

**WTF? Detecting Students who are Conducting Inquiry
Without Thinking Fastidiously**

Michael Wixon, Ryan S. J. d. Baker, Janice Gobert,
Jaclyn Ocumpaugh, & Matthew Bachmann.

User Modeling, Adaptation, and Personalization 2012 286-296

1. Introduction

- 伝統的な生徒のモデル化は対象とされた課題の実行やただ単に知らないスキルに関連した正しくないパフォーマンスになりがちである。
- 期待されたパターンから逸れる生徒の振る舞いを考えている
 - Game the System
 - ✧ システムの特性や規則性に対してシステムチックな利点を取り教育的課題を成功させようとするものである (Baker, R. S., Corbett, A. T., Koedinger, K. R., Wagner, A. Z., 2004)
 - 学習を全く異なるタスクに変化してしまう (Magnussen, R., Misfeldt, M, 2004)
 - 多くの生徒は学習システムの目標から離れた行動に従事。
 - ✧ 例えばオンライン学習システムでは生徒は関係ない振る舞いに時間を費やす
- 本論文の著者らは学習には関係ないのに学習中に起きてしまう全く学習に関連していないようなオフタスクな振る舞いを“WTF(Without Thinking Fastidiously) = 「考えるのほんと無理」”として言及する。
- WTF は学習においてネガティブなインパクト。
 - ログファイルから直接的にこれを明確化
 - この振る舞いに適合した検出器の必須性
- この論文では WTF 検出器をマイクロワールド上のデータで開発を行う。

2. Data Set

- データについて
 - 144 人の 8 回生 (12-14 歳)
 - 理科の授業で Science ASSISTment’s Phase Change Microworld を利用 (図 1)
 - ✧ 生徒は固体、液体、気体を変化に関する調査を実施するためにシミュレーションを行う。
 - ✧ 現象からの仮説はシミュレーション内で実験を走らせることによってそれらの仮説をテストすることができる。
 - ✧ マイクロワールド上の実験結果を解析し、それらの仮説 (主張) を正当化したり、発見を交換する。
 - ✧ 分析中のいかなる時点でも実験や仮説フェーズを戻すことが可能。
- 本研究では、仮説と実験フェーズのアクションに焦点を当てログを取った。
- 仮説生成、実験の設計と実施、仮説と実験と他のアクティビティ間の切り替えを 144 人の生徒の総 841 のログデータとしてとった。
 - ログにはアクションタイム、シミュレーション関数値、タイムスタンプが含まれている。

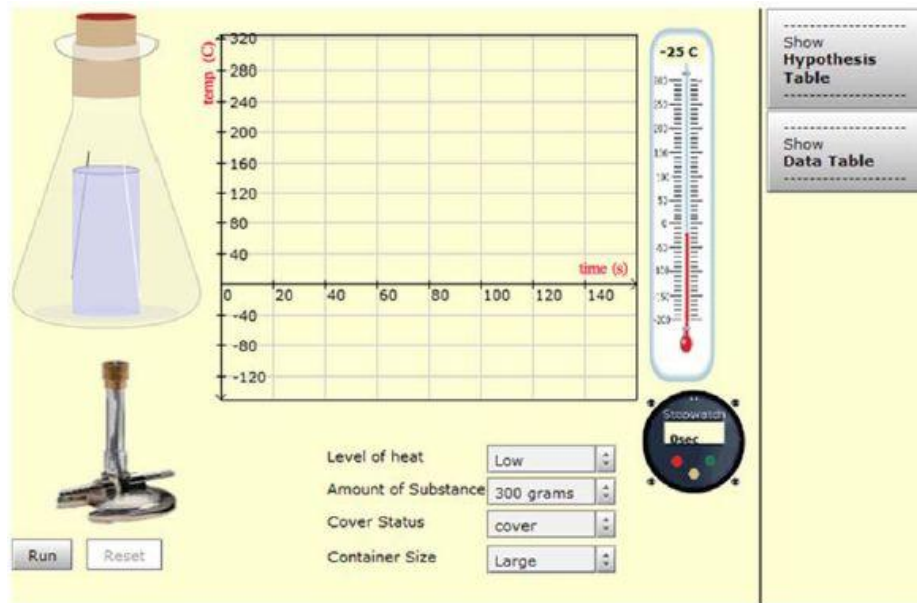


Fig. 1. A screen shot of the Phase Change microworld

3. WTF Detector

3.1 Text Replay を用いた WTF の分類ラベルの獲得

- WTF の検出器を開発する第一歩として、Text Replays を用いて、ラベル分類を開発する。
- Text Replays を用いる利点
 - 人間はかなりの量のログファイルを示される。
 - 学習から関係ない振る舞いは比較的容易に人間の手によって認識することが出来る。
 - 図 2 はかなり多くの回数同じ実験、変数設定を行き来したり、繰り返しあちこちしたりしているデータの例である。
- 生徒のデータは生徒の振る舞いのシーケンスである” clips” で分割。
 - 生徒のデータ収集の振る舞い（マイクロワールド上の実験）のシーケンスで分割。
 - “clip” は生徒がデータ収集フェーズに突入してからフェーズ終了まで
 - 仮説形成、データ収集、データ解釈、仮説の正当化、発見を交換、
- 生徒はデータ解釈の後データ収集に戻ることが出来るため Clip は以下で始まる。
 - 生徒が仮説を形成した後
 - データ収集することの決定
 - データを解釈し、さらにデータを収集することに試みた後
- 2 人の人間のコーダーが WTF のコーディングを練習
 - 練習に用いたデータは検出器の開発には用いない
 - 初めに、Clips のコーディングについてお互いに議論し規範化する
 - 200 の clip をそれぞれ独立してコーディングする。
 - 2 人のコーダーは Choen の Kappa 値 0.66 を出した（お互いのコーディングがどれくらい一致してるかみたいなもの？）
- 2 番目の著者が 501 の clip をコーディングする（これは WTF 検出器開発に用いる）
 - ✧ 全体の 3% の clip が WTF としてラベリングされた。
 - ✧ 上記の WTF は全体の 15% の生徒（複数回 WTF を行う生徒は存在しない）から引き出されている

```

Coding File  Replay File  Help
Text Replay  Coding  Coding Scheme
time since start:0 Hypothesis iv: 18752
time since start:65 Hypothesis iv: current="container.containerSize" previous="unknown" 18756
time since start:69 Hypothesis iv: direction: current="increases" previous="unknown" 18757
time since start:79 Hypothesis dv: current="time the water took to boil" previous="unknown" 18758
time since start:89 Hypothesis dv: direction: current="doesn't change" previous="unknown" 18759
time since start:103 Adding hypothesis: iv="container.containerSize"; iv direction="increases"; dv="time the water took to boil"; iv direction="doesn't change" 18760
time since start:120 Experiment 18761
time since start:129 Microworld variable: Container Size' current="Medium" previous="Large" 18762
time since start:133 Microworld variable: Container Size' current="Large" previous="Medium" 18763
time since start:141 Microworld variable: Container Size' current="Medium" previous="Large" 18764
time since start:149 Microworld run: Amount of Substance=300 grams; Container Size='Medium'; Cover Status='cover'; Level of heat='Low'; 18765
time since start:164 Microworld complete
time since start:171 Microworld reset
time since start:176 Microworld run: Amount of Substance=300 grams; Container Size='Medium'; Cover Status='cover'; Level of heat='Low'; 18772
time since start:191 Microworld complete
time since start:199 Microworld reset
time since start:203 Microworld variable: Container Size' current="Large" previous="Medium" 18779
time since start:204 Microworld run: Amount of Substance=300 grams; Container Size='Large'; Cover Status='cover'; Level of heat='Low'; 18780
time since start:219 Microworld complete
time since start:224 Microworld reset
time since start:242 Microworld variable: Container Size' current="Medium" previous="Large" 18786
time since start:244 Microworld run: Amount of Substance=300 grams; Container Size='Medium'; Cover Status='cover'; Level of heat='Low'; 18787
time since start:259 Microworld complete
time since start:268 Microworld reset
time since start:271 Microworld variable: Container Size' current="Large" previous="Medium" 18793
time since start:275 Microworld run: Amount of Substance=300 grams; Container Size='Large'; Cover Status='cover'; Level of heat='Low'; 18794
time since start:291 Microworld complete
time since start:295 Go to analyze 18797
1.No_Trial 2.WTF 3.Other Done
    
```

3.2 Data Features

- コーダーによってラベルされた振る舞いの clips に一致したデータの特徴を抽出。
- 初めは 77 のイニシャルな特徴が前もって開発されたコードを用いて抽出される。
 - しかし、特徴の多くは WTF に関連して出現しない上に、過剰な適合をしてしまう可能性[16]
 - 24 の特徴までに削減した。
 - ◇ 大きく分けて 6 群のデータ特徴抽出(最終ページに特徴掲載)

3.3 Detector Development

- WTF を 11 の分類アルゴリズムを用いて検出器を適合
 - ベイジアン、J48 決定木など
 - 一番良いパフォーマンスを出したのは PART アルゴリズム
 - PART は C4.5 の決定木のルールを生成する
 - アルゴリズムの実装は WEKA (データマイニングソフトウェア) 用いた。
- PART アルゴリズム
 - 決定木と最適パスを導くルールを繰り返し生成する
 - 生徒モデリングではあまり用いられていないが、成功的な先行研究がある
- モデルの評価か交差検定 (cross-validation) で行う
 - 6 つのグループに生徒を分ける
 - 第 6 のグループでテストする前に 5 つのグループのデータから開発する
- 検出器は A' , Kappa, Precision, Recall で評価
- A' での評価
 - WTF が含まれていない clip と WTF が含まれている clip を検出器が区別する可能性
 - $A'=0.5$ のモデルはチャンスレベル、 $A'=1$ は完全なパフォーマンス。
 - Z スコアを用いて A' と生徒ごとの標準誤差をモデルごとに算出
 - ◇ しかし著者らのラベリング手続きにおける生徒ごとに獲得した小さなサンプル数では実行不可能

◇ A' の統計的なテストは本論文では述べない。

- Chen's Kappa での評価
 - WTF が含まれる clip を検出器が有意に認識しているかどうかを評価。
 - ◇ Kappa=0 はチャンスレベルで実行している事を示す。
 - ◇ Kappa=1 は確実に実行している事を示す。
- Precision と Recall (適合率と再現率)
 - 誤検出や検出漏れを避けるのがどれくらい良いかを示している。

4. Results

- WTF 検出器は 6-fold の交差検定で行った PART アルゴリズム (一番パフォーマンスが良かった)
 - 表 1 に結果を示す
 - A' =0.979 という高い結果⇒Clip に WTF が含まれているかどうかを 97.9% 区別できる。
- Kappa=0.4。チャンスよりも 40% 優れていることを示す。
 - このレベルの Kappa 値は先行研究での検出器に匹敵している(Sao Pedro, M., Baker, R., Gobert, J., Montalvo, O., & Nakama,),(Mitchell, T.M.1997)
- Precision=38.9% (正と予測したデータのうち実際に正だった割合)
Recall=46.7% (実際に正であるもののうち生徒予測されたものの割合)
 - 介入を選択するとき、フェールソフトな介入や確実性を考慮したりすることが重要である。
 - が、検出器が WTF の検出に反抗しているか賛同しているかの強いバイアスを持っていることの証拠は存在しない。

Table 1. WTF Detector Confusion Matrix

	Clips Coded as WTF by Humans	Clips Coded as NOT WTF by Humans
Detector Predicted WTF	7	11 (false positives)
Detector Predicted NOT WTF	8 (false negatives)	475

- モデルを実行すると、最初のルールから最後のルールまでを順番に実行していく。
 - WTF を分類するために 8 つの特徴を採用し、6 つのルールを用いた。
 - 以下のうち 4 つは WTF でないことの振る舞いの特徴を識別し、2 のみが WTF の振る舞いを識別している。
- 1) 独立変数の変化の総数が 7 以下、かつ実験の実行が 3 以下なら WTF ではない。
 - 2) 中途半端な実行及び、その時間の浪費の最大が 10 秒以下、かつ独立変数の変化の総数が 11 以下かつ、停止時間の平均が 6 秒以内なら WTF ではない。
 - 3) 独立変数の変化の総数が 1 回以上、かつアクション間の 441 秒以下、かつ停止またはリセット以外の実行の総数が 4 回以下なら WTF ではない。
 - 4) 独立変数の変化が 1 2 回以下なら WTF である。
 - 5) それぞれの実験トライアルの実行の前、しかしその前のアクションの実行の後の最大時間浪費が 1.8 秒以上なら WTF ではない。
 - 6) 残りの全てのインスタンスは WTF と分類。

5 Discussion and Conclusions

- 学習環境において活発的でなく生徒がラーニングタスクから完全にディスコネクト(Without Thinking Fastidiously=WTF)したときを識別できる検出器を紹介。
 - この研究は以下の可能性がある。
 - ◇ どれほど WTF が学習に関連しているか
 - ◇ 生徒が WTF にいく要因は何か
 - ◇ 学習へと再び軌道にのせるための介入を自動化して開発することである。

- WTF 検出器が生成したルールは何が WTF であるかよりも、何が WTF でないかを認識する事を目的としていた。
 - このモデルが採用したルールのデータ特徴も議論の価値がある
 - 特徴 21 である 6 つのうち 4 つのルールにこの特徴が用いられた、
 - 実験を多数実行し、シミュレーションを長時間停止した場合、変数の変化が少ない生徒は WTF として分類
 - 特にシミュレーションを実行する前に考える事をやめず何度も変数を変化させるのは WTF である。
 - 沢山の実験トライアルや長い時間のシミュレーションの停止をしているなら WTF である。
 - ◇ 停止のパターンはシミュレーションについて勉強するための時間を利用しているという事を示していない。だって多くの変数の変化を正当化できてない

- WTF にはさまざまな理由がある
 - 態度的、課題に価値、やることを避けるという目的、混乱やいらつき、退屈の即時的な影響
 - ◇ 学習の状況や特殊な特徴を持った生徒に対して最も適切で効果的な介入を決定することが必要
 - ◇ 例えば、学習環境からどれくらい学んでいるかを理解する追加的なサポートなど。
 - ◇ 退屈で学習課題に価値を見いだせなく WTF をしている生徒には生徒のゴールの内外目で見えた価値をデモンストレーションする介入が必要である「28」

- WTF の原因を理解するためには、どれくらい WTF に反応するか、違う手法をとりながら生徒との全面的なインタラクションを行う効果的な学習ソフトウェアを開発する。

Clipに含まれているデータの特徴

- 以下4つはclips全体の統計に含まれている

- (1) the total number of actions
- (2) the average time between actions,
- (3) the maximum time between actions,
- (4) the total number of experimental trials run by the student

- 停止を基にした特徴とした以下の3つ

- (5) the number of times a student paused the simulation during runs
- (6) the average duration of student-initiated pauses of the simulation (i.e., total time spent paused, divided by number of pauses),
- (7) the duration of the longest single instance when the student paused the system.

- 実験中に時間経過に関連している10個以上の特徴を用いる

- (8) the total amount of time spent before running each experimental trial but after performing the previous action
- (9) the average time spent by the student before running each experimental trial but after performing the previous action
- (10) the standard deviation of the time spent by the student before running each experimental trial but after performing the previous action
- (11) the maximum time spent before running each experimental trial but after performing the previous action.

- 特徴の数個は実験装置のリセットor停止に関連した物が含まれる。

- 多くの状況においてシミュレーションを停止することは適切であるが、何度も何度も停止することはWTFである。

- (12) the number of experimental trials run without either pauses or resets,
- (13) the average time spent by the student before running each experimental trial which was completed without being reset but after performing the previous action,
- (14) the number of trials where the system was reset,
- (15) the average time spent before running each experimental trial that were reset but after performing the previous action
- (16) the maximum time spent before running an experimental trial that was reset before completion but after performing the previous action

- 以下に示す特徴は仮説を形成する間、どれくらい変数を変化させたかどうかが含まれている。

- (17) the number of times a variable was changed, and three measures of the period of time that elapsed before the student changed a variable (measured from the previous action, whatever it was)
- (18) the sum total of time elapsed in all these periods
- (19) the mean time elapsed across these periods
- (20) the standard deviation of time elapsed across these periods.

- 以下は実験の実施との間で独立して変数を変化することが含まれている特徴

- (21) the number of times an independent variable was changed during the experiment phase, and three measures of the period of time that elapsed before the student changed a variable (measured from the previous action, whatever it was), namely
- (22) the sum total of time elapsed in all these periods
- (23) the mean time elapsed across these periods
- (24) the standard deviation of time elapsed across these periods.