Webページのユーザビリティとパフォーマンスに注目した信頼性評価手法の提案

山田 健太† Eint Sandi Aung† 山名 早人‡

†早稲田大学大学院基幹理工学研究科 〒169-8555 東京都新宿区大久保3-4-1
‡早稲田大学理工学術院 〒169-8555 東京都新宿区大久保3-4-11

E-mail: {yamada, esdaung, yamana}@yama.info.waseda.ac.jp

あらまし 昨今，フェイクニュースをはじめとする信頼性の低いWebページが問題視されている．信頼性の低いWebページを自動検知する研究の最新事例では，自然言語情報の利用やSNSなどのAPIを用いた特徴量を用いることが主流となっている．しかし，自然言語は利用の際に言語ごとに自然言語処理をしなくてはならないことや，APIは利用に制限がかかっていることが多い現状を踏まえ，信頼性評価システムの汎用性や継続性を損ねてしまうと考えられる．そこで，本研究では自然言語やAPIを利用せず，Google Lighthouseで定義される特徴量の中から特徴量選択により選択された特徴量をWebページの信頼性評価に用いることを提案する．Google LighthouseはWebページのユーザビリティやパフォーマンスを判定するためにGoogleによって開発されたツールである．評価実験ではFake News Corpusデータセットに対する既存手法と提案手法の2クラス，10クラスにおける分類精度と独自に収集した8,230ページに適用した場合の評価の2つの観点から評価を行った．2クラス分類の精度はF値0.898，独自に収集したページに適用した場合の結果は0.650となった．10クラス分類の精度評価ではマクロ平均，加重平均をとり，F値はそれぞれ0.340，0.468となった．2クラス分類と同様に独自に収集したページに適用した場合の結果は0.095となった．

キーワード 情報信頼性，フェイクニュース

1. はじめに

近年，個人用PCやスマートフォンの普及とともに様々な人が手軽にWebページにアクセスできるようになった．一方で悪質なデマや誤解を生むような表現やいわゆるフェイクニュースといわれる信頼性の低いWebページに対する対処はより一層重要視されている[1].

Webサイトの信頼性評価手法としては，Olteanuら[2]のWebページより得られるテキストに関する情報からなるContents特徴量とSNSのAPIや検索エンジンのページランクの情報を用いたSocial特徴量を併用し，機械学習によって評価する手法や，Wawerら[3]によるOlteanuらの提案する特徴量に加え自然言語処理によって得られる情報からなるLinguistic特徴量を使用する手法などが提案されている．

しかし，APIの利用に依存するSocial特徴量は，昨今のAPI利用に関する制約に伴い利用が困難となっている．さらに，Linguistic特徴量は言語毎に用意しなければならず汎用性の点で問題がある．そこで，大谷ら[4]はSocial特徴量やLinguistic特徴量を使用することなくWebページの信頼性評価をする新たな特徴量としてWebページに埋め込まれた広告に代表される外部コンテンツに着目したAD/NonAD特徴量を提案している．しかし，大谷らの手法は，外部コンテンツを含むWebページに特化している点， Linguistic特徴量を使ったモデルに対して信頼性評価の精度が劣ることが課題としてあげられる．

こうした背景のもと，本稿では，Social特徴量及びLinguistic特徴量を用いることなく，Webページの信頼性評価を高精度で行うことを目指している．具体的には，これまで信頼性判定の特徴量として着目されてこなかったWebページのユーザビリティとパフォーマンスに注目した特徴量を用いる．Webページのユーザビリティやパフォーマンスに関する特徴量として，GoogleLighthouse[1]を採用する．GoogleLighthouseは，ユーザビリティとパフォーマンスに関連する特徴量として合計132個の特徴量を持つ．これら特徴量に対してRFE(Recursive Feature Elimination)[5]を適用し，特徴量選択を行いWebページの信頼性評価を行う．著者らは，これまでにGoogle Lighthouseより得られる特徴量がWebページの信頼性に有効であることを示しており[6]，本稿ではより厳密に有効である特徴量を抽出し，Webページの信頼性評価のための特徴量として提案することを目指す．

本稿は次の構成をとる．まず，第2節で関連研究について述べる．次に第3節で提案手法の説明を行い，第4節で評価実験と結果を説明する．最後に，第5節でまとめを行う．

2. 関連研究

本項では，関連研究として，近年提案されたWebページの信頼性評価手法について述べる．

2.1 Content, Social特徴量

Olteanuら[2]はWebページの信頼性を評価するための特徴量として，表1に示すWebページより得られるテキストに関する情報からなるContents特徴量とSNSのAPIや検索エンジンのページランクの情報を用いたSocial特徴量の2種類の特徴量を提案し，統計手法に基づいて下線で示す22の特に有効である特徴量を選択している．検証用のデータセットとして被験者がランダムに抽出されたWebページの信頼性評価をReliable，Unreliableの2値でおこなったMicrosoft credibility corpus[7]を用い，Extremely Randomized Trees (ERT)[8]を使用して分類している．分類精度はF値で0.75であった．

一方，Social特徴量はSNSのAPIや検索エンジンのページランクの情報の利用が中心であるが，表2に示すように現在APIやページランクの提供は制限されてしまっていることが多く，Social特徴量の利用は信頼性評価システムの継続性を損ねてしまうことが懸念されている．

[1] Google Lighthouse, https://developers.google.com/web/tools/lighthouse

表1. Olteanuら[2]による提案特徴量

カテゴリ	タイプ	特徴量	説明
Content	Text	#Exclamations	テキスト中の“!”の数
		#Questions	テキスト中の“?”の数
		#Commas	テキスト中の”,”の数
		#Dots	テキスト中の”.”の数
		Token Count	単語数
		?Polarity	文章の感情極性
		#Positive	positiveな文章数
		#Negative	negativeな文章数
		#Subjective	客観的な文章数
		#Objective	主観的な文章数
		#Spelling Errors	スペルエラーの数
		Text Complexity	テキストの分散
		Informativeness	テキストの情報量 [9]
		SMOG	テキストの可読性 [10]
		Category	Webページのカテゴリ
		#NN	テキスト中の名詞数
		#VB	テキスト中の動詞数
		#JJ	テキスト中の形容詞数
		#RB	テキスト中の副詞数
		#DT	テキスト中の限定詞数
	Appearance	#AD	Webページ中の広告数
		AD Max Size	最大面積の広告のpixel
		AD Body Ratio	Webページ中の広告面積
		CSS Definition	CSSの定義数
	Meta Information	Domain Type	URLのドメイン
Social	Social Popularity	#fb_share	Facebook[2]の共有数
		#fb_like	FacebookのLike数
		#fb_comment	Facebookのコメント数
		#fb_click	Facebookのクリック数
		#fb_total	Facebookの共有, like, コメント, クリック数の合計
		#tweets	Twitter[3]のツイート数
		#bitly_clicks	Bitly[4]の短縮URLのクリック数
		#bitly_referrer	Bitlyのリファラー数
		#delicious_bookmarks	Delicious[5]のブックマーク数
	General Popularity	Alexa Rank	Alexa[6] Rank
		Google page Rank	Google PageRank[7]
	Link Structure	Alexa links in page rank	被リンク数

[2] Facebook, https://www.facebook.com/
[3] Twitter, https://twitter.com/
[4] Bitly, https://bitly.com/
[5] Delicious API has been abolished.
[6] Alexa, https://www.alexa.com/
[7] PageRank cannot be retrieved in general.

表2. API, ページランクの提供状況

サービス/API	制限
Google PageRank	2016に公開終了
Alexa	1,000/月[$]
Delicious	APIなし
Bitly	100/時 or 1,000/月[$]
Twitter	450/15分[$]
Facebook	4,800/日

$ … 課金次第で制限の緩和が表れるもの

2.2 Linguistic特徴量

Wawerら[3]は，Olteanuら[2]の特徴量をアップデートする形で自然言語処理を用いたLinguistic特徴量を提案している．Lingistic特徴量は文章中の感情評価をするGeneral Inquirer (GI) [11]と文章中の単語の重要度を扱うTF-IDFから成る特徴量である．Olteanuらの提案特徴量であるContent特徴量とSocial特徴量と併用する形で同一のMicrosoft credibility corpus [7]を用いて評価実験をおこなっている．分類器としてlogistic regression (LR)を用い，分類精度はF値で0.74～0.83でありOlteanuらの実験から0.02～0.05の向上を示している．

このように，Linguistic特徴量を用いることで分類精度を向上させることができるが，言語毎の形態素解析器構築が必須となる．しかし，現在のところ世界には100を超える言語が存在し，その全ての言語において形態素解析器を構築することは現実的でなく，Linguistic特徴量を利用することが信頼性評価システムの汎用性を損ねてしまう．

2.3 AD/Non-AD特徴量

Olteanuら[2]のSocial特徴量やWawerら[3]のLinguistic特徴量が信頼性評価システムの汎用性や継続性を損ねてしまうことを受けて，大谷ら[4]はSocial特徴量やLinguistic特徴量を使用することなくWebページの信頼性評価をする新たな特徴量としてWebページに埋め込まれた広告に代表される外部コンテンツに関する特徴量であるAD/Non-AD特徴量を提案した．評価用データセットとしては，Fake News Corpus[8]を用いている．分類アルゴリズムとしてGradient Boosting Decision Tree (GBDT)を用い，最も良い分類精度はLinguistic特徴量とAD/Non-AD特徴量を併用したものであり，F値で0.828であった．

しかし，AD/Non-AD特徴量は，外部コンテンツを含むWebページを対象にしている点において汎用性が劣る．また，AD/Non-AD特徴量とLinguistic特徴量の併用で最も高いF値を出しており，分類精度を高めるためにはLinguistic特徴量が必要になるという問題がある．

3. 提案手法

本項では本研究の提案手法であるWebページのユーザビリティとパフォーマンスに注目した特徴量について述べる．

3.1 Google Lighthouseより取得できる特徴量を用いた信頼性評価手法

Social特徴量やLinguistic特徴量を使用することなくWebページの信頼性評価をする新たな特徴量として，Webページのユーザビリティ及びパフォーマンスを用いることを提案する．具体的には，ユーザビリティ及びパフォーマンスを計測するために開発されたツールであるGoogle Lighthouseより得られる特徴量の一部を用いる．また，関連研究[2][3][4]と同様に信頼性評価に機械学習アルゴリズムであるExtreme Gradient Boosting (XGboost) を用いる．

[8] Fake news corpus, https://github.com/several27/FakeNewsCorpus

3.2 Google Lighthouseについて

Google LighthouseはPerformance，Accessibility，Best Practices，SEO，PWAの5つの観点からWebページの評価をするGoogleによって開発されたツールである．PerformanceはWebページ表示までの時間やページ読み込みの際の処理などWebページのパフォーマンスに関する指標である．Accessibilityはhtml要素の定義や命名などサイト全体のAccessibilityに関する指標である．Best Practicesはhttpsを使用しているかどうか等，最近のWebページ構築におけるベストプラクティスを考慮している．SEOはWebページのSEO対策に関する指標である．PWAはProgressive Web Appsの略称であり，アプリケーションのようにWebページを利用できる仕組みのことであり，指標としている．図1にGoogle Lighthouseの外観を，表3にGoogle Lighthouseそれぞれの指標中に含まれる特徴量を示す．

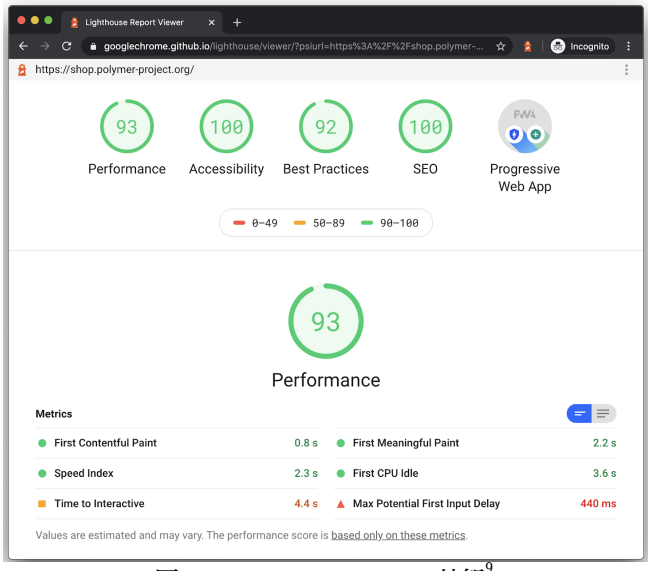

図1 Google Lighthouseの外観[9]

表3．Google Lighthouse[10]の各指標と特徴量

指標	*特徴量*
Performance	first-contentful-paint, first-meaningful-paint, speed-index, interactive, first-cpu-idle, max-potential-fid, estimated-input-latency, total-blocking-time, render-blocking-resources, uses-responsive-images, offscreen-images, unminified-css, unminified-javascript, unused-css-rules, uses-optimized-images, uses-webp-images, uses-text-compression, uses-rel-preconnect, time-to-first-byte, redirects, uses-rel-preload, efficient-animated-content, total-byte-weight, uses-long-cache-ttl, dom-size, critical-request-chains, user-timings, bootup-time, mainthread-work-breakdown, font-display, performance-budget, resource-summary, third-party-summary, network-requests, network-rtt, network-server-latency, main-thread-tasks, diagnostics, metrics, screenshot-thumbnails, final-screenshot
Accessibility	accesskeys, aria-allowed-attr, aria-required-attr, aria-required-children, aria-required-parent, aria-roles, aria-valid-attr-value, aria-valid-attr, audio-caption, button-name, bypass, color-contrast, definition-list, dlitem, document-title, duplicate-id, frame-title, html-has-lang, html-lang-valid, image-alt, input-image-alt, label, layout-table, link-name, list, listitem, meta-refresh, meta-viewport, object-alt, tabindex, td-headers-attr, th-has-data-cells, valid-lang, video-caption, video-description
Best Practices	is-on-https, uses-http2, uses-passive-event-listeners, no-document-write, external-anchors-use-rel-noopener, geolocation-on-start, doctype, no-vulnerable-libraries, js-libraries, notification-on-start, deprecations, password-inputs-can-be-pasted-into, errors-in-console, image-aspect-ratio
SEO	viewport, document-title, meta-description, http-status-code, link-text, is-crawlable, robots-txt, image-alt, hreflang, canonical, font-size, plugins, tap-targets, structured-data
PWA	load-fast-enough-for-pwa, works-offline, offline-start-url, is-on-https ，service-worker, installable-manifest, redirects-http, splash-screen, themed-omnibox, content-width, viewport,without-javascript, apple-touch-icon, pwa-cross-browser, pwa-page-transitions, pwa-each-page-has-url

[9] https://developers.google.com/web/tools/lighthouse

[10] Google Lighthouse, https://developers.google.com/web/tools/lighthouse

4. 評価実験

本項では，提案手法に対する評価実験の手法と実験結果について述べる．Google Lighthouseより取得できる特徴量を用い，関連手法[2][3][4]同様にReliable, Unreliableの2クラス分類のデータセットに対する分類精度を確認する．加えて被験者実験をおこない現行のWebページに対する適応性も確認する．また，10クラス分類での分類精度を確認することで提案手法の多クラス分類への応用可能性を探る．

4.1 Google Lighthouseの各特徴量の取得

Google Lighthouseより取得できる特徴量の取得にはGoogleの提供するGoogle LighthouseのNode CLI版を用いる．Google Lighthouseは本来Webページ作成者用に開発されたWebページの質を向上させるための自動検査ツールであり，Webサイトを5つの観点からそれぞれ0-100点でスコア評価するものである．CLI版においては点数の根拠となる表3に示した特徴量に対する評価を出力することができる．本研究では実験対象URL全てに対してGoogle Lighthouseの検査を走らせ，表3に示す特徴量を取得した．

4.2 2クラス分類

4.2.1 使用するデータセット

大谷ら[4]と同様にFake News Corpusを用いる．Fake News Corpusは10クラスのラベルから成るWebページの評価結果のデータセットである．表4に示すように10クラスのうちCredible，Unreliable，Fake，Conspiracyを使用し，CredibleをReliableとし正例に，Unreliable，Fake，ConspiracyをまとめてUnreliableとし負例として学習した．また，偏りを防ぐため1つのドメインから判定するURLは最大20個に限定しランダムに取得し，アクセスのできないURLは除外している．

表4. 2クラス分類におけるデータセット

ラベル		*URL数*	*domain数*
本実験におけるタグ	*Fake news corpus中のタグ*		
Reliable	Credible	1,533	77
Unreliable	Fake Unreliable Conspiracy	1,683	112

4.2.2 特徴量選択

表3に示されたGoogle Lighthouseより取得できる特徴量のうち有効であるものをRFE(Recursive Feature Elimination)[12]と呼ばれる選択手法によって選択する．RFEは機械学習での予測の際にそれぞれの特徴量の重要度が導出できるアルゴリズムを採用している場合に使用でき，指定した数になるまで特徴量を重要度が低い順に削っていく特徴量選択のための手法である．本論文ではRFEで使用する特徴量数の指定を1つずつ減らしていき，データセットに対して最も高いF値でありかつ使用特徴量数が最も少ないとき選択された特徴量群を選択する．選択された特徴量を表5，特徴量の重要度を図2に示す．

表5. 選択された特徴量

指標	特徴量	説明
Performance	uses-webp-images	画像フォーマットにwebpを使用しているか
Performance	uses-long-cache-ttl	効率的なキャッシュポリシーを使用しているか
Performance	first-contentful-paint	ナビゲーションからブラウザがDOMのコンテンツの最初のビットをレンダリングするまでの時間
Performance	speed-index	ページのコンテンツが表示されるまでの時間
Performance	uses-rel-preconnect	<link rel = preconnect>でフェッチリクエストを優先しているか
Performance	total-blocking-time	ユーザー入力に対するページの応答がブロックされる時間
Performance	unused-css-rules	使用されていないcssがないか
PWA	installable-manifest	pwaのマニフェストを設定しているか
PWA	redirects-http	HTTP トラフィックを HTTPS にリダイレクトしているか
Performance	uses-responsive-images	レスポンシブイメージを使用しているか
Performance	first-meaningful-paint	ウェブフォントが読み込まれる前のペイントの表示までの時間
Performance	font-display	ロード中のWebフォントの動作
Performance	offscreen-images	オフスクリーン画像への対処がされているか
Performance	bootup-time	javascriptのbootup時間
SEO	canonical	ドキュメントに有効なrel = canonicalがあるか
Best Practices	is-on-https	HTTPSで配信されているか
Performance	uses-optimized-images	最適化された画像であるか
Performance	dom-size	適切なdom sizeであるか
Accessibility	frame-title	<frame> or <iframe> elementsがタイトルを持っているか
Performance	max-potential-fid	サイトの最大入力遅延
Performance	redirects	リダイレクトの使用
Accessibility	tabindex	タブインデックスを使っているか
Accessibility	aria-valid-attr-value	attrに有効な値が入っているか
Accessibility	html-lang-valid	言語情報が入っているか
Performance	total-byte-weight	通信されるバイト数

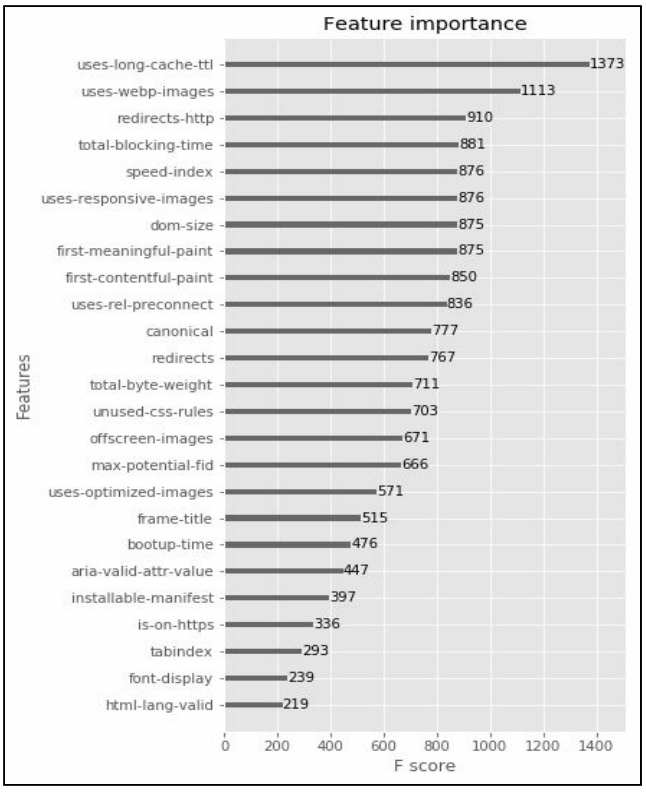

図2 選択された特徴量の重要度

4.2.3 XGBoostによる2クラス分類

4.2.2にて選択した特徴量を用い，信頼性評価データセットの分類を行う．比較手法としてOlteanuら[2]，Wawarら[3]，大谷ら[4]の提案特徴量であるContent, Linguistic, AD/Non-AD特徴量を実装しそれぞれ組み合わせて実験をおこなった．使用した特徴量とF値を表6に示す．

表6．2クラス分類の精度

	特徴量				F値
	Content [2]	*Linguistic*	*AD/Non-AD [4]*	*提案特徴量*	
Olteanu らの手法[2]	✓				0.727
Wawarらの手法 [3]	✓	✓			0.797
大谷らの手法[4]	✓	✓	✓		0.811
提案手法				✓	**0.898**
提案手法+[2]	✓			✓	0.874
提案手法+[3]	✓	✓		✓	0.878
提案手法+[4]			✓	✓	0.882

4.2.4 独自に取得したURL群の分類

提案特徴量による分類がデータセット中のWebページに対してだけでなく，現行の任意のWebページに対して有効であることを示すために独自に取得したURL群の分類をおこなった．まず，著者の研究室にて独自に収集した8,230件のURL群について提案手法で自動分類を行った．分類結果を表7に示す．次に被験者に収集したURLのうちドメインが異なるように無作為に抽出した200件にアクセスしてもらい，それぞれのWebページの信頼性についてReliable(信頼できるデータ）とUnreliable（信頼できると言い切れないデータ）の2値で評価してもらった．被験者は早稲田大学内から募った12人であり，1つのURLにつき3人の被験者による分類をおこない，最終分類結果として多数決を採用した．実験の結果，自動分類の分類精度は0.650となった．また，提案手法による自動分類と被験者の分類結果の混同行列を表8に示す．

表7．独自に取得したURL8230件の提案特徴量による自動分類結果

ラベル	*自動分類*
Reliable	2,144
Unreliable	6,086

表8．提案手法による自動分類と被験者実験の分類結果の混同行列

		被験者実験	
		Reliable	*Unreliable*
自動分類	*Reliable*	16	46
	Unreliable	24	114

4.2.5 2クラス分類結果に対する考察

最初に，4.2.2における特徴量選択において選択された25特徴量について考察する．選択された25特徴量のうち，16件はWebページのパフォーマンスに関連する特徴量であった．また，パフォーマンスに関連する特徴量の内容として画像やWebページの表示に関するもの，待ち時間に関するものが16件中12件を占めており，Webページ表示までの時間，表示そのものが信頼性評価における重要な特徴量となると考えられる．

次に，XGBoostによる2クラス分類においては，提案特徴量を用いることで，表6に示す通り，従来手法[2][3][4]と比較して0.087～0.171のF値向上が確認できた．さらに，被験者実験によるWebページの評価においても，被験者による評価と提案特徴量を用いた自動評価において精度が0.650となり，提案特徴量を用いた自動評価が人間による手動評価と近い評価ができていることが示唆される．また，表8の混同行列より，UncredibleなWebページの検知において一定の効果があることが示されており，提案手法を検索エンジンやブラウザなどに組み込むことによって応用的な使い方ができると考えられる．

また，図3, 図4に誤判定されたWebページの例を示す．図3の誤ってReliableと判定された例に関しては，サイト自体の情報量が少なくシンプルであるためWebページ全体のパフォーマンスが良く，提案手法においてReliableと判定されたと考えられる．図4の誤ってUnreliableと判定された例については，教育機関のホームページであることから信頼性が高いものであると考えられるが，Webページの更新がされていないことや，読み込み速度などのパフォーマンスが低いことが原因となって誤判定されたものと考えられる．

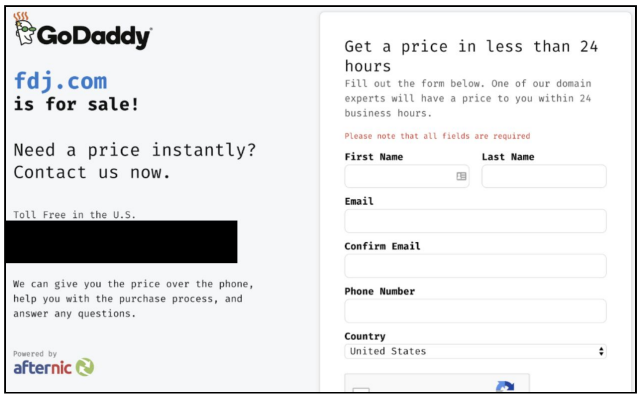

図3 誤ってRELIABLEと判定された例

図4 誤ってUNRELIABLEと判定された例

4.3 10クラス分類

提案手法がWebページの信頼性評価においてReliableとUnreliableの2値だけでなく，データセット内にある各種ラベルにも適用することができるのか確認するため，10クラス分類をおこなった．

4.3.1 使用するデータセット

2クラス分類と同様にFake News Corpusを用いる．10クラス分類ではFake News Corpusの全てのラベルを使用する．また，2クラス分類同様1つのドメインから判定するURLは最大20個に限定しランダムに取得し，アクセスのできないURLは除外している．

表9．10クラス分類におけるデータセット

ラベル	*URL数*	*domain数*
Bias	2,588	129
Conspiracy	2,188	109
Reliable	1,980	99
Fake	1,880	94
Satire	1,720	86
Political	1,522	76
Unreliable	832	42
Clickbait	600	30
Janksci	570	29
Hate	440	22

4.3.2 XGBoostによる10クラス分類

10クラス分類結果のprecision，recall，F値を表10に示す．マクロ平均ではラベルごとにprecision，recall，F値を個別に計算し、平均を取っている．加重平均では各ラベルのドメイン数で加重平均を計算している．F値はマクロ平均で0.340、加重平均で0.468である．

表10. 10クラス分類のPRECISION/RECALL/F

ラベル	*Precision*	*Recall*	*F*
Bias	0.506	0.284	0.364
Conspiracy	0.494	0.464	0.479
Reliable	0.696	0.784	0.737
Fake	0.326	0.376	0.346
Satire	0.224	0.242	0.233
Political	0.420	0.372	0.395
Unreliable	0.196	0.476	0.278
Clickbait	0.232	0.602	0.335
Janksci	0.160	0.510	0.244
Hate	0.096	0.250	0.139
Macro Avg.	0.340	0.460	0.391
Weighted Avg.	0.468	0.412	0.438

4.3.3 独自に取得したURL群の分類

4.2.4項同様に独自に取得したURL群の分類をおこなった．まず，著者の研究室にて独自に収集した8,230件のURL群について提案手法で10クラスに自動分類した．分類結果を表11に示す．次に被験者に収集したURLのうちドメインが異なるように無作為に抽出した200件にアクセスしてもらい，それぞれのWebページの信頼性について10クラスに分類してもらった．表12に分類に際しての指示を列挙する．1つのURLにつき3人の被験者による分類をおこない，最終分類結果として多数決を採用した．多数決の結果分類ラベルが一意に定まらなかった物に関してはUnknownラベルとして除外した．また，Unknownラベルとして判定されたものは200件中52件であった．

表11. 独自に取得したURL8230件の
提案特徴量による自動分類結果

ラベル	*自動分類*	*ラベル*	*自動分類*
Bias	1,283	Political	1,404
Conspiracy	1,968	Unreliable	102
Reliable	563	Clickbait	86
Fake	502	Janksci	135
Satire	2,177	Hate	10

表12. 被験者実験に際しての分類指示

ラベル	*指示*
Bias	偏った観点で物事を見ており、プロパガンダ、文脈から切り離された情報、事実として歪められた意見に依存している
Conspiracy	陰謀論。ある出来事について、広く人々に事実として認められている公の情報やその解説とは別に、特定の組織や人物にとっての利益に繋がった策謀や事実の存在を指摘している
Reliable	ジャーナリズムの伝統的かつ倫理的な慣行と一貫した方法でニュースや情報を流している
Fake	フェイク。情報を完全に作成したり、虚偽のコンテンツを流布したり、実際のニュースを著しく歪めたりしている
Satire	風刺。ユーモア、皮肉、誇張、および虚偽の情報を使用して現在のイベントにコメントしている
Political	特定の視点または政治的オリエンテーションをサポートする一般的に検証可能な情報を提供する
Unreliable	信頼できるかもしれないが、その内容にさらなる検証が必要
Clickbait	一般的に信頼できるコンテンツを提供するが、誇張された、誤解を招く、または疑わしい見出し、ソーシャルメディアの説明、画像を使用するソース。
Janksci	擬似科学、形而上学、自然主義的、およびその他の科学的に疑わしい主張を促進する
Hate	ヘイト。人種差別、女性差別、同性愛嫌悪、およびその他の形態の差別を積極的に促進する

4.3.4 10クラス分類結果に対する考察

マクロ平均で0.340，加重平均で0.468のF値を達成した．また，被験者による評価と提案特徴量を用いた自動評価において精度が0.095となり，提案手法が多クラスの信頼性ラベルの自動付与に有効であるとは言えない結果となった．

5. おわりに

本論文では，Webページの信頼性を自動評価するためにWebページのユーザビリティとパフォーマンスに注目した特徴量を提案し，提案特徴量が信頼性評価に有効であることを確認した．また，提案特徴量を用いることにより，APIや自然言語に関連する特徴量を使わずにWebページの信頼性評価をおこなうことができ，評価システムの汎用性や継続性が向上することを確認した．新規Webページを対象にした被験者実験でも，2値の分類において提案特徴量が有効であり，人間の評価と同等の評価をすることが確認された．しかし，10クラス分類においては，分類精度が著しく低く，Webページのユーザビリティとパフォーマンスに注目した提案手法のみでは多クラスの信頼性ラベルの自動付与に有効であるとは言えないと考えられる．今後の研究を進めるにあたって，より広範囲のWebページの取得や均一なラベルの付与がされているデータセットを用いること，評価システムの汎用性や継続性を損なわない形での他情報の使用などを検討し，Webページがより汎用性，継続性を持って評価されるシステムの開発・提案を行う予定である．

謝辞

本研究を進めるにあたり，研究の相談や助言をしていただいた山名研究室の方々に厚く御礼を申し上げます．本研究は、科学研究費助成事業17KT0085によるものである．

参 考 文 献

[1] M. Kakol, R. Nielek, and A. Wierzbicki, “Understanding and predicting web content credibility using the content credibility corpus,” Information Processing and Management, vol. 53, no. 5, pp. 1043–1061, 2017.
[2] A. Olteanu, S. Peshterliev, X. Liu, and K. Aberer, “Web credibility: features exploration and credibility prediction,” LNCS, vol. 7814, pp. 557–568, 2013.
[3] A. Wawer, R. Nielek, and A. Wierzbicki, “Predicting web page credibility using linguistic features,” Proc. of WWW '14, pp. 1135–1140, 2014.
[4] K. Ootani and H. Yamana, “External content-dependent features for web credibility evaluation,” Proc. of IEEE BigData 2018, pp. 5414–5416, 2018.
[5] Guyon, I., Weston, J., Barnhill, S., & Vapnik, V., “Gene selection for cancer classification using support vector machines”, Mach. Learn., vol. 46, Issue 1–3, pp. 389–422, 2002.
[6] Kenta Yamada, Hayato Yamana, “Effectiveness of Usability & Performance Features for Web Credibility Evaluation”, IEEE BigData 2019, 2019.
[7] J. Schwarz and M. Morris, “Augmenting web pages and search results to support credibility assessment,” Proc. of ACM CHI '11, pp. 1245-1254, 2011.
[8] P. Geurts, D. Ernst, and L. Wehenkel, “Extremely randomized trees,” Machine Learning, vol. 63, no. 63, issue 1, pp. 3–42, 2006.
[9] C. F. Hsu, E. Khabiri, and J. Caverlee, “Ranking comments on the social web,” Proc. of CSE '09, vol. 4, pp. 90–97, 2009.
[10] G. H. McLaughlin, “SMOG grading: A new readability formula,” Journal of Reading, vol. 12, no. 8, pp. 639–646, 1969.
[11] P. Stone, D. Dunphy, M. Smith, and D. Ogilvie “The General Inquirer: A Computer Approach to Content Analysis,” The MIT Press, 1966.