選択するかを判断
- necessity に応じてループ
 - ◇ satisfaction strategy(Simon & Langley)
 - ◇ DDM における満足度の見積りは残り時間による

フィードバック (feedback)

- 意思決定の結果、ポジティブなこと、ネガティブなことが起きたときに、記憶中の SDU の utility を更新
- 常にフィードバックが起きるわけではない
- 意思決定の結果はすぐにはわからない

4. CogIBLT: an ACT-R implementation of IBLT

ACT-R4.0 アーキテクチャ上に実装

- 多くの実証研究に支えられたアーキテクチャ
- IF-THEN ルール(プロダクション)
- チャンク

CogIBLT では両方の知識を利用．学習はチャンクの蓄積により生起

4.1. Dynamic decision-making task -Water Purification Plan (WPP) Fig. 2

WPP の実験課題としての特徴

- 水の分配システムのシミュレーター
- 現実場面のスケジューリング課題（郵便局の仕分けなど）と同型
- 1 時間以内の教示で理解可能．数分で 1 トライアルが終わる

WPP の構成要素

- タンク(0 ~ 22)：水を貯蔵する(各タンクに黒く記された部分が水量)
- チェーン：タンクの系列．各チェーンにはデッドラインが割り振られる
- ポンプ：それぞれのタンクに付属(右隣のタンクから毎分 0.5 ガロンの水を吸い上げる)

WPP の目標

- 全体のデッドライン(10:00)までに分配された水をタンク 0 に集める

WPP の制約

- 一度に開けるポンプの数は 5(開いているポンプは赤く表示，現在利用しているポンプ数は左

上に表示)

- 失った水の量がパフォーマンス(なるべく失う水を少なくする)
- 各チェーンのデッドラインまでに取り除くことができなかった水量は左上に表示
- タンクの水は突然増えることもある

WPP における意思決定

- ポンプを開ける/ポンプを閉じる

WPP における誰でも気づくヒューリスティクス

- タイムヒューリスティクス：デッドラインに近いポンプを開ける

4.2. Structure of SDUs and the decision process in CogIBLT

SDU が表現するもの

- 22 のタンク一つ一つに対応して SDU が存在
- SDU が集まることで各時間における WPP の状況が表現
- 各意思決定(ポンプを開けたか閉じたか)のたびにインスタンスとして記憶に貯蔵

SDU の構造

Time	Tank	Wateramt	Chain	Utility	Decision
138	0	5.0	4	162	Activate

Fig. 3. An example of a SDU for the WPP task. The situation is defined by: time, tank, water amount, and chain. The decision is to activate, and the utility is 162 min.

- Time：意思決定の時間
- Tank：タンク番号
- Wateramt：水量
- Chain：記載無し
- Utility：評価値
タイムヒューリスティクス/インスタンスにより割り振られる
タイムヒューリスティクスの場合はデッドラインまでの残り時間
- Decision：開いた(Activate)/閉じた(Deactivate)

プロダクション(Fig. 4)

ゴールにガイドされてプロダクションが配置 (発話プロトコルから仮定)

ゴール to meet deadline

- 環境から最も早いデッドラインを探し, current deadline とする
- 環境中の時間が current deadline を過ぎたら, 環境中の変化に基づいて, 選択された SDU の utility を変更(プロダクション 1)
- 次の current deadline を探す .

ゴール to make decision

Necessity レベル(デッドラインが差し迫っているか)によってプロダクションを選択

- デッドラインが差し迫っていない
プロダクション 2・3：評価するタンク(選択肢)の選択
- デッドラインが差し迫っている
プロダクション 4・5：現在までに評価された最良の選択肢を実行

ゴール to evaluate tank

選択されたタンクの utility を見積もる

- 過去に蓄積されたインスタンスとの類似度に応じて典型か非典型か判断
典型：過去の SDU を利用して見積もる(プロダクション 6)
非典型：タイムヒューリスティクスを利用して見積もる(プロダクション 7)

IBLT とプロダクションの関係, パラメータ

再認(Recognition)...環境中の手がかりから選択肢を認識

プロダクション 2・3

ランダムに選択/差し迫ったタンクから選択

判断(Judgment)...典型/非典型の判断, ストラテジーによる utility の評価

ゴール to evaluate tank にて行われる

典型か非典型かの判断...similarity rate

Highe(1.0)...完全に一致する SDU が存在したときのみ典型

Low(0.1)...常に典型と判断

典型的なとき(プロダクション 6)

類似した SDU を検索し, 検索された SDU によって utility の見積り

類似性は SDU の各スロットの重みにより計算

SDU の検索は ACT-R の活性化メカニズムにより影響

最近の SDU は検索されやすい

複数の SDU からの utility の見積りが行われる

Blending...類似性の高い SDU からより影響を受ける(Fig. 5)

非典型的なとき

各タンクの残り時間を計算し utility とする

選択(Choice)...選択肢の探索を継続するか, 意思決定を行うかを評価

to make decision にて Necessity に応じて行われる

necessity なし：全ての選択肢を評価してから意思決定

necessity あり：探索を打ち切るタイミングは変更可(1min ~ 8h)

フィードバック(Feedback)

プロダクション 1 にて行われる

現在の時間がデッドラインに到達するたびに起きる

失った水量が少なければ, その間に行った SDU の utility が向上

失った水量が多ければ, その間に行った SDU の utility が向上

フィードバックを受ける程度...learning rate

Highe(1.0)...常にフィードバックを受ける

Low(0.1)...常にフィードバックを受けない

Trial	Time	Tank	Water Amount	Chain	Decision	Utility
1	129	9	3.19	1	Activate	306
5	172	15	12.00	1	Activate	309
8	130	9	3.19	0	Activate	293

↓

Trial	Time	Tank	Water Amount	Chain	Decision	Utility
10	128	9	3.19	1	Activate	297.83

Fig. 5. Blending the utility into a new decision. The current situation in the bottom SDU compares to three SDUs from the past with activation above threshold. The utility of the current situation is calculated by Blending using the utility of the previous three SDUs.

5. シミュレーション実験

IBLT によって提案された理論的概念(再認, 判断, 選択, フィードバック)を検討

5.1. Human data collection

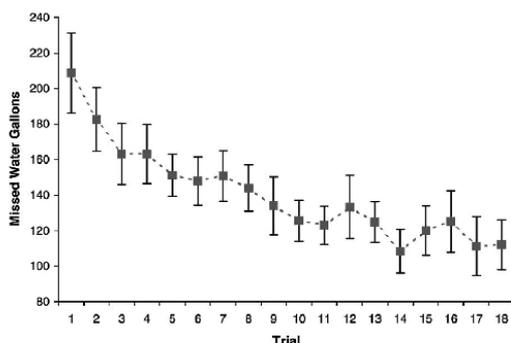


Fig. 6. Participants' average performance for 18 trials. The performance measure is the number of gallons of water missed. The overall mean is 140.55 and the average standard deviation is 20.85.

大学生 14 名が WPP を 18 試行(6 × 3 日)プレイ
1 試行の所用時間は 8 分
試行数の効果が有意 $F(1,17)=6.871, p < .001$

5.2. Data collection from CogIBLT

14 回のシミュレーション(人間の被験者数に対応)

心理実験とシミュレーションのデータのフィッティング

ピアソンの決定係数 r^2 学習曲線の傾きの重なり

RMSSD Root Mean Squared Standard Deviation

$$\frac{((data-model)/data SE)^2}{試行数}$$

人間とモデルのパフォーマンスの距離の平均を標準化したもの

実験 1: recognition process

再認の方法を操作したモデル 1 ~ 3

モデル 1

- 選択肢をランダムに検索
- SDU のユーティリティを time ヒューリスティクスに従って算出(プロダクション 6 はない... インスタンスからの utility の見積りをしない)

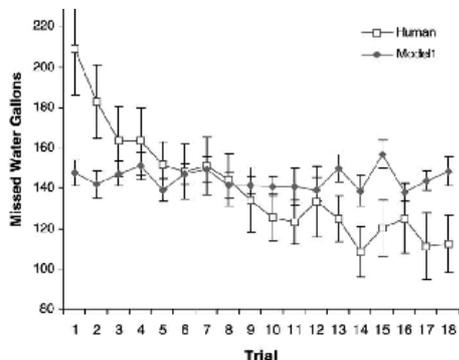
- Necessity なし(全ての選択肢を探索)・フィードバックなし

モデル 2

- 差し迫った選択肢を優先して評価

モデル 3

- はじめの 2 試行はランダム, その後に差し迫った選択肢を優先して評価



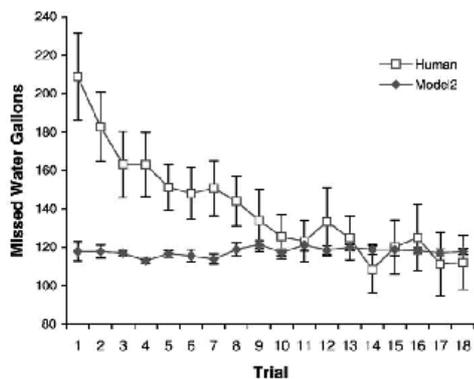
モデル 1 の結果

- $r^2 = 0.02$
- RMSSD = 1.64

初期は人間とモデルが類似

後期になるとモデルは人間より悪い

Fig. 7. Model 1. Random search of alternatives, utility determined by the time heuristic, no Blending, no necessity value, and no feedback mechanisms.



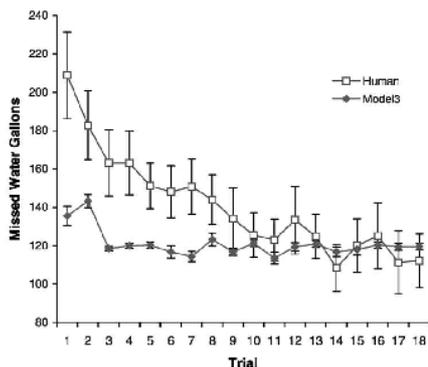
モデル 2 の結果

- $r^2 = 0.12$
- RMSSD = 2.02

初期が良すぎる(初期 2 試行のみの

RMSSD は 14.67)

Fig. 8. Model 2. Sorted search of alternatives, utility determined by the time heuristic, no Blending, no necessity value, and no feedback mechanisms.



モデル 3 の結果

- $r^2 = 0.46$
- RMSSD = 1.73

相関は高いが実際の値は違う

Fig. 9. Model 3. Random search of alternatives in the first two trials, sorted search after Trial 3. Utility determined by the time heuristic, no Blending, no necessity value, and no feedback mechanisms.

実験 2: Judgment

Judgment(Blending と similarity-rate)の効果を検討

モデル 4

- プロダクション 6 を付加:類似インスタンスによる utility の見積もり
- Similarity-rate は 1.0(完全に一致した SDU のみを検索)

モデル5

- Similarity-rate を 0.01(ほとんど全ての SDU を検索)

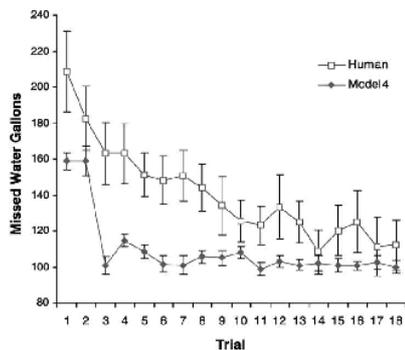


Fig. 10. Model 4. Random search of alternatives in the first two trials, sorted search after Trial 3. Utility determined Blending with one similarity-rate, no necessity value, and no feedback mechanisms.

モデル4の結果

- $r^2 = 0.64$
- RMSSD = 2.32

3-8 試行で人間と違う

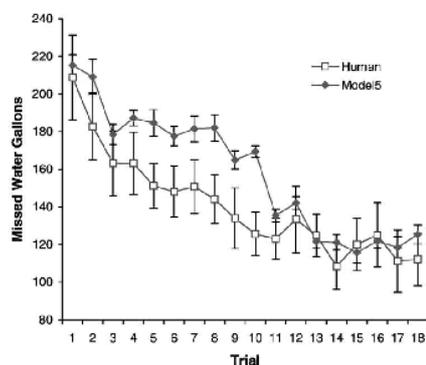


Fig. 11. Model 5. Random search of alternatives in the first two trials, sorted search after Trial 3. Utility determined Blending with 0.01 similarity-rate, no necessity value, and no feedback mechanisms.

モデル5の結果

- $r^2 = 0.81$
- RMSSD = 1.71

3-10 試行で人間よりも悪くなる

実験 3: choice

Choice におけるパラメータ(necessity)の操作

necessity = (1 分 , 1 時間 , 8 時間)

- 8 時間の場合は全ての選択肢を評価した後に実行
- 1 時間の場合はある程度の選択肢を評価した後に実行
- 1 秒の場合は十分な選択肢の評価が行われないうちに実行

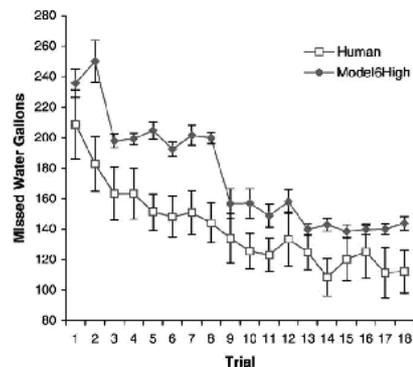


Fig. 12. Model 6-High. Random search of alternatives in the first two trials, sorted search after Trial 3. Utility determined Blending with 0.01 similarity-rate, 8 h necessity level, and no feedback mechanisms.

モデル6-8H

- $r^2 = 0.86$
- RMSSD = 2.62

人間よりも悪いパフォーマンス

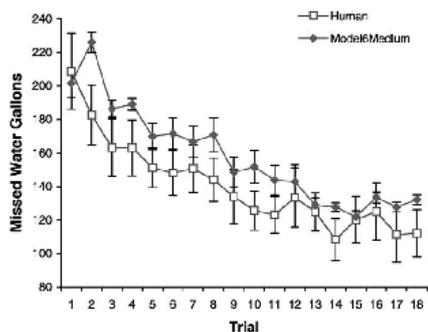


Fig. 13. Model 6-Medium. Random search of alternatives in the first two trials, sorted search after Trial 3. Utility determined Blending with 0.01 similarity-rate, 1h necessity level, and no feedback mechanisms.

モデル 6-1H

- $r^2 = 0.85$
- $RMSSD = 1.43$

大まかに人間と同程度のパフォーマンス

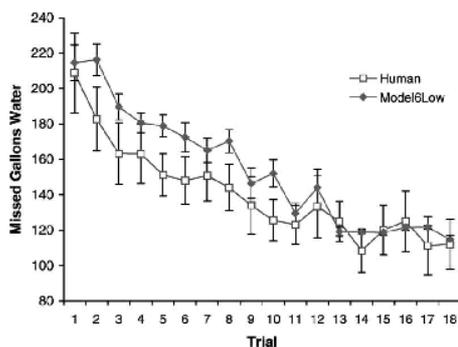


Fig. 14. Model 6-Low. Random search of alternatives in the first two trials, sorted search after Trial 3. Utility determined Blending with 0.01 similarity-rate, 1 min necessity level, and no feedback mechanisms.

モデル 6-1 分

- $r^2 = 0.90$
- $RMSSD = 1.26$

大まかに人間と同程度のパフォーマンス

実験 4: feedback

Learning rate=1.0 環境の変化に応じて記憶中の utility の値を更新

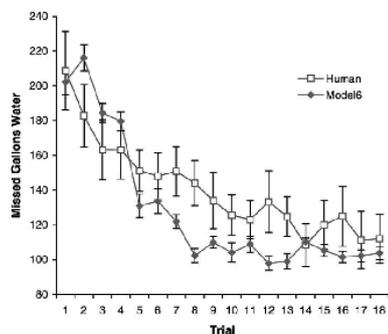


Fig. 15. Model 7. Random search of alternatives in the first two trials, sorted search after Trial 3. Utility determined Blending with 0.01 similarity-rate, 1 min necessity level, feedback learning-rate of 1.0.

モデル 7

- $r^2 = 0.78$
- $RMSSD = 1.57$

人間よりもかなりよいパフォーマンス

5.7. Summary of experiments

- IBLT の仮定するメカニズムを段階的に付加
- 人間のデータとのフィッティングも段階的に向上
- 最も人間とフィットしたモデルはモデル 6-Low
- IBLT の理論を完璧に実装したモデルはモデル 7
- モデル 7 は人間のパフォーマンスをうまく捉えることが出来なかった

Table 3
Summary of experimental conditions, and results from r^2 and RMSSD compared to human data

	Recognition	Judgment	Choice	Feedback	r^2	RMSSD	r^2 model comparison
Model 1	Random	Time heuristic, no similarity	No stopping rule	No	.02	1.64	0.04
Model 2	Sorted	Time heuristic, no similarity	No stopping rule	No	-.12	2.02	0.15
Model 3	Random/sorted	Time heuristic, no similarity	No stopping rule	No	.46	1.73	0.48
Model 4	Random/sorted	Blending, similarity-rate 1.0	No stopping rule	No	.64	2.32	0.63
Model 5	Random/sorted	Blending, similarity-rate 0.01	No stopping rule	No	.81	1.71	0.59
Model 6-High	Random/sorted	Blending, similarity-rate 0.01	Necessity level = 480 min	No	.86	2.62	0.71
Model 6-Medium	Random/sorted	Blending, similarity-rate 0.01	Necessity level = 60 min	No	.85	1.43	0.81
Model 6-Low	Random/sorted	Blending, similarity-rate 0.01	Necessity level = 1 min	No	.90	1.26	0.75
Model 7	Random/sorted	Blending, similarity-rate 0.01	Necessity level = 1 min	Yes	.78	1.57	1.00

6. Process analysis

これまでの分析はパフォーマンスのみの分析

パフォーマンスのみではモデルが人間の学習をどの程度シミュレートしたのかわからない

人間とモデルはどの程度、蓄積されたインスタンスから影響を受けたのか

6.1. Average fit to decision rules

タイムヒューリスティクスをつかった程度(fit) の比較(人間 vs. モデル 6-Low)

$$\text{Fit} = 1 - \frac{\text{actual decision} - \text{worst decision}}{\text{best decision} - \text{worst decision}}$$

- actual : 意思決定者(モデル/人間)が実際に行った決定(開いたポンプ)の値(恐らく水量とデッドラインまでの残り時間)
- worst : タイムヒューリスティクスにより予測されるもっとも悪い決定(開いたポンプ)の値
- best : タイムヒューリスティクスにより予測されるもっとも悪い決定(開いたポンプ)の値

Fit の値が 0 に近い = タイムヒューリスティクスにより予測される悪い決定を意思決定者は行った

Fit の値が 1 に近い = タイムヒューリスティクスにより予測される良い決定を意思決定者は行った .

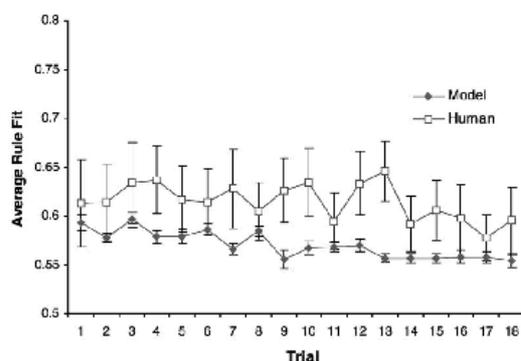


Fig. 16. Average rule fit, comparison between human data and Model 6-Low.

$r^2 = 0.08$

RMSSD = 1.47

相関はないけど同程度のフィットであっ

た

モデルははじめはフィットが高くて後は

低くなった

モデルがタイムヒューリスティクスを使った割合

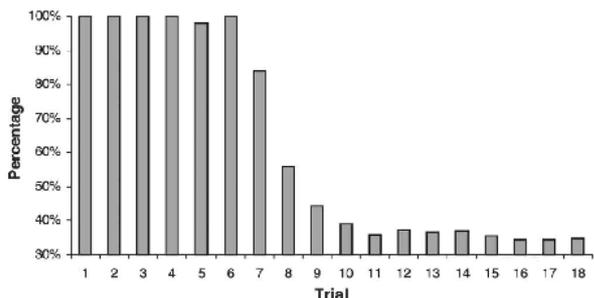


Fig. 17. Percentage of judgments using the time heuristic by trial from Model 6-Low.

はじめの試行はほとんどタイムヒューリスティクス
 後の試行ではタイムヒューリスティクスを使わない(インスタンスによる決定)
 人間もおなじような学習をしたのでは？

6.2. Instance similarity

直面する状況がどの程度過去のインスタンスと類似したか(インスタンスとの類似に基づく意思決定をどの程度の割合で行ったか)

$$\text{Similarity} = 1 - [\alpha(\text{decision}_i - \text{decision}_{(i-1)}) + (1 - \alpha)(\text{accumulated similarity}_{(i-1)})]$$

decision_i : 時系列 i において決定したポンプの値(水量とデッドラインまでの残り時間)

decision_{i-1} : 時系列 i の直前で決定したポンプの値

: 直前の決定との類似度にどれだけの重みを与えるか

similarity が 1 に近いほどインスタンスに依存した意思決定を行ったことに

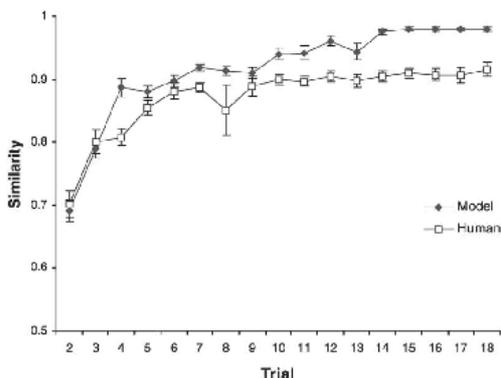


Fig. 18. Average similarity with alpha = 0.99.

が 0.99(直前の事例との類似度)のとき
 r²=0.91
 RMSSD=5.13
 相関は出ているが、RMSSD は高い
 人間よりもモデルのほうが直前の事例に影響されている

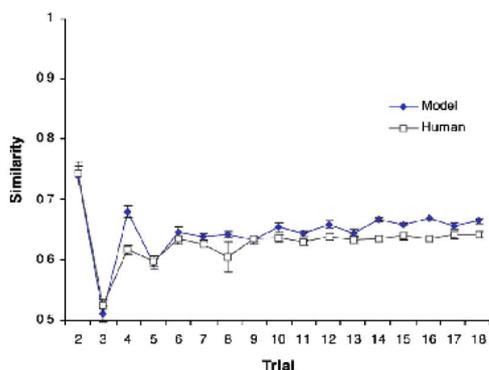


Fig. 19. Average similarity with alpha = 0.5.

が 0.5(過去に蓄積された事例との類似度を考慮した)のとき
 r²=0.84
 RMSSD=3.76
 相関は低くなったが、RMSSD は低い
 人間とモデルは同様のパターンに
 人間は直前の事例だけでなく、蓄積された事例との類似度に基づいて意思決定を行っている

6.3. Exploring individual data

これまでの分析は被験者 17 人の平均とモデル 17 回の実行の平均とのフィッティング

個々人のデータとモデル個々の実行とのフィッティングを行うことで学習の個人差についての知見が

見えてくるかも

この分析の詳細については <http://cogsci.psy.utexas.edu/supplements/>

7. Discussion of results

実験結果の意味

CogIBLT(SDU の蓄積, 再認に基づく検索, ストラテジーの利用, フィードバックによる SDU の更新)は IBLT 理論の WPP 課題に特化したインプリメンテーション

しかし, IBLT 理論は他の DDM 課題にも適用できると信じる

実験 1 の結果から

評価する選択肢の選択方法の変更はエキスパート/ノービスの違いと類似

- エキスパートは課題の重要な特徴(時間)に応じて選択
- ノービスはランダムに選択

はじめと最後で選択肢の選択方法を変更したモデル 3 は人間のデータに近い

モデル 3 は人為的にパラメータを変更したが, 人間は徐々にランダムから時間に応じた選択に移行するはず 徐々に移行するメカニズムの構築が今後の課題

実験 2 の結果から

インスタンスが蓄積されることによってインスタンスベースの意思決定が増加

提案されたインスタンスベースの utility の見積り方法(Blending)の妥当性が示唆

今後の課題はタイムヒューリスティクス以外のヒューリスティクスと比較

実験 3 の結果から

Necessity が 1 分のときに人間と同様のパフォーマンス

人間も割りと直前まで意思決定の結果を実行することを待つのかもしれない?

人間は Necessity をトライアルの経過と共に変更することも考えられる

はじめは時間に余裕をもって意思決定をし, 慣れてくると早い意思決定

より多くの実証研究が必要

実験 4 の結果から

フィードバックを付加したモデルは人間とのフィッティングが悪くなった

人間は環境からのフィードバックを上手に取りこむことができないのでは?

8. Concluding remarks

DDM における学習メカニズム IBLT を提案し, CogIBLT を構築した. CogIBLT は心理実験のデータと詳細にフィッティングがされた

今後, IBLT が DDM 研究を活性化することを期待したい

また, CogIBLT が認知モデルや AI の研究者を刺激して, 今後, ダイナミックな人間の認知をシミュレートするモデルが作られることを期待する