

読書行動の検出における有効な視線特徴量の選定

中嶋 一樹^{1,a)} 内海 ゆづ子^{1,b)} 岩村雅一^{1,c)} 黄瀬 浩一^{1,d)}

概要：読書行動は学力や知識と相関関係があると言われており、読書行動を記録することで新たな知見を得られると考えられる。しかし、現状では日常生活での読書の記録方法は自動ではない。よって、その自動記録化に向け、本稿では実環境を想定した読書行動の検出手法を提案する。これまでの読書行動に関する研究に Bulling らの研究、吉村らの研究、Kunze らの研究がある。それらで用いられている視線特徴量は読書行動の検出に有効である可能性をもつ。そこで本稿では、それらの特徴量から読書行動の検出に有効な特徴量を Backward stepwise selection を用いて選定する。今回の手法の評価のためにデータセットを作成し、それを用いて評価を行ったところ、特徴量は 30 個から 24 個に選定され、検出精度は 90.5% となった。

1. はじめに

近年、リーディングライフログに関する研究が行われている [1], [2]。リーディングライフログとは、人が日常生活中で「いつ」「何を」「どのように」読んだかなどを長期にわたりデジタルデータとして記録すること、あるいはその記録自体のことである [3]。読書の量や質などは、学力や知識と相関関係があると言われており [4]。よって、リーディングライフログは、読書の質や量と相関関係のある能力の推定なども可能になると考えられる。しかし、現状では「いつ」「何を」「どのように」読んだかを記録する方法は自動ではない。そのため、何時から何時まで、何の本を、どれくらいの理解しながら内容を読んだかなどを手動でしか記録することができず、長期的な記録は大変困難である。よって、我々はリーディングライフログを自動記録化することを最終的な目標とし、本稿では日常生活での読書行動を検出する問題を扱う。

読書行動を検出する研究に、一人称視点画像を用いたもの [5] や、加速度センサを用いたもの [6] などがある。しかし、複数種類ある行動の中から読書行動を判別するような限定的な条件での研究が多く、それではただ文字を眺めて実際には読んでいない場合も検出してしまふ。読んでいないか読んでいないかまでを判断するためには、視線情報などに注目する必要がある。

これまでに視線情報を用いた読書行動に関する研究がいくつかされている。Kunze らは、小説や漫画などを含む 5 種類の文書の中から、ユーザの読んでいる文書の種類を識別した [7]。ただし、読書行動の検出は行っていない。読書行動検出の研究では、Bulling らは、視線情報を用いて読書行動を含む行動認識をしている [10]。しかし、ユーザに負荷がかかるデバイスを使用していたので実環境には適さない。他に、ブラウザ上の文書のみを扱った研究 [8] や、横書きで書かれた文書のみを用いた吉村らの研究 [9] などがある。ただし、限定された文書を用いているので、実環境での読書行動の検出には適していない。

本稿では、視線情報を用いて、日常生活を想定した実環境での様々な文書の読書行動を検出する。視線情報を用いた読書行動の研究のうち、Bulling らの研究 [10]、吉村らの研究 [9]、Kunze らの研究 [7] は、モバイルアイトラッカなどの装着型デバイスから特徴量を抽出しており、実環境における読書行動は検出できていないが、用いている視線特徴量は読書行動の検出に有効なものが含まれている可能性がある。そこで、その 3 つの従来研究で用いられている特徴量を全て使用し、読書行動の検出を行う。ただし、全ての特徴量が読書行動の検出に必要であるとは限らないので、全ての特徴量の中から読書行動の検出に有効な特徴量を Backward stepwise selection を用いて選定する。

視線情報を用いる実環境での読書行動の検出は従来の研究では行われていなかったため、評価用のデータセットを作成した。作成したデータセットを使って特徴量の選定を行ったところ、特徴量は 30 個から 24 個に選定され、検出精度は 90.5% となった。

¹ 大阪府立大学大学院工学研究科, 〒599-8531 堺市中区学園町 1-1 Graduate School of Engineering, Osaka Prefecture University

a) nakajima@m.cs.osakafu-u.ac.jp

b) yuzuko@cs.osakafu-u.ac.jp

c) masa@cs.osakafu-u.ac.jp

d) kise@cs.osakafu-u.ac.jp

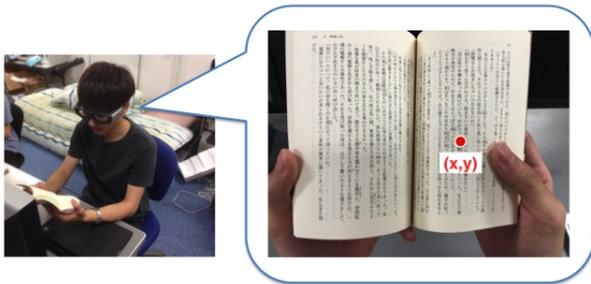


図 1 モバイルアイトラッカによる視線情報の取得

Fig. 1 A gaze point obtained from a mobile eye tracker.

2. 関連研究

これまでに読書行動を検出する研究が行われているが、行動認識の中の一つの行動として読書が扱われていることが多い [5], [6]. Jalal らは depth camera を用いて読書行動を含めた行動認識を行っている [11]. ただし、その場合ある特定の行動との識別は可能だが、実環境を想定することを考えれば、ただ文字を眺めていて実際に読んでいないような場合と読書行動を区別することは難しい。したがって、読書行動に関する研究では、正しい読書行動を検出するため、多くの研究が視線情報に注目している。

Campbell らは、据え置き型のデバイスを使用し、モニターを見るユーザの視線情報を用いて読み判定を行った [8]. しかし、固定されたブラウザ上の文書しか扱っていないので、検出できる読書行動は限定されている。また、据え置き型のデバイスから特徴量を取得しているので、特徴量は実環境での読書行動の検出には適していないと考えられる。Bulling らは、目の周りの電位を測定し、眼球運動を抽出することで、読書行動を含めた 6 クラスの行動認識を行った [10]. デバイスには、装着型のものを用いていたが、目の周りの複数箇所には電極をつけ電位を計測するものであったため、ユーザに負荷がかかった。特徴量には、眼球運動から得られる様々な特徴量のうち、特に行動認識に有効であった特徴量を用いているので、読書行動の検出に有効な特徴量が含まれている可能性がある。吉村らは、視線情報を得ることができるモバイルアイトラッカを使用して、文書の読み判定を行った [9]. 横書きの文書のみを対象としており、日常で読むあらゆる文書は対象としていない。文書の読み方や非読書行動に制約を与えていないので、実環境を想定した読書行動の検出にも有効であると考えられる。

Kunze らは、モバイルアイトラッカから得られる視線情報を用いて、ユーザが、漫画、教科書、雑誌、小説、新聞の 5 種類のうち、どの文書を読んでいるかを識別をした [7]. ただし、読書行動の検出は行っていない。様々な種類の文書を扱っているため、実環境における読書行動検出に有効な特徴量を用いている可能性があると考えられる。

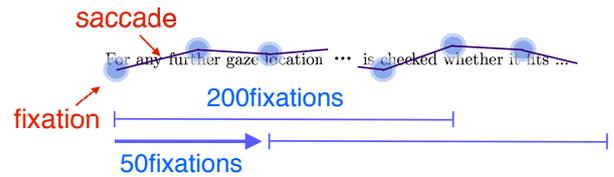


図 2 Sliding window approach による視線情報のグループ化

Fig. 2 Forming groups of eye movements by sliding window approach.

3. 読書行動検出手法

読書行動の検出には、ユーザの視線の動きから読書行動を検出する方法を提案する。ユーザの視線情報を取得し、得られた fixation をグループ化することで、そこから特徴量を抽出する。抽出した特徴量を用いて Support Vector Machine(SVM)[12] により読書行動の検出を行う。

3.1 視線情報の取得

日常生活での読書行動を検出するため、デバイスは使用場所や行動を制限せず視線情報を取得できるものが望ましい。そこで、視線情報の取得にはモバイルアイトラッカを用いる。モバイルアイトラッカはシーンカメラと両目を撮影するカメラが付いており、それらによって図 1 のように一人称視点画像とそれに対応したユーザの視点を取得することができるメガネ型のデバイスである。

アイトラッカで得られた視点座標にはノイズが含まれていることがある。そこで、アイトラッカで得られた視点座標に対して、Buscher らの手法 [13] を用いて fixation・saccade 判定を行い、得られた結果を視線情報として用いる。ここで、fixation とは視点が、ある小さい範囲の中に一定時間とどまる状態を表す。それに対し、saccade とはその fixation と fixation を線分で結んだ素早い目の動きのことである。

3.2 視線情報のグループ化

読書行動は瞬間的な行動ではなく、ある一定時間連続して続ける動作である。よって、時系列順に連続した fixation を 1 つのグループとし、それを特徴量の抽出に用いる。グループ化には Sliding window approach を使用する。図 2 に示すように、時系列順に並んだ fixation から、連続した 200 個の fixation を一つのグループとする。この 200 個をウィンドウサイズとする。次に、ウィンドウサイズはそのまま、時系列順に 50 個分 fixation をずらして、それを次のグループとする。この操作を最後まで繰り返していくことで、fixation をグループ化する。予備実験ではウィンドウサイズを 200 個、ずらし幅を 50 個にすると、最も検出精度が高かったため、今回はこれらの値を用いる。なお、

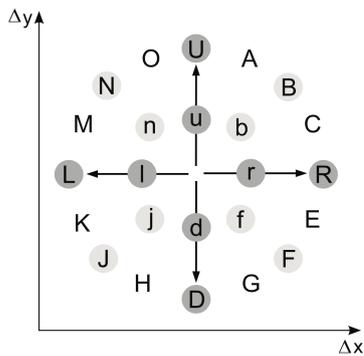


図 3 Bulling らによる saccade の分類 [10]

Fig. 3 Classification of saccade by Bulling's method [10].

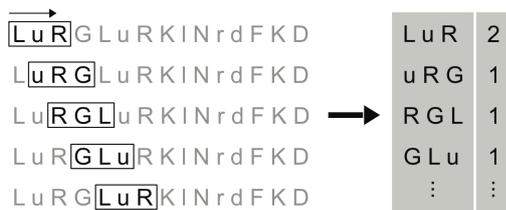


図 4 Bulling らによる n-gram 法の例 (n=3)[10]

Fig. 4 An example of n-gram by Bulling's method (n=3) [10].

連続した 200 個の fixation は時間に換算するとおよそ 1 分から 1 分半である。

3.3 特徴量の抽出

前節で説明したグループ一つ一つから特徴量を抽出する。本稿では、吉村らの手法と、Kunze らの手法と、Bulling らの手法で用いられている特徴量を用いる。表 1 に手法ごとの特徴量を示す。以降、それぞれの特徴量について詳しく説明する。

3.3.1 Bulling らの手法

Bulling らは論文の中で、眼球付近の電位を測定することで、saccade, fixation, まばたきを抽出し、その情報を解析することで特徴量を計算していた。Bulling らが用いた特徴量について、詳しく説明する。saccade の長さの平均と分散、視線運動における fixation 時間の割合、まばたき時間の割合を特徴量として用いている。また、時系列順に並んだ saccade は方向と長さを持っているので、一つずつ図 3 のように 24 種類に分類し、アルファベットの列を取得する。そのアルファベットを図 4 のように n-gram 法を用いて、現れる出現語パターンの頻度を数えていく。そして、出現語回数の最大値、そして全ての出現語の出現回数の平均を特徴量として用いる。この特徴量は、同じ視線運動が繰り返し行われる読書行動には適しているのではないかと考えられる。

抽出する特徴量について、本稿では Bulling らと同じデバイスを使用しないため、アイトラッカのソフトウェアから得られた fixation, saccade, まばたきを用いる。

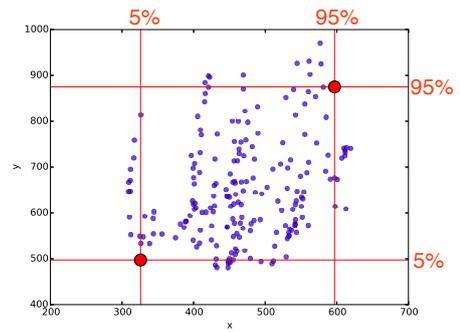


図 5 分位点間のマンハッタン距離

Fig. 5 Manhattan distance between quantile points.

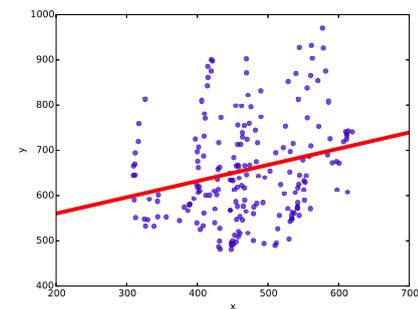


図 6 fixation から得られる近似直線

Fig. 6 Linear regression of fixations.

3.3.2 吉村らの手法

吉村らは論文の中で、モバイルアイトラッカを用いて視線情報を取得し、それを解析することで特徴量を得ていた。fixation に関するものは、fixation の合計時間、fixation の平均時間を用いている。saccade に関するものは、saccade の長さの平均、saccade の長さの最小値、saccade をベクトルとみなした際の合成ベクトルの x 成分と y 成分を用いている。また、x 軸方向の視点座標に対して一次元連続ウェーブレット変換した後の振幅の大きさの平均を特徴量としている。fixation の特徴量からは主に読書時の注視時間、saccade の特徴量からは読書時の視点の動きの大きさを抽出していると考えられる。

3.3.3 Kunze らの手法

Kunze らは論文の中で、モバイルアイトラッカを用いて視線情報を取得し、それを解析することで特徴量を得ていた。用いた特徴量は、saccade の角度の平均と分散、saccade の上下左右方向の頻度、分位点間のマンハッタン距離、近似直線の傾きである。上下左右の方向の頻度は、saccade の角度が、x 軸の正の方向に対して 335°~25°, 65°~115°, 155°~205°, 245°~295° のいずれかに入っていれば、その方向について頻度を数える。次に、図 5 に示す青い点は、ある連続した fixation である。分位点間のマンハッタン距離は、図 5 の赤い線で示す通り fixation の x 座標の分布の 5% の分位線と y 座標の分布の 5% の分位線の交点、そして、

表 1 手法ごとの特徴量 (() 内は特徴量数を表す)
Table 1 A table of each features.

	特徴量の詳細
Bulling らの特徴量 (15)	saccade の長さの平均と分散, fixation の割合, まばたきの割合, saccade を大きさで 24 クラス分類したときのクラス内での値の平均と分散, n-gram 法を適応したときの出現回数の最大値と平均値
吉村らの特徴量 (7)	fixation の合計時間と平均時間, saccade の長さの平均と最小値, saccade の x 成分の合計値, y 成分の合計値, ウェーブレット変換後の x 成分の平均
Kunze らの特徴量 (8)	saccade の角度の平均と分散, 上下左右方向の回数, 分位点間のマンハッタン距離, 近似直線の傾き

x 座標の分布の 95% の分位線と y 座標の分布の 95% の分位線の交点のマンハッタン距離とする。このマンハッタン距離は、視点を動かす領域の大きさを表していると考えられる。また、近似直線の傾きは、図 6 に示すようにある連続した fixation を最小二乗法を用いて直線に近似し、その直線の傾きを特徴量としている。縦書きの文書だと傾きは水平になると考えられる。

3.4 特徴量の選定方法

学習データを用いて識別器を生成する際、検出などに有効な特徴量のみでなく不要な特徴量が含まれている場合がある。その場合、識別精度の向上を妨げるおそれがあるので、有効な特徴量のみを用いることが望ましい。特徴量の中から検出などに有効なものを選定する方法の一つに、Backward stepwise selection[14]がある。本稿では、読書行動の検出における有効な特徴量を選定するために、この手法を用いる。Backward stepwise selection は、まず、全ての特徴量の中から一種類ずつ特徴量を取り除いて精度を比較する。特徴量を取り除く前と取り除いた後の結果を比べ、精度が向上していれば、その特徴量を削除する。そして、さらに一種類ずつ特徴量を取り除き、精度を比較する。これを精度が向上する限り、繰り返すことで、最終的に特徴量が選定される。

3.5 読書行動の検出

3.4 までで得られた特徴量に対し、SVM[12]を用いることで読書行動を検出する。SVM は現在知られている手法の中でも認識性能が優れた手法の一つである。学習データから 2 クラスの識別器を生成する。その際、サポートベクトルを基準として、ユークリッド距離が最大になるように識別境界を設定することで、高精度な識別を行う。本稿では、学習データを入力とし、SVM によって識別器を生成する。その識別器にテストデータを入力することで、読書行動か非読書行動かを出力し、その結果から読書行動を検出する。

4. 実験

作成したデータセットの説明をしたのち、Bulling らの手法、吉村らの手法、Kunze らの手法とそれらの特徴量を

全て用いた手法の比較実験、そして、それらに Backward stepwise selection を実行した実験について述べ、その結果と考察について述べる。

4.1 データセット

読書行動検出のためのデータセットを作成した。被験者は 20 歳から 24 歳の日本人男性 10 人である。SMI 社の装着型アイトラッカ Eye Tracking Glasses 2.0 (ETG2.0) を使用した。サンプリング周波数は 30Hz である。被験者には読書行動として、雑誌、漫画、小説、教科書、新聞、論文の 6 種類の文書を一度ずつ読んでもらった。また、非読書行動として、歩く、会話する、動画を見る、食事をする、ゲームをする、パズルをとくなど、自由な行動を複数種類行ってもらった。一つの行動は連続した 6 分以上という時間制限を設定した。行動の順番は全て被験者の自由とした。一人あたりの撮影時間は連続した約 50 分から 80 分で、そのうちの読書時間はおよそ半分である。読書行動の正解データは、実験撮影中に被験者本人がラベル付けている。図 7 はアイトラッカのシーンカメラから得られた画像と視線情報の一例である。

4.2 実験条件

データセットを用いて、Bulling らの手法、吉村らの手法、Kunze らの手法、そして全ての特徴量を用いた手法の検出精度を出した。precision, recall の調和平均である F 値を被験者ごとに出し、その平均を結果とした。検出精度の評価には Leave-one-subject-out Cross-validation を用いた。そして、それぞれの手法ごとに Backward stepwise selection を行い、特徴量を選定して検出精度を計算した。実験に使用した計算機の OS は Mac OS 10.10.5, CPU は 2.7GHz Intel Core i7, メモリは 16GB 1600MHz DDR3 であった。

4.3 結果と考察

Bulling らの手法、吉村らの手法、Kunze らの手法、そして全ての特徴量を用いた手法で読書行動の検出を行った。検出精度の結果を表 2 に示す。従来手法よりも、全てを用いた手法の方が F 値が高いことがわかる。このことから、それぞれの手法に、読書行動検出に有効な特徴量が



図 7 アイトラッカのシーンカメラから得られた画像と視線情報の一例。
画像中の赤い点は被験者の視点を表す。

Fig. 7 Examples of images captured by a scene camera on the eye tracker.
Red dots show gaze points.

表 2 読書行動の検出精度 (%)
Table 2 Reading detection results.(%)

	precision	recall	F 値
Bulling らの手法	87.6	87.8	86.3
吉村らの手法	80.3	92.0	84.0
Kunze らの手法	84.5	92.6	87.6
全てを用いた手法	89.1	90.7	88.8

表 3 Backward stepwise selection による特徴選択結果

Table 3 Feature selection results by the backward stepwise selection method.

	特徴量選択前		特徴量選択後	
	特徴量数	F 値 (%)	特徴量数	F 値 (%)
Bulling らの手法	15	86.3	12	88.2
吉村らの手法	7	84.0	3	84.3
Kunze らの手法	8	87.6	4	88.3
全てを用いた手法	30	88.8	24	90.5

含まれている可能性があると考えられる。次に、それぞれの手法に Backward stepwise selection を実行した。結果を表 3 に示す。表 3 を見ると、どの手法も特徴量数が減少し、精度が向上していることがわかる。よって、有効な特徴量が選定されたと考えることができる。また、選定後の F 値を比較すると、全てを用いた手法が最も高いことがわかる。さらに、全ての手法を用いた特徴量では、特徴量数が 30 から 24 に減少したことがわかる。削除された特徴量を表 4 に示す。全ての手法を用いた特徴量の中で、削除された特徴量について考察する。まばたきの割合は、個人差があることや、そのときの健康状態など様々な要因によって異なってくることが示されている [15], [16]。そのため、読書行動か非読書行動かを判別する特徴量としては有効でなかったと考える。fixation の合計時間は、fixation の平均時間と特徴量の内容が重複しているのが有効でなかったと考えられる。次に、saccade の角度の平均について、縦書

表 4 backward stepwise selection により削除された特徴量
Table 4 Deleted features by backward stepwise selection.

Bulling らの手法	まばたきの割合 n -gram 法 ($n=2$) 適用時の最大出現語回数 saccade 右方向の長さの平均値
吉村らの手法	fixation の合計時間
Kunze らの手法	saccade の角度の平均 saccade 下方向の回数

きの文書や横書きの文書など、様々な文書を用いたことから、角度の平均からは読書行動を検出することは難しいと考える。その他、saccade 右方向の長さの平均値と saccade 下方向の回数、 n -gram 法適用時の最大出現語回数に関しては、用いた文書の種類やユーザなどに依存していたため、読書行動の検出には有効でなかったと考えられる。

5. まとめ

本稿では、リーディングライフログの自動記録化を最終的な目標とし、モバイルイトラッカを用いて、日常生活を想定した様々な文書を用いた読書行動の検出を行った。そして、Bulling らの手法、吉村らの手法、Kunze らの特徴量と、それらの特徴量を全て用いた手法の読書行動の検出精度の比較を行った。また、全てを用いたときの 30 個の特徴量に Backward stepwise selection を実行することで、読書行動の検出に有効な特徴量を 24 個選定した。選定後、全てを用いた手法よりも検出精度が向上した。

今後の課題は、モバイルイトラッカの一人称視点画像を読書行動の検出に同時に用いることで、検出精度をさらに向上させることである。

参考文献

- [1] 志賀優毅, 内海ゆづ子, 岩村雅一, カイクンツェ, 黄瀬浩一: 一人称視点画像を用いた文書画像の分類, 情報処理学会研究報告コンピュータビジョンとイメージメディア

- (CVIM), Vol. 15, pp. 1–7 (2014).
- [2] 川市仁史, Kunze, K., 黄瀬浩一: Reading-Life Log の実現に向けた万語計の提案, 電子情報通信学会技術研究報告. PRMU, パターン認識・メディア理解, Vol. 113, No. 493, pp. 79–84 (2014).
 - [3] Kimura, T., Huang, R., Uchida, S., Iwamura, M., Omachi, S. and Kise, K.: The Reading-Life Log–Technologies to Recognize Texts That We Read, *Proceedings of 12th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 91–95 (2013).
 - [4] 村山 功, 長崎栄三, 益川弘如, 酒井宜幸, 藤井宜彰: 読書活動と学力・学習状況調査の関係に関する調査研究., 全国学力・学習状況調査の分析・活用の推進に関する専門家検討会議 (第 17 回) (2010).
 - [5] Wan, S. and Aggarwal, J.: Mining Discriminative States of Hands and Objects to Recognize Egocentric Actions with a Wearable RGBD Camera, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 36–43 (2015).
 - [6] Pirttikangas, S., Fujinami, K. and Nakajima, T.: Feature selection and activity recognition from wearable sensors, *Proceedings of 3rd International Symposium on Ubiquitous Computing Systems*, pp. 516–527 (2006).
 - [7] Kunze, K., Utsumi, Y., Shiga, Y., Kise, K. and Bulling, A.: I know what you are reading: recognition of document types using mobile eye tracking, *Proceedings of the 2013 International Symposium on Wearable Computers*, pp. 113–116 (2013).
 - [8] Campbell, C. S. and Maglio, P. P.: A robust algorithm for reading detection, *Proceedings of the 2001 workshop on Perceptive User Interfaces*, ACM, pp. 1–7 (2001).
 - [9] 吉村和代, 川市仁史, Kunze, K., 黄瀬浩一: アイトラッカで取得した視点情報と文書理解度の関係, 電子情報通信学会技術研究報告, Vol. 112, No. 495, pp. 261–266 (2013).
 - [10] Bulling, A., Ward, J. A., Gellersen, H. and Tröster, G.: Eye movement analysis for activity recognition using electrooculography, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 4, pp. 741–753 (2011).
 - [11] Jalal, A., Kamal, S. and Kim, D.: A depth video sensor-based life-logging human activity recognition system for elderly care in smart indoor environments, *Sensors*, Vol. 14, No. 7, pp. 11735–11759 (2014).
 - [12] Chang, C.-C. and Lin, C.-J.: LIBSVM: A library for support vector machines, *ACM Transactions on Intelligent Systems and Technology (TIST)*, Vol. 2, No. 3, p. 27 (2011).
 - [13] Buscher, G. and Dengel, A.: Gaze-based filtering of relevant document segments, *Proceeding of International World Wide Web Conference*, pp. 20–24 (2009).
 - [14] Guyon, I. and Elisseeff, A.: An introduction to variable and feature selection, *The Journal of Machine Learning Research*, Vol. 3, pp. 1157–1182 (2003).
 - [15] Prause, J. U. and Norn, M.: Relation between blink frequency and break-up time?, *Acta ophthalmologica*, Vol. 65, No. 1, pp. 19–22 (1987).
 - [16] Stern, J. A., Boyer, D. and Schroeder, D.: Blink rate: a possible measure of fatigue, *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 36, No. 2, pp. 285–297 (1994).