文書レビュー時の視線計測情報によるレビュー品質の解析

Analysis of Review Quality by Using Gaze Data During Document Review

斉　藤　功　樹

要　約　ソフトウェア開発では，上流工程における仕様書や設計書の品質が重要であり，様々なレビュー手法が存在する．レビューの定量的な品質評価は，実施率や不具合検出率が用いられるが，それらの指標だけでは正確に評価できない．

近年，視線計測情報を用いた文書の読み方に関する研究が盛んに行われている．そこで，本研究では，要件定義書レビューを模した実験を行い，視線情報から品質を評価した．

その結果，レビュー品質には「瞬目時間の割合」と「瞬目回数の割合」の二つの特徴量が重要であり，品質の高いレビューアの瞬きは，平常時から増減された状態にあることが分かった．平常時よりも減っている場合には，レビュー時間に比例して品質の向上が確認された．一方，平常時よりも瞬きが増す場合の品質は分類できていないため，今後の研究課題となった．

Abstract　In software development, it is important to ensure the quality of specification and design document in upstream process. And there are many review methods. Although defect detection rate and review rate are used for the quantitative evaluation of review quality, the review quality can't be accurately evaluated only with those indices.

Recently, researches on how to read documents using gaze data have been actively conducted. Therefore, in this paper, we paid attention to reviewer's gaze data during document review, conducted experiments simulating the requirement definition document review and evaluated the review quality by using gaze data.

As a result, we found that two feature values of "the proportion of the number of blinks" and "the proportion of time spent blinking" are important for review quality and the blinks of the reviewer with high review quality is in a state fewer/more from normal time. In the case that the blinks fewer than the normal time occur, the improvement of review quality was confirmed in proportion to the review time. On the other hand, in the case that the blinks more than the normal time occur, it can't classify the review quality, which is a future research question.

1. は　じ　め　に

ソフトウェア開発では，上流工程における仕様書や設計書の品質が下流工程の成果物の品質にも影響を及ぼすため，当該仕様書の品質を担保する様々なレビュー手法が存在する．しかし，レビューそのものの品質（以降，特に記載しない限り，品質とはレビューの品質を指す）を左右するのは，手法の違いよりも個人の能力差のほうが大きい[1]．また，同一人物であっても，時間的な制約や集中度合いなどにより，品質が異なることも問題となっている．さらに，品質の評価において，レビュー実施率や不具合検出率による定量的な評価だけでは，その妥当性を正確に判断できない．例えば，不具合検出率の低さの原因が，レビューアの能力の低さ/レ

ビュー対象文書の品質の高さのどちらに起因するものであるのか考慮する必要がある．

近年，lifelog[*1] に関する様々な研究が盛んに行われている中，文章を読んでいる時の理解度を視線計測情報から推定する研究[2]や後述する様々な研究にみるように，視線計測情報を応用した文書の読み方の分析が試みられている．そこで，本研究では文書レビュー時のレビューアの視線と品質との間には何か特徴的な関係があるのではないかとの考えから，視線計測情報の分析を行った．視線情報を応用した先行研究としては，文書を流し読みしているか否かリアルタイムで判別する研究[3]がある．しかし，品質の評価には，流し読みの有無だけでは不十分で，他にもレビュー時の注意力や集中力，さらには文書への理解度などが影響すると考えられる．注意力や集中力が増すほど，瞬目回数が減ることは研究[4]で報告されている．JINS 社では，瞬きの情報から集中力を可視化するアルゴリズムが開発されている[5]．また，文書の理解度と視線情報の関連を分析した研究[2][6]や視線情報により英語のスキルを推定する研究[7]もある．

上述の研究[4]に関連して，瞬きの特徴としては，次のように言われている：

（A） 平常時の成人（20 代－90 代）の場合，1 分間の瞬きの回数（瞬目回数）は約 20 回であり，1 回の瞬きの時間（瞬目時間）は約 100－130ms である[8][9]．

（B） 好きな本を読んで集中する時や PC モニタを見ている時，瞬目回数は減る[4]．

（C） 何らかの緊張状態やストレスのかかった状態では，瞬目回数は増える[10][11]．

そこで，本研究では次の問題を考える：

文書レビュー中のレビューアの瞬きと品質の間にはどのような関連があるのか？

以下，2 章で提案手法を記述し，3 章で実験結果を示す．4 章で考察を述べ，5 章で本稿をまとめる．

2. 方　法

本章では，アイトラッカで収集する実験データの設定法（2.1 節）を説明し，収集した視線情報を用いた品質の評価手法（2.2 節）を記述する．

2.1　レビュー文書の作成方法と実験条件および被験者特性

実験に使用するレビュー対象文書は，社内で実際に使用された 3 種類の要件定義書を基にして，概要/機能要件/非機能要件の 3 頁の構成に改編し，サンプル文書を二つ加えた計 11 頁とした．各頁には後工程での障害に繋がる欠陥を意図的に含ませているため，欠陥が多いと文章の不自然さが増し，通常のレビュー時の視線の動きが計測されない可能性がある．そこで，1 頁あたり欠陥を含む文章は最大 2 個までとし，11 頁全体では欠陥を含む文章は 16 個とした．また，フォントの違いが読みやすさに影響を与えることが報告されているため[12]，レビュー対象文書の本文のフォントを MS 明朝に統一した．

被験者 19 名に対して，上述のレビュー対象文書をモニタ上に提示し，レビュー時の視線をアイトラッカで計測した．アイトラッカは視線情報を計測する専用のデバイスであり，本実験では gazepoint 社の GP3 HD Eye Tracker を用いた（図 1）．被験者には，モニタ上の要件定義書をレビュー後，紙の要件定義書に対して改善すべき箇所に下線を引くように指示した．下線を引いた箇所が，意図的に含んだ欠陥部分と一致した場合に欠陥を検出できたとする．なお，途中休憩は挟まず，レビューの時間や形式に制限は設けなかったが，測定の精度を確保するためになるべく頭部を動かさないように指示した．

実験終了後にアンケートを実施し，被験者特性（年齢/性別/要件定義書レビュー経験/文書レビュー経験/レビュー時の集中度/レビュー対象文書への理解度）を取得した．被験者 19 名の年齢は 30 ～ 50 代で満遍なく分布したが，約半数の 9 名の被験者は要件定義書レビュー経験が全くなかった（図 2）．被験者 19 名のレビュー時間の平均/最小/最大は，それぞれ 21 分/8 分/40 分であった．

図 1　実験の様子（モニタ下部にアイトラッカを設置して視線を計測）

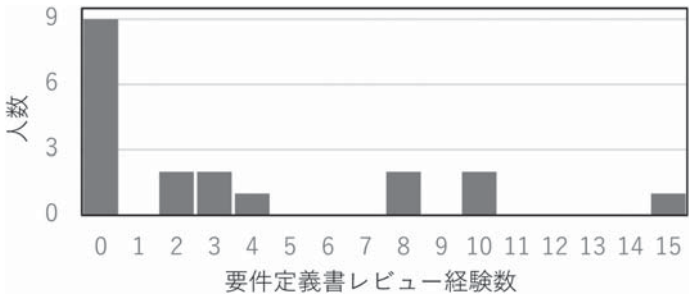

図 2　被験者の要件定義書レビュー経験分布

被験者 19 名に対して 11 頁分の視線情報を取得した計 209 頁のデータには，視線情報が正しく計測できていない頁が存在する可能性がある．その原因としてはアイトラッカが被験者の目の検知に失敗することが考えられるが，その場合，後述する固視（fixation）の情報が無効となる．そこで，スミルノフ・グラブス検定[13]を行い，有効な固視の割合が 0.6 未満の 4 頁分（図 3）を視線情報の計測不良データとして削除した．

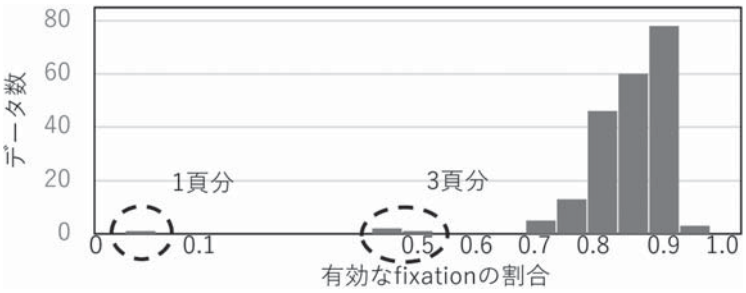

図 3　頁ごとの有効な fixation の割合のヒストグラム

2.2　品質評価手法

アイトラッカを用いてレビュー時の視線情報を取得し，レビュー対象文書の頁ごとに特徴量を算出する．視線情報から算出できる特徴量は多数あるが，品質と相関のない特徴量もある．そこで，まず，品質に影響を及ぼしている特徴量を抽出する（2.2.1 項）．その後，抽出した特徴量を使って機械学習を行い，品質を評価するモデル（以降，品質評価モデル）を構築する（2.2.2 項）．

なお，特徴量の抽出や品質評価モデルの構築に必要な統計解析は，R 言語（R version 3.4.1）[14] で行った．

2.2.1　特徴量の抽出

アイトラッカで取得できる次の四つの基本情報を基に，計 47 個の特徴量を算出する．

① 固視（fixation）：1 箇所を注視している視線の集まり
② 跳躍（saccade）：固視間の素早い目の動き（視線が移動した軌跡がわかる）
③ 瞬目（blink）：瞬きの有無
④ 瞳孔径（pupil）：瞳孔の大きさ（興味や関心がある際に瞳孔が大きくなる）

47 個の特徴量の内訳は，Bixler らの研究[15]で定義された 46 個の特徴量と本研究で追加した「瞬目回数の割合」である．頁ごとにレビュー時間が異なるため，瞬目回数はレビュー時間による影響を受ける．そこで，本研究ではレビュー時間による影響を排除するため瞬目回数をレビュー時間で割った「瞬目回数の割合」を追加した．

英語の理解度と視線の関連を調査した研究[6]では，特定の視線情報が理解度と相関した．これと同様に，全 47 個の特徴量の内，特定の特徴量が品質に影響すると考えた．そこで，Random forest（以降，RF）[16]による機械学習アルゴリズムを使って，品質に影響を及ぼす特徴量を抽出した．RF は各特徴量の重要度も同時に算出するので，重要度の低い特徴量は品質への影響は少ないとして削除できる．

2.2.2　品質評価モデル

抽出した特徴量を用いて，品質評価モデルを複数のアルゴリズムで構築する．本研究で採用したアルゴリズムは，RF，Decision tree（以降，DT），k-nearest neighbor（以降，kNN），および Weighted support vector machine（以降，WSVM）[17]である．特に，WSVM の場合は，データの重み付けにより，品質が悪い人と良い人の分布が不均一な場合にも適切な対応ができる．

各アルゴリズムによるレビュー評価モデルの有効性の確認は交差検証[*2] を行い，正解率（Accuracy），再現率（Recall）を算出して比較する．正解率は予測した品質と実際の品質との一致率であり，再現率は品質が悪い人を正とした場合，正の人を正しく予測できた割合である．それぞれ式(1)，(2)にて導出する．

$$\mathrm{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

$$\mathrm{Recall} = \frac{TP}{TP + FN} \tag{2}$$

ここで，TP，TN，FP，FN は表 1 に示す混合行列の各要素である．

表 1　真の結果と予測結果の混合行列

		真の結果	
		正	負
予測結果	正	*TP*	*FP*
	負	*FN*	*TN*

3. 結　果

本章では，前章に示した方法により得られた実験結果を示す．まず，使用する実験データについて説明する（3.1 節）．その後，評価モデルに使用する特徴量の抽出結果を示し（3.2 節），品質評価モデルの性能結果を示す（3.3 節）．なお，R 言語による統計解析は，Intel/Core i7 920 の CPU を搭載したデスクトップ PC で行った．

3.1　使用する実験データ

使用する実験データは，2.1 節で述べた方法で収集した．全 16 個の欠陥を含む 11 頁の実験データから 19 名の被験者が欠陥を幾つ検出できたのか，欠陥検出数と要件定義書レビュー経験の関係を図 4 の左図に示す．この結果から次の事実が確認できる：

- 要件定義書レビュー経験は 0 であるが，欠陥を検出できる人が存在する．
- 反対に，経験は多いが欠陥を検出できない人も多い．

要件定義書レビュー経験が多い人は high グループと予想したが，以上の事実から本研究の実験データでは，レビュー経験数は品質には関係していないことを確認した．

以降，欠陥検出数が T（＞0）個以下の被験者を品質が低い low グループ，検出数が T より大きいグループを high グループとする．本研究では，T = 1，2，3 の 3 ケースを扱う．図 4 の右図は T = 1 の場合である．

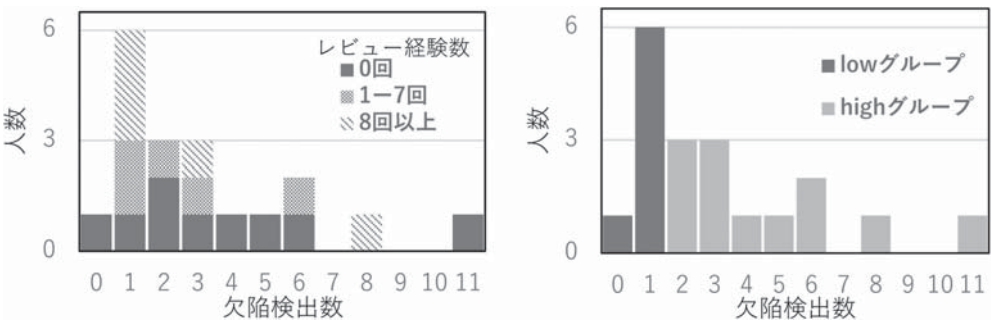

図 4　欠陥検出数ごとの要件定義書レビュー経験の内訳（左）と T = 1 の被験者分布（右）

3.2　特徴量の抽出

47 個の特徴量について，RF を 3 回実行して得られた重要度上位 4 個を表 2 に示す．RF の計算時間は約 6 秒（1 回の計算時間は約 2 秒）であった．

T が変化しても，「瞬目時間の割合」と「瞬目回数の割合」は上位 4 位以内で共通の特徴量

であった．したがって，当該二つの特徴量が品質に影響を与える重要な特徴量であると考えた．ここで，「瞬目時間（回数）の割合」とは，瞬目時間（回数）をレビュー時間で割った値である．

表 2　RF により得られた平均重要度の上位 4 項目（RF の実行は 3 回）

順位	T = 1）特徴量（重要度）	T = 2）特徴量（重要度）	T = 3）特徴量（重要度）
1	**瞬目時間の割合(12.31)**	**瞬目時間の割合(9.49)**	**瞬目時間の割合(6.84)**
2	**瞬目回数の割合(10.77)**	跳躍の持続時間の平均(6.24)	跳躍の持続時間の歪度(4.90)
3	跳躍の持続時間の平均(5.87)	**瞬目回数の割合(6.24)**	跳躍の回数(4.33)
4	跳躍の持続時間の標準偏差(5.42)	跳躍の角度の尖度(5.18)	**瞬目回数の割合(4.27)**

3. 3　品質評価モデル

「瞬目時間の割合」と「瞬目回数の割合」の二つの特徴量に基づいて，品質評価モデルを構築し，採用するアルゴリズムごとの性能を評価した．得られた結果を表 3 に示す．表 3 から次の事実が確認できる：

- T = 1 と T = 2 の場合，DT による品質評価モデルの正解率が最も高い．
- T = 3 の場合，kNN による品質評価モデルの正解率が最も高い．
- T の増加につれて，アルゴリズムによらず正解率は下がる．
- T = 1 と T = 3 の high と low のデータ数は逆転し，WSVM 以外のアルゴリズムは再現率の性能も同様に逆転しているが，WSVM の high と low に対する再現率の性能差は小さい．

なお，モデル構築の計算時間は RF/DT/kNN/WSVM のそれぞれに対し，6 秒/3 秒/2 秒/3 秒であった．

表 3　アルゴリズムごとの品質評価モデルの性能

T	評価指標	項目	データ数	モデル構築アルゴリズム			
				RF	DT	kNN	WSVM
1	Recall	high	131	84.3%	**90.9%**	89.8%	86.3%
		low	74	75.1%	77.3%	75.4%	**80.3%**
	Accuracy	all	205	79.6%	**84.0%**	82.4%	83.1%
2	Recall	high	99	73.1%	**78.2%**	71.8%	73.7%
		low	106	65.7%	63.9%	**66.3%**	62.5%
	Accuracy	all	205	69.3%	**71.0%**	69.0%	68.0%
3	Recall	high	66	37.3%	43.4%	43.0%	**56.1%**
		low	139	73.8%	76.5%	**78.9%**	58.5%
	Accuracy	all	205	55.1%	59.5%	**60.5%**	57.1%

以上の結果をまとめると，47 個の特徴量のうち，「瞬目時間の割合」と「瞬目回数の割合」は T = 1，2，3 に対して，常に上位の重要度を示し（特に，前者は常に 1 位），当該二つの特

徴量によって構築した品質評価モデルの性能はTの増加につれて低下したものの，T = 1の場合には84%の精度で分類できた．

4. 考　察

本章では，前章で着目した「瞬目時間の割合」と「瞬目回数の割合」の二つの特徴量に対して，さらなる実験結果によりその有効性を示し（4.1節），関連する実験データを示しながら文書レビューの品質に関わる瞬目の特徴を議論する（4.2節と4.3節）．

4.1　二つの特徴量の有効性

T = 1の場合，本研究で着目した二つの特徴量「瞬目時間の割合」と「瞬目回数の割合」は重要度の上位2位であり，これらで構築した品質評価モデルは，アルゴリズムによって数%の性能のバラツキはあったが，良い性能を示した（3.3節）．そこで，当該二つの特徴量で構築したモデルの有効性を確認するために，47個の全特徴量で構築した品質評価モデルの結果を表4に示す．この結果から次の事実が確認できる：

- 2個の特徴量による評価モデルの結果（表3）と比較して，RF以外のアルゴリズムでは，性能が著しく劣化した．
- RFでは再現率の性能は向上したが，正解率が同程度に劣化した．

以上の事実から，抽出した2個の特徴量が品質評価には有効だと分かった．品質評価モデルは，品質に無関係な特徴量を利用せず，適切な特徴量で構築すべきであることを示唆している．

表4　47個の全特徴量による品質評価モデルの性能（T = 1の場合）

評価指標	項目	データ数	モデル構築アルゴリズム			
			RF	DT	kNN	WSVM
Recall	high	131	**88.1%**	81.5%	74.9%	54.0%
	low	74	61.5%	**76.0%**	54.1%	61.4%
Accuracy	all	205	74.6%	**78.7%**	64.6%	57.4%

4.2　二つの特徴量と品質の関係

1章で述べた(B)や既存研究[4]に従えば，注意深く文章を読んでいる時，瞬目は平常時よりも減ると考えられる．(A)により，本研究で着目する特徴量「瞬目時間の割合」に対する平常時の値は以下で与えられる：

下限値：B_L = 0.033（= 0.10[s/回] × 20/60[回/s]）

上限値：B_U = 0.043（= 0.13[s/回] × 20/60[回/s]）

そこで，実験で得られた「瞬目時間の割合」と「瞬目回数の割合」の散布図とそれを楕円で抽象化した図との関係を図5に示す．この結果から次の事実が確認できる：

1. 瞬目時間の割合と瞬目回数の割合は，値が大きくになるにつれ（図の右上に向かって）バラツキが大きくなる．
2. lowグループは，T = 1の場合，平常時の瞬きの状態（= B_L と B_U の間）に存在したが，Tの増加につれて全域に広がっている．

3. T = 1，2，3 のすべて場合において，high グループは平常時の瞬きの状態には存在せず，B_L 以下と B_U 以上の 2 領域に存在した．

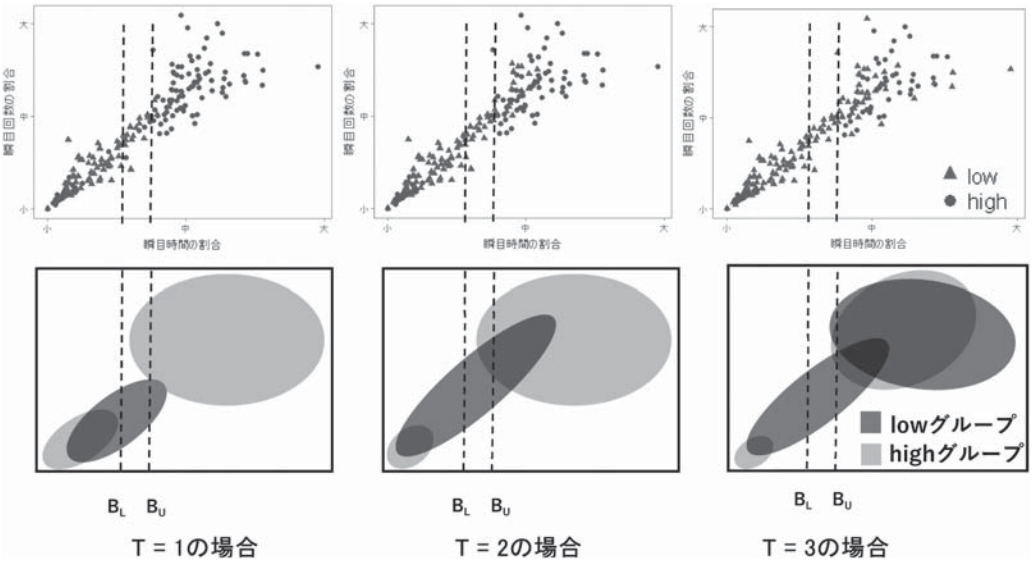

図 5　瞬目時間の割合と瞬目回数の割合の散布図（上段）とその抽象図（下段）

1 の事実に関して，「瞬目時間の割合」だけでなく「瞬目回数の割合」を加えた 2 個の特徴量を用いることの有効性を支持するものである．

3 の事実に関して，high グループが B_L 以下に存在したという結果は，1 章で述べた事実(B)や既存研究[4]の結果に合致する．一方，high グループが B_U 以上にも存在したという結果は，関連研究での報告はないが，(C)に合致する．

2 と 3 の事実に共通して，T = 1 の結果は重要な意味がある．T の増加につれて，low (high) グループの分布は拡大（縮小）する一方であり，T = 1 で得られた low の事実，すなわち，平常時の瞬きの状態は low グループであったという事実は，T = 2，3 でも成立する．また，排反事象である high グループについては，low と逆のことが成立する．すなわち，T = 1 で平常時の瞬きの状態は low グループであるのだから，high グループは平常時の瞬きの状態でない範囲にしか存在しない．しかしながら，T の増加につれて high グループと low グループの混在範囲が増え，二つのグループを適切に分割することが難しくなるという問題がある．この問題により，3.3 節の品質評価モデルで T の増加につれて性能が劣化してしまった．

次節では，B_L 以下の high グループと low グループを分割する方法を示す．

4.3 「瞬目時間の割合」とレビュー時間による品質の分類

「瞬目時間の割合」と「瞬目回数の割合」の二つの特徴量は共に，それぞれの瞬き情報をレビュー時間（秒）で割っており，図 5 からはレビュー時間による瞬きの変化は確認できない．high グループは注意深く文書を読むため，レビュー時間が長くなると想定される．そこで，19 名の被験者ごとの「瞬目時間の割合」とレビュー時間の関係を図 6 に示す．縦軸の被験者は，レビュー時間で下から昇順とし，T = 3 の場合として high と low を分類した．この結果から，

4.2 節の図 5 と同様の事実が確認できるが，新たな事実としては以下の点が確認できる．

- 「瞬目時間の割合」が B_L 以下の被験者はレビュー時間で high と low に分かれる．
- 「瞬目時間の割合」が B_U 以上かつレビュー時間が短い（18 分以下）被験者は，high と low の分類によらず存在しなかった．

以上の事実から，レビュー時間を考慮して品質を分類できるのは，「瞬目時間の割合」が B_L 以下の被験者だけであり，B_U 以上の被験者を分類するという問題は解決できておらず，今後の研究課題である．

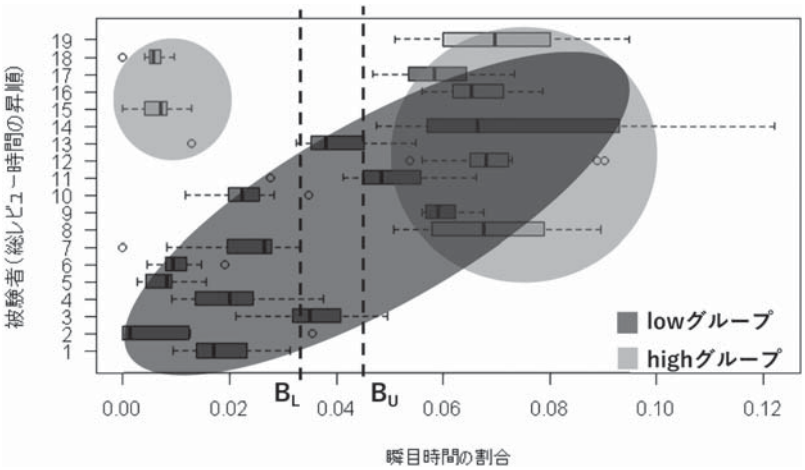

図 6　被験者ごとの瞬目時間の割合の箱ひげ図（T ＝ 3 の場合）

最後に，「瞬目時間の割合」とレビュー時間に関する実験データの分析結果を表 5 に示す．この結果から次の事実が確認できる：

- 図 6 の「瞬目時間の割合」が B_L 以下で high グループを示す被験者 No. 15 と No. 18 は，レビュー時間が長い（27 分と 34 分）にも関わらず，レビュー開始時（初めの 3 頁）と終了時（最後の 3 頁）で瞬目時間の割合の値があまり変化していない．
- 同様に，T ＝ 3 で「瞬目時間の割合」が B_U 以上の high グループの被験者も，レビュー開始時と終了時で瞬目時間の割合の値があまり変化していない．
- 逆に，レビュー開始時と終了時で瞬目時間の割合の値が大きく変化した（終了時の方が大きくなった）のは No. 14 と No. 17 である．これらは T ＝ 1，2 では high グループに，T ＝ 3 では low グループに分類された被験者である．

以上の事実から，次のことが推察される：レビュー時間が長い人は後半には疲労度が増して集中力が下がり，それが瞬目の回数や時間を増す要因になった可能性が考えられる．しかし，T ＝ 3 の high グループの被験者はレビュー開始と終了時の「瞬目時間の割合」に変化は見られず，レビュー後半にパフォーマンスが落ちることは確認されなかった．そのようなことが確認されたのは，（T ＝ 2 では high グループだったが）T ＝ 3 で high グループにならなかった被験者であった．つまり，真に high グループの被験者の「瞬目時間の割合」は，疲労や集中力の低下に左右されず，レビュー中は一定であると考えられる．

表5　レビュー時間とレビュー開始 / 終了時の瞬目時間の割合

被験者 No	レビュー時間(分)	瞬目時間の割合		T = 1	T = 2	T = 3
		初めの3頁の平均値	最後の3頁の平均値			
1	6	0.018	0.020	low	low	low
2	7	0.021	0.001	high	low	low
3	9	0.036	0.038	low	low	low
4	12	0.019	0.020	low	low	low
5	12	0.007	0.006	high	high	low
6	14	0.014	0.010	low	low	low
7	18	0.022	0.027	low	low	low
8	19	0.074	0.060	high	high	high
9	20	0.058	0.061	high	low	low
10	20	0.028	0.021	low	low	low
11	21	0.054	0.051	high	low	low
12	22	0.067	0.064	high	high	high
13	22	0.038	0.046	low	low	low
14	26	**0.054**	**0.095**	high	high	low
15	27	0.005	0.008	high	high	**high**
16	30	0.064	0.065	high	high	high
17	33	**0.054**	**0.067**	high	high	low
18	34	0.008	0.005	high	high	**high**
19	40	0.085	0.070	high	high	high

4. 4　制限事項

本研究で使用したアイトラッカの H/W 上の制約から，①実験で使用した要件定義書は，通常よりも行間を広めに設定した．それ以外では，②要件定義書の図表は削除して文章のみにしていることに加え，③当該文書は PC モニタ上に表示して視線情報を取得している．

①については，通常の行間の文書でも本研究の提案手法は適用できると考えているが，②と③の条件を解除した場合の適用については，今後の研究で検証する．

5.　結　論

文書レビュー時の視線計測情報に基づいて，品質の高いレビューアにはどのような特徴があるのかを分析した．その結果，アイトラッカによる計測で取得できる四つの情報（固視/跳躍/瞬目/瞳孔径）の内，瞬目が重要であることが分かった．そこで，本研究では，特に，二つの特徴量「瞬目時間の割合」と「瞬目回数の割合」に着目して品質評価モデルを構築し，性能を評価したところ，T = 1 の場合には高精度の結果が得られたが，T の増加につれて精度は劣化

した．また，重要度の高い「瞬目時間の割合」とレビュー時間の関係を分析したところ，次の結果を得た：

- レビューアの瞬きが平常時の場合，レビューアは low グループに属した．
- high グループの瞬きは，平常時よりも少ない時と多い時があった．
- high グループの瞬きが平常時よりも少ない場合，レビュー時間は長かった．

特に，2 番目の“high グループの瞬きは，平常時よりも多い時があった”という結果は，既存の関連研究では示されておらず，本研究の重要な成果であると考える．

6. お わ り に

本稿では文書レビュー時の視線情報を用いて，品質評価モデルを構築した．視線情報の内，瞬きが重要であり，瞬きの違いにより品質を分類できた．いくつかの制限事項はあるものの，レビュー時の視線情報を用いて，頁ごとの品質を評価できるモデルを構築できたことは有益である．

今後の研究の展望は次の通りである．T の増加に伴う品質評価モデルの性能劣化の問題の解決には，「瞬目時間の割合」が B_U 以上の場合，二つの特徴量とレビュー時間では high と low グループを適切に分類できないため，キーとなる特徴量を見出し，新たな特徴量を組み込んだ品質評価モデルによって性能向上を目指す．また，制限事項の解除に向けた取り組みの中でも，特に，③のモニタ上ではなく紙でレビューする場合については早期に検討したい．

＊ 1　人間の生活行動をデジタルデータとして記録すること，またはそのデータのこと．
＊ 2　交差検証（Cross validation）は標本データから一部を検証データとして抜き出し，残りのデータでモデルを構築し，検証データでモデルの妥当性を検証する手法である．

参考文献
［1］ H. Uwano, M. Nakamura, A. Monden, and K. Matsumoto, “Analyzing Individual Performance of Source Code Review Using Reviewers’ Eye Movement”, In Proceedings of the 2006 Symposium on Eye Tracking Research & Applications, 2006.03, pp. 133–140
［2］ 吉村和代，川市仁史，K. Kunze，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：信学技報 112(495)，2013 年 3 月，P261 ～ 266
［3］ R. Biedert, J. Hees, A. Dengel, and G. Buscher, “A robust realtime reading-skimming classifier”, In Proceedings of the Symposium on Eye Tracking Research and Applications, vol. 1, no. 212, 2012.03, p. 123
［4］ H. Ledger, “The effect cognitive load has on eye blinking”, The Plymouth Student Scientist, Vol. 6, No. 1, 2013, pp. 206–223
［5］ Y. Uema and K. Inoue, “JINS MEME algorithm for estimation and tracking of concentration of users”, In Adjunct Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers, 2017.09, pp. 297–300
［6］ A. Okoso, T. Toyama, K. Kunze, J. Folz, M. Liwicki, and K. Kise, “Towards Extraction of Subjective Reading Incomprehension: Analysis of Eye Gaze Features”, In Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, 2015.04, pp. 1325–1330
［7］ O. Augereau, K. Kunze, H. Fujiyoshi, and K. Kise, “Estimation of english skill with a mobile eye tracker”, In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct, 2016.09, pp. 1777–1781
［8］ 杉山敏子，田多英興，「成人における内因性瞬目の年齢差と性差」，生理心理学と精

神生理学，25(3)，2007 年 12 月，P225 ～ 265
［9］ G. GORDON, “Observations upon the movements of the eyelids”, The British journal of ophthalmology, vol. 35, no. 6, 1951.06, pp. 339–351
［10］ 山田冨美雄，「瞬目による感性の評価」，Japanese Psychological Review，vol. 45, no. 1，2002 年 7 月，P20 ～ 32
［11］ 坪田智子，「（修士論文）ヒューマンインターフェースのための顔情報の計測に基づくユーザの心理状態推定」，奈良先端科学技術大学院大学，2001 年 3 月
［12］ 三枝竜，佐藤歩，竹本雅憲，窪田悟，佐々木愛，石坂博司，「電子書籍リーダー用の日本語フォントの読みやすさの比較評価」，日本人間工学会大会講演集，vol. 48，2012 年，P414 ～ 415
［13］ F. E. Grubbs, “Procedures for Detecting Outlying Observations in Samples”, Technometrics, vol. 11, no. 1, 1969.02, pp. 1–21
［14］ The R Project for Statistical Computing
［15］ R. Bixler and S. D’Mello, “Automatic Gaze-Based Detection of Mind Wandering with Metacognitive Awareness”, User Modeling, Adaptation and Personalization, 2015.06, pp. 31–43
［16］ L. Breiman, “Random forests”, Machine Learning, vol. 45, no. 1, 2001.10, pp. 5–32
［17］ C.-C. Chang and C.-J. Lin, ¨Libsvm¨ , ACM Transactions on Intelligent Systems and Technology, vol. 2, no. 3, 2011.04, pp. 1–27

執筆者紹介 斉 藤 功 樹（Koki Saito）
2009 年日本ユニシス(株)入社．金融機関向けのバックシステムの開発・保守を担当．2013 年に総合技術研究所に異動．大規模データ処理技術や衛星画像のデータ処理・データ分析に関する研究に従事し，現在は視線情報を用いた文章の読解に関する研究に従事．