



Using Psycho-Physiological Measures to Assess Task Difficulty in Software Development

Thomas Fritz

Andrew Begel

Sebastian C. Muller

Serap Yigit-Elliott

Manuela Zuger

目的とアプローチ

- 先行研究はバグを埋め込んだ後に分析する
 - この研究ではバグを埋め込む前に「やめとけよ」といえるようにする
- タスクのむずかしさを分類
 - 目の動き, 肌の温度, 脳波の記録を計測

リサーチクエスト

- RQ1
 - 目線の動き, 皮膚の動き, 脳波からタスクのむずかしさがわかるのか
- RQ2
 - 3センサのどの組合せが予測に向いているのか
- RQ3
 - 開発者の作業中にタスクのむずかしさを予測することができるのか

実験：被験者

- 被験者
 - 15名の2年以上開発経験のある開発者
 - 27歳から60歳（男14：女1）
 - シアトルに住んでいる人でMSの研究に興味がある人
 - もちろん報酬あり

実験：タスク

- コード理解をする
 - 1つのコードを理解するのは数分で終わる
 - 練習で2つ解いて、実験で8つ解く
 - C#でコメントなし、実行もなし
 - コードを理解したかを確認する質問がある
- 2種類のプログラム
 - 2つの物体(長方形など)を生成するプログラム
 - 4つの物体(円など)を生成するプログラム
 - 複数ネストがあるなど比較的複雑

実験手順

1. 被験者に実験装置をつけ、センサのチェックをする
2. 被験者を落ち着かせるためにビデオをみる
3. 被験者の効き目を明らかにする
4. こまのない椅子に座らせ、またビデオをみる
5. 一つ目のタスクをやる
6. 作業負荷を測るテスト(NASA TLX)を行う
7. そしてまたビデオをみる
8. 二つ目のタスクをやる
9. 作業負荷を測るテストを行う
10. 実験後のアンケートに答える

機械学習

- 3タイプの予測
 - 人ごとの分類
 - 新しい人が入ってきても予測できる
 - タスクごとの分類
 - 新しいタスクが追加されても予測できる
 - 人とタスクのペアごとの分類
 - 誰かが一度やったタスクを他人がやるときに予測可能
- ナイーブベイズ, J48, SVM

RQ1:3つのセンサーからタスクの むずかしさを予測できるのか

Prediction	Sensors	Precision	Recall	F-Measure
By Participant	Eye	69.16%	65.83%	65.10%
	EDA	55.18%	55.77%	51.99%
	EEG	53.05%	56.73%	50.82%
	Eye+EDA	68.37%	64.42%	61.92%
	Eye+EEG	68.58%	63.46%	60.89%
	EDA+EEG	68.02%	64.58%	62.01%
	Eye+EDA+EEG	64.99%	64.58%	62.21%
By Task	Eye	79.17%	66.67%	69.65%
	EDA	75.12%	58.65%	63.80%
	EEG	81.97%	59.62%	63.40%
	Eye+EDA	78.59%	66.35%	70.37%
	Eye+EEG	82.42%	66.35%	69.89%
	EDA+EEG	82.79%	65.63%	69.76%
	Eye+EDA+EEG	84.38%	69.79%	73.33%
By Participant-Task	Eye	66.67%	66.67%	66.67%
	EDA	59.62%	59.62%	59.62%
	EEG	56.73%	56.73%	56.73%
	Eye+EDA	68.27%	68.27%	68.27%
	Eye+EEG	62.50%	62.50%	62.50%
	EDA+EEG	62.50%	62.50%	62.50%
	Eye+EDA+EEG	67.71%	67.71%	67.71%

RQ2: 3センサのどの組合せが 予測に向いているのか

Prediction	Sensors	Precision	Recall	F-Measure
By Participant	Eye	69.16%	65.83%	65.10%
	EDA	55.18%	55.77%	51.99%
	EEG	53.05%	56.73%	50.82%
	Eye+EDA	68.37%	64.42%	61.92%
	Eye+EEG	68.58%	63.46%	60.89%
	EDA+EEG	68.02%	64.58%	62.01%
	Eye+EDA+EEG	64.99%	64.58%	62.21%
By Task	Eye	79.17%	66.67%	69.65%
	EDA	75.12%	58.65%	63.80%
	EEG	81.97%	59.62%	63.40%
	Eye+EDA	78.59%	66.35%	70.37%
	Eye+EEG	82.42%	66.35%	69.89%
	EDA+EEG	82.79%	65.63%	69.76%
	Eye+EDA+EEG	84.38%	69.79%	73.33%
By Participant-Task	Eye	66.67%	66.67%	66.67%
	EDA	59.62%	59.62%	59.62%
	EEG	56.73%	56.73%	56.73%
	Eye+EDA	68.27%	68.27%	68.27%
	Eye+EEG	62.50%	62.50%	62.50%
	EDA+EEG	62.50%	62.50%	62.50%
	Eye+EDA+EEG	67.71%	67.71%	67.71%

RQ3:開発者の作業中にタスクのむずかしさを予測することができるのか

- 開発者が難しいと感じている部分を色づけることができるツールを開発したらできそうだ
- 3つのセンサーすべてを使うことで予測精度が向上するとは限らないことがわかった
 - ツールを作る際のコストを削減できる

まとめ

- 精神心理学的な尺度でタスクの難しさを予測することができた
- どのセンサーの組合せを用いると予測精度が向上するのかを明らかにした

所感

- 実験の手順をものすごく丁寧に記述
- 圧倒的サーベイ(参考文献74)
- 松本君のシステムの評価するのに, NASA TLX surveyを使ってみてはどうか?

Sliding time-window

- 入力: センサー, 出力: タスクの難しさ で機械学習するとセンサーの値の変化に対応できない

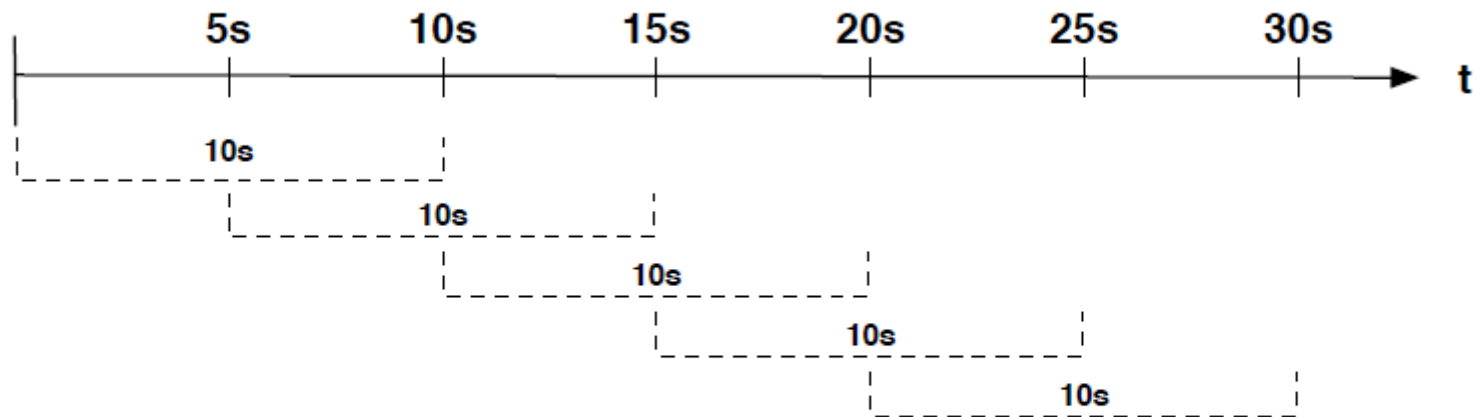


Figure 1: Sliding windows of size 10 seconds with 5 second offsets.