

科学研究費助成事業 研究成果報告書

平成 30 年 6 月 20 日現在

機関番号：32714

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K00318

研究課題名(和文) パーソナルなセンシングデータも活用するソフトウェアリポジトリマイニング

研究課題名(英文) Datamining of Software Repository using Personal Sensing Data

研究代表者

松本 一教 (Matsumoto, Kazunori)

神奈川工科大学・情報学部・教授

研究者番号：40350673

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：ソフトウェア開発技術者の身体に装着したセンサーからデータを収集し、時系列データのデータマイニングの応用により、技術者の開発時の集中度を推定する技術を開発した。さらに、技術者のソフトウェア文書やプログラム読解時の視線データを収集し、視線移動パターンのデータマイニングにより読解時の疲労度や集中度を推定する技術を開発した。また、マウスやキーボード操作の状況から集中度を推定する技術も開発した。これらの技術により、開発技術者の開発時における個人的な状況を推定できるようになる。また本研究では、推定した技術者の集中度や理解度を用いるソフトウェア品質管理手法も開発した。

研究成果の概要(英文)：We collect sensing data with sensors attached to the bodies of software engineers. By using this data, we develop a method, based on datamining techniques of time-series data, to estimate the degree of concentration during their work. We also develop a datamining method to estimate degree of comprehension and concentration during work from gaze movement data patterns. Further we develop a method to estimate concentration degree from the situation of mouse and keyboard operations. With these developed methods, it becomes possible to estimate engineers' personal situation at the time of development. We finally develop a software quality management method that uses the personal situation.

研究分野：情報学

キーワード：データマイニング 集中度推定 視線データ 個人識別

1. 研究開始当初の背景

ソフトウェアの総合的な品質管理や生産性向上のために、開発プロセスで発生する多種多様なデータを格納したソフトウェアリポジトリを構築することの有効性は、本研究の開始時点で既に知られていた。さらに、そのリポジトリに対するデータマイニング技術の開発も行われていた。しかし、その時点でのリポジトリは、ドキュメント、コード、ログ等、テキストや図式を蓄積するものであった。開発に関わる人間（技術者）についてのセンシングデータも格納対象としたリポジトリは当時実現されていなかったし、そのようなデータを格納するという研究もほとんどなされていなかった。

本研究では、ソフトウェア開発が人間（技術者）の知的活動の成果であり、技術者の個人的な状況に左右される場面も多いことに着目した。例えば技術者の理解度が低い仕様書からは、品質の低い成果物が生産される。同様に技術者が開発に集中できていない場合にも品質低下が起きる。このような技術者に関する状況をソフトウェアの品質管理に役立てていくというシステムを構想した。そして、技術者の身体に測着したセンサーからのデータもリポジトリに格納することを着想した。センサーデータに対してデータマイニングを行うことで、作業時の技術者の状況を推定し、さらに技術者の状況とソフトウェア成果物との関係性を知識抽出できる新たなデータマイニング実現を目指すことにした。

2. 研究の目的

本研究の目的は図1に示すように大きく2つに分けられる（図中の①と②）。以下にその内容を説明する。

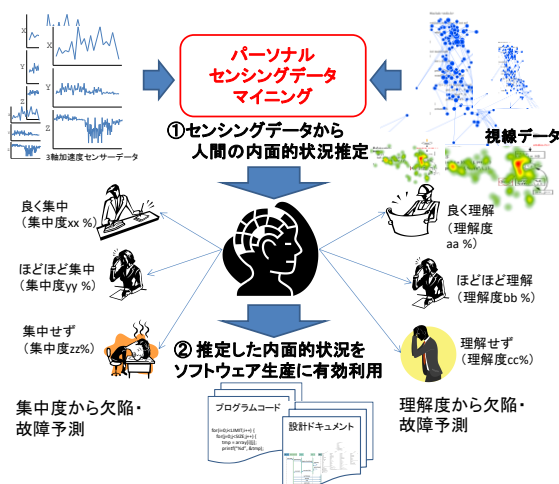


図1 研究の目的

(1) センサーデータからの状況推定

身体に装着するセンサーデータのデータマイニングにより、作業中の人間がどの程度の集中度で作業に取り組んでいるか推定す

る手法を開発する。また視線データ中の視線移動パターンから、読解作業中の作業者の理解度や集中度を推定する手法を開発する。両手法の組み合わせにより、推定精度の向上を実現する手法も開発する。集中度や理解度は人間の陽には見えない内面的な状況であるが、状況を限定すれば、高価で特殊な脳観測装置を用いることなく、入手容易なセンサーのデータだけから推定可能であることを実証して明らかにする。

(2) 推定した状況のソフト生産への活用

ソフトウェアの品質や生産性に、作業時の人間の内的な状況が関係していることを実証する。従来は内定状況を外部から推定する手段がないために、ソフトウェア工学ではほとんど検討されてこなかった問題である。そして本研究では、人間の内的な状況推定結果を積極的に利用し、生産性や品質向上につながる手法を見出す。なお、内面的情報だけを唯一利用するわけではなく、従来ソフトウェア工学手法と相補的に用いるため、ある程度の精度での推定で十分実用に供することができることも示す。

上記の(1)と(2)の技術を組み込んだ試作システムを開発し、本研究で開発手法の有効性を評価し、実用化が可能であることを実証することが本研究の目的である。

3. 研究の方法

身体に装着するセンサーとしては、スマートフォンに内蔵の3軸加速度センサーを用いて胸ポケットに装着することにした。これにより技術者にほとんど負担を強いることなくデータを収集できる。加速度センサーとしての精度や信頼性は本研究で用いる範囲では十分耐えうることも予備実験で判明した。また装着位置により収集できるデータが著しく異なってくるが、本研究では測定に限界があることを前提として胸ポケットにのみ装着する方式とした(図2)。

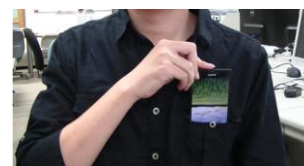


図2 胸ポケットへの装着

視線データの収集にはディスプレイの下部に装着するアイトラッカー装置を用いる。身体には全く機器を装着しないので、計測時の負荷はほとんど無いに等しい。この装置では装着したディスプレイ内での視線移動のみを測定することができる。実際の開発現場では、ディスプレイと他の資料等を交互に見て作業を進めるが、そのような視線移動は本研究では収集することができずに対象外として扱った。

4. 研究成果

(1) 加速度データからの集中度推定

加速度データからの集中度推定にはデータマイニングによる行動推定技術を応用している。本研究で開発した手法の概要を図3に示す。

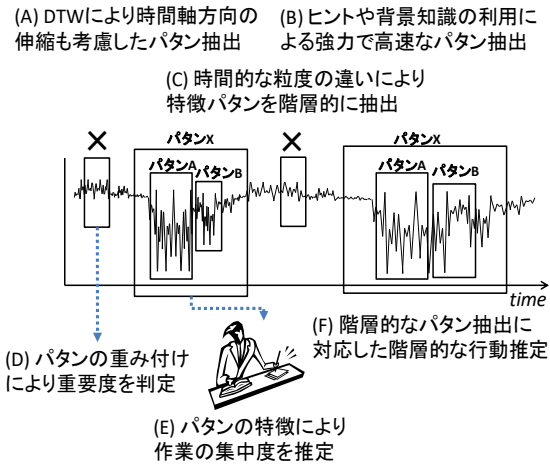


図3 集中度推定に関する技術

開発した最終的な手法は図中に示す(A)~(F)の要素技術を組み合わせたものである。(A)の技術は、時間軸方向のデータの伸縮を行いながらパターン抽出を可能にするものであり、個人毎のデータの違いを吸収する役割を果たす。この技術は時系列データマイニングの技術に基づいており、DTW(Dynamic Time Warping)を用いている。(B)の技術はパターン発見を効率的に行うための知識ベースに関する技術である。(C)や(D)や(F)の技術は、パターンの階層的な構造に従って対応する人間の行動を推定したり、パターンの重要度を決定したりする技術である。(F)はこれらの技術によって見出されたパターンを使って、人間の集中度を推定する技術である。時間軸センサーデータのパターンにより技術者の集中度を推定できる。推定した集中度は10秒~100秒程度の時間間隔で記録されるようにしている。

	ConcentrationDegree 集中度測定結果	01/11 01:06:06
	Nomal 判定:3 集中度:50 01/11 01:06:06	
	ConcentrationDegree 集中度測定結果	01/11 01:05:55
	Nomal 判定:3 集中度:50 01/11 01:05:55	
	ConcentrationDegree 集中度測定結果	01/11 01:05:45
	Nomal 判定:3 集中度:50 01/11 01:05:45	

図4 ライフログとしての集中度管理

推定された集中度の妥当性については、個人差が大きいため、多くの技術者に共通する指標として用いることは困難であることが

判明した。そこで技術者個人毎に推定された集中度を個人ライフログとしてツイッターに投稿する方式を採用した(図4)。技術者個人の過去の履歴をリポジトリ内の成果物と対応付けて管理できるようにしている。

(2) 視線データからの集中度推定

視線データについては、予備実験の結果から、疲労度を比較的良く推定できることが判明した。疲労度については、フリッカー値による計測が良く知られている。本研究で行った実験では、タイピングのような比較的単純な作業時において、タイプミス回数(y)とフリッカー値(x)の間には、

$$y = -3.4x + 135.1$$

という線形回帰が可能であることが明らかになった($R^2=0.72$)。このようにフリッカー値は、疲労度=ミスが生じる可能性の近似として用いることができるが、その測定のためには作業を中断してフリッカー装置による測定を行う必要がある。本研究では、動画で提示される文書に対して、その内容とは無関係のノイズを挿入し、そのノイズへの反応時の視線パターンからフリッカー値、すなわち疲労度を推測することとした。その結果、フリッカー値 40Hz のとき、すなわち疲労がほとんどない状況では 53%の被験者がノイズ混入を感知できるが、軽い疲労状態であるフリッカー値 35Hz のときには 39%の被験者が感知でき、疲労状態であるフリッカー値 34Hz 以下においては 20%以下の被験者しか感知できないノイズパターンを得ることができた。このことにより、本研究で得たノイズパターンを文書中に混入させることにより、ある程度の精度で作業者の疲労度を視線データから推定できることが明らかになった。先に述べたように疲労度と作業ミスには相関関係があるので、本研究では疲労度を集中度の近似として扱っている。

(3) 視線データからの個人識別

ソフトウェア開発現場では、実際に作業を行った人間が確定できないという状況がよく起きる。作業の途中で担当者が変わるという状況もある。実際の作業者を特定できるだけでも、品質をある程度推測できることもある。

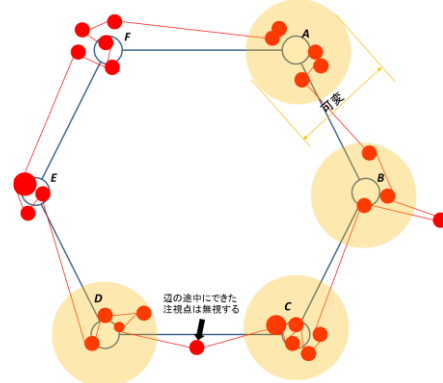


図5 個人識別の多角形パターン

そこで、視線データを用いて、実際に作業を行った個人を特定する技術を開発した。開発した手法では、文書中に単純な多角形を表示し、その全ての頂点を順番に注視するように指示する。個人によって、表示されている図形とはやや異なった形の視線パターンが得られる。正六角形の場合に、この状況を図5に示す。なお、図形の違いや大きさ、色などが最終的な識別率に影響を及ぼすことも判明した。

各頂点付近での視線のズレ（座標位置および停留時間）をパラメータとして、データマイニング（機械学習）の手法による個人識別を行う。用いる手法としては、サポートベクターマシン(svm)、畳み込みニューラルネットワーク(cnn)、多層パーセプトロン(mlp)、決定木(decision tree)、K近傍法(knn)を用いた。また、精度の比較を行うために、本人拒否率(False Rejection Rate, FRR)および他人受率(False Acceptance Rate)を用いた。その結果を表1に示す。高い精度とはいえないので、これらの学習器を組み合わせるアンサンブル学習の手法も試みたが、明確な精度向上は得られなかった。

表1 視線での個人識別精度

	a		b		c		d	
	FRR	FAR	FRR	FAR	FRR	FAR	FRR	FAR
svm	0.60	0.08	0.20	0.05	0.75	0.10	0.40	0.17
cnn	0.80	0.07	0.00	0.10	0.45	0.07	0.20	0.13
mlp	0.75	0.10	0.10	0.07	0.85	0.07	0.35	0.17
decisiontree	0.75	0.15	0.65	0.27	0.65	0.23	0.45	0.23
knn	0.75	0.22	0.05	0.58	1.00	0.00	0.80	0.18

(4) その他のデータからの集中度推定

作業中のマウスクリック数やキーストローク数も作業者の集中度を推定する指標として用いることができる。本研究では、これらのデータからの推定値も用いているようにしている。

(5) ソフトウェア開発管理への適用

ソフトウェア開発では進捗管理が重要な課題となる。本研究では、製造の現場で良く活用されているカンバン方式を応用する方式を開発した。この方式では、表示板（カンバン）に現在のタスク状況を3つの状態(TODO, Doing, Done)に分けて示す。さらに、TestやReviewといったステージの状態を付加することもできる。いずれにしても、カンバンというディスプレイ上に簡潔な状況のみを示すことがポイントである。本研究では、このカンバン上に先に述べた手法で推定した作業者の集中度や疲労度などのデータを補足的に示すことができるようにした。このようなカンバンを用いることで、進捗管理をより個人の状況を踏まえた詳細なものにできる。

(6) 研究の総括

本研究では、ソフトウェア開発時の開発者に関する個人的な状況を推測する方法を開発した。スマートフォンに内蔵される単純な加速度センサーを用いてある程度の集中度推定が可能となる方式を開発した。視線データを用いて、疲労度や集中度を推定する方式も開発した。また視線データから、個人を識別する技術も開発した。この技術を用いることで、実際に文書を読んだ開発者を特定できるようになる。これらの個人に関する推定値を組み込んだソフトウェア開発管理システムをカンバン方式に基づいて拡張することで実現した。カンバン上に開発者の個人的な状況が表示されるので、従来とは全く異なる開発管理を行うことができるようになる。

(7) 今後の課題

視線による疲労度測定では、動画を使う手法を開発した。視線データを用いる場合には、動画を用いる手法が今後期待できる。本研究で開発した様々な手法をソフトウェア開発プロセスの中に組み込み品質向上を達成できる方法論やツールの開発は今後の課題である。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計1件)

- (1) 中澤 舜, 鈴木 孝幸, 田中 哲雄, 松本一教, ウィンドウの枠線描画機能を持つクラウドサービス向けウィンドウマネージャ, 電気学会論文誌C, 査読有, 136巻, 8号, 1237 - 1245, 2016, DOI: 10.1541/ieejieiss.136.1237

[学会発表] (計17件)

- ① Yu Takahata, Tetsuo Tanaka, and Kazunori Matsumoto, Personal Authentication Method based on Eye-Tracking Data Mining, 23th International Symposium on Artificial Life and Robotics (AROB 23th), Bcon Plaza, Beppu, Japan, January 19, 2017
- ② Shun Nakazawa, Tetsuo Tanaka, and Kazunori Matsumoto, Development and Application of Monitoring Functions in a Software Development Project Carried Out by Students, 6th Asian Conference on Information Systems (ACIS2017), Phnom Penh, Cambodia, December 13, 2017
- ③ 渡辺優太, 中澤 舜, 田中哲雄, 松本一教, 学生の学習行動を可視化する穴埋め式ワークブックにおける基本機能の試作, 電気学会情報システム研究会資料, IS-17-053, 利尻町交流施設「どんと」, 利尻町, 2017年8月2日

- ④ Shun Nakazawa, Kazuki Komatsu, Tetsuo Tanaka, and Kazunori Matsumoto, Development and Evaluation of Large-Screen Digital Kanban with Smartphone Operation, 6th IIAI International Congress on Advanced Applied Informatics (IIAI AAI2017), ACT City Congress Center, Hamamatsu, Japan, July 12, 2017
- ⑤ 中澤 舜, 小松和樹, 田中哲雄, 松本一教, スマートフォン操作が可能な大画面デジタルカンバンの開発, 電気学会 情報システム研究会資料, IS-17-017, 大濱信泉記念館, 石垣市, 2017年3月24日
- ⑥ 高畑 悠, 前川凌佑, 後藤翔太, 松本一教, 類似性に基づく視線データの分析, 電気学会 情報システム研究会 IS-17-019, 大濱信泉記念館, 石垣市, 2017年3月24日
- ⑦ 渡辺優太, 中澤舜, 田中哲雄, 松本一教, 学生の学習行動を可視化する穴埋め式ワークブックの設計, 電気学会情報システム研究会資料, IS-17-034, 大濱信泉記念館, 石垣市, 2017年3月24日
- ⑧ Kazunori Matsumoto, Tetsuo Tanaka, Considerations on Web Document Designs, 5th IIAI International Congress on Advanced Applied Informatics (IIAI AAI2016), City International Center, Kumamoto, Japan, July 10, 2016
- ⑨ 杉村 博, 田中哲雄, 松本一教, 文書の著者推定方式の実装と比較, 電気学会情報システム研究会資料, IS-16-020, はこだてみらい大学, 函館市, 2016年8月2日
- ⑩ 中澤 舜, 田中哲雄, 松本一教, 実働時間計測機能とコスト見積り機能を持つデジタルカンバンの開発と適用, 電気学会 情報システム研究会資料, IS-16-022, はこだてみらい大学, 函館市, 2016年8月2日
- ⑪ Kazunori Matsumoto, Tetsuo Tanaka, Shigenori Ioroi, Possible Worlds Framework for Decision Making of Intelligent Agents, 5th Asian Conference on Information Systems (ACIS2016), Krabi, Thailand, October 28, 2016
- ⑫ Shun Nakazawa, Tetsuo Tanaka, Kazunori Matsumoto, Prototype of Functions for Actual Work Time Measurement and for Task Cost Estimation on the Digital Kanban, 5th Asian Conference on Information Systems (ACIS2016), Krabi, Thailand, October 28, 2016
- ⑬ Atsushi SHIMIZU, Keisuke TAKAHASHI, Daisuke YAGI, Tetsuo TANAKA, Kazunori MATSUMOTO, Analyzing a Relation

- between Skill Levels and Eye Tracking Patterns: Towards Effective Learning System that Suggests Weak Points, 22nd International Symposium on Artificial Life and Robotics (AROB 2016), Bcon Plaza, Beppu, Japan, January 21, 2016
- ⑭ Keisuke TAKAHASHI, Atsushi SHIMIZU, Daisuke YAGI, Tetsuo TANAKA, Kazunori MATSUMOTO, Investigating the Relation of Eye Tracking Patterns and Preferences: Towards a Methodology of Digital Documents, 22nd International Symposium on Artificial Life and Robotics (AROB 2016), Bcon Plaza, Beppu, Japan, January 21, 2016
- ⑮ Shun NAKAZAWA, Tetsuo TANAKA, Kazunori MATSUMOTO, Towards Visualization of Software Project using Digital KANBAN Log Analysis, 22nd International Symposium on Artificial Life and Robotics (AROB 2016), Bcon Plaza, Beppu, Japan, January 21, 2016
- ⑯ Takayuki SUZUKI, Tomohiro TAKAHASHI, Tetsuo TANAKA, Kazunori MATSUMOTO, Prototyping and Trial Experiment of Lecture and Learning Support System that uses Students', 5th Asian Conference on Information Systems (ACIS2015), Penang, Malaysia, October 16, 2015
- ⑰ Daisuke YAGI, Takayuki SUZUKI, Tetsuo TANAKA, Kazunori MATSUMOTO, Towards Personal Identification using Eye Tracking Patterns, 5th Asian Conference on Information Systems (ACIS2015), Penang, Malaysia, October 16, 2015

[図書] (計1件)

- (1) 松本一教, 宮原 哲浩, 永井 保夫, 市瀬 龍太郎, IT text, 人工知能 第2版, 全244ページ, オーム社, 2016

6. 研究組織

(1) 研究代者

松本 一教 (MATSUMOTO KAZUNORI)
 神奈川工科大学・情報学部・教授
 研究者番号: 40350673

(2) 研究分担者

大須賀 昭彦 (OHSUGA AKIHIKO)
 電気通信大学・大学院情報理工学研究科・教授
 研究者番号: 90393842