

ニュース記事の品質が広告消費行動 に与える影響の調査

飯塚洸二郎¹ 加藤誠² 関喜史³

本論文では、推薦するニュース記事の品質がユーザの広告消費活動に与える影響を調査する。まず我々は、ニュースアプリケーションの実サービスログを分析することによって、記事の品質と広告の消費傾向には相関があることを確認した。次に、ニュース記事の品質と広告の消費に関する因果関係を調査するために、A/Bテストを用いた大規模なオンライン実験を行った。結果として、高品質な記事を掲出したユーザ群の広告の消費指標は全体的に上昇し、特にコンバージョンに関する指標が顕著に増加することを確認した。最後に、これらの知見を活かし、広告の消費情報を活用することが、コンテンツ品質の分類性能の向上につながることを示した。

1 はじめに

近年、ニュース推薦システムは、ユーザの情報ニーズを満たす重要な役割を果たしている。ウェブ上には数え切れないほどのニュースが毎日生み出されており、そうした中で推薦システムは、ユーザの興味にあった記事や、有用性の高い記事を提供している [1]。このようなニュースを提供するサービスの多くは、推薦する記事コンテンツと広告（製品、サービス、情報などを広告するアイテム）を混合したランキングをユーザに掲出する。サービス上で、ユーザに継続的にサービス価値を提供し続けるためには、ユーザが満足するコンテンツを提供すると同時に、収益をあげていく必要がある。このユーザ満足度と収益のバランスを取るためには、ユーザ行動において、コンテンツと広告が相互に与える影響を解明することが重要な研究課題である。

情報検索や情報検索の分野において、クリック情報に基づくユーザ行動分析は古くから行われてきた [2] [3] [4]が、コンテンツと広告に相互作用をもたらす要因については未解明な部分がある。一般に、ユーザが起こすクリック行動は、ポジション [5] [6]、信頼性 [7]、品質 [7]、プレゼンテーション [8]、配信手段 [9]などのバイアスに作用されることが知られている。近年では、これらのバイアスを考慮し、ユーザの主観的な知覚に基づく行動分析が行われている [10]。また、推薦コンテンツに対するユーザ満足度と、広告に対する収益を直接的に同時に最適化する試みも始まっている [11] [12]。しかしながら、これらの研究は推薦するコンテンツと広告を同時に最適化しているものの、コンテンツ自体が広告の消費に与える影響は考慮していなかった。そのため、コンテ

ンツが広告の消費に影響を与える場合、全体的なパフォーマンスの最適化が行えない懸念が残る。

そこで本研究では、ニュース記事の品質がユーザの広告の消費行動に与える影響を調査する。本研究において、記事の品質が広告消費に影響を与えるとは、記事の品質に応じて広告の消費が変化することを指す。例えば、記事の品質が広告の消費に影響を与える例は、図 1 のような場合である。この図は、ニュースサービスにおけるニュースと広告が混ざったランキングにおいて、記事の品質に応じて広告の消費が変化する例を表している。図の左側では、高品質の記事の下に広告が位置して広告消費が起きる例を表し、右の図は低品質の記事の下に広告が位置して広告消費が起これない例を表している。この例は、ユーザーはランキングのアイテムを上から下に向かって消費していくため、直前に接した記事の品質に応じて広告への行動に変化があるという仮説に基づいている。

本研究では、記事の品質と広告の消費に関するオフラインでの相関分析からはじめ、A/Bテストによるオンライン実験によって、記事の品質と広告消費の因果関係を検証した。オフライン調査に用いたデータには、大規模なオンラインニュースサービスにおけるニュース記事と、記事と広告に対するユーザフィードバックログを用いた。ニュースの記事品質は、専門家と単語のマッチングによる機械的な判別を交えて高品質と低品質の2つに事前に分類した。結果としては、記事の品質は広告の消費、特にコンバージョン率に対して相関が見られた。A/Bテストでは、専門家が判定した高品質の記事を優先的にユーザに提示する群と、既存のシステムの出力をユーザに提示する群の2つに分け、広告消費の変化を観測した。結果として、高品質の記事をユーザに提示することで、コンバージョン率をはじめとする広告指標が改善することを確認した。これは、高品質なコンテンツを提供することがサービスの信頼度を高め、結果としてコンバージョンという深いユーザアクションにつながったのではないかと考えられる。これらの知見を活かし、広告のコンバージョンなどのフィードバックは、記事品質の分類精度に寄与するのではないかと考えた。実際に、ツリーベースの予測モデルを用いて記事品質の分類精度を計測したところ、広告の特徴量を加えることで、分類精度が向上することを確認した。

本研究の貢献は以下のようにまとめられる。

- ニュース記事の品質が広告の消費に与える影響を調査する新しい設定の問題に取り組んだ。
- オフラインでの分析と、A/Bテストを用いたオンライン実験によって、ニュース記事の品質は広告の消費に影響を及ぼすことを確認した。特に、高品質の記事を優先的にユーザに提示することは、コンバージョン率をはじめとする広告指標の改善につながった。
- 広告の消費情報を活用することで、ニュースの品質の予測精度が向上することを示した。

本論文の構成は下記のとおりである。第2章では、関連研究と本研究での差分を記述する。第3章では、分析に用いるニュース記事の分類の定義とニュースサービスにおける行動形式について

¹ 非会員 株式会社Gunosy/筑波大学大学院人間総合科学学術院

kojiro.iizuka@gunosy.com

² 正会員 筑波大学図書館情報メディア系/JST さきがけ

mpkato@acm.org

³ 非会員 株式会社Gunosy

yoshifumi.seki@gunosy.com

	ランキング	品質	広告消費
記事	記事A	高	
	記事B	高	
	広告		有り

	ランキング	品質	広告消費
記事	記事C	低	
	記事D	低	
	広告		無し

図1 記事の品質が広告の消費に与える影響例

述べる。第4章では、サービスログを用いた分析と結果について報告する。第5章では、A/Bテストを用いたオンライン実験の設定と結果について報告する。第6章では、広告のフィードバックを用いた記事品質の分類について述べる。第7章では、結論と今後の展望を述べる。

2 関連研究

この章では、推薦コンテンツと広告の最適化、ユーザ行動分析、記事品質の予測のトピックに関して、研究の方向性をレビューする。

2.1 推薦コンテンツと広告の最適化

ソーシャルネットワークサービスやニューサービスを始めとするウェブサービスは、ユーザを惹き付けるコンテンツと、企業の収益源となる広告の両方をユーザに掲出している。そこで、推薦コンテンツによるユーザ満足度と広告による収益の両方を最適化する問題が研究の対象になっている。Zhaoらは、処理の第一段階で長期的なユーザ体験を考慮した推薦ランキングを生成し、第二段階で利益とユーザ体験のバランスを取りつつ広告を推薦ランキングに挿入する手法を提案した [11]。Wangらは、広告を挿入する位置や数を固定ではなく、動的に挿入する強化学習ベースの手法を提案した [13]。Yanらは、ニュースフィードにおいて、記事と記事の間に挿入する広告の位置を最適化する手法の提案を行った [12]。

これらの研究は、コンテンツと広告を独立のものとして扱い、推薦コンテンツと広告の位置や数を最適化している。これらの研究の前提として、コンテンツと広告の相互作用が存在するのかわ、あるとしたらどの程度存在するのかわを明らかにすることは、重要な研究トピックであると考えられる。

2.2 ユーザ行動分析

情報推薦や情報検索の分野において、クリック情報に基づく行動分析は古くから行われてきた [2] [3] [4]。ユーザが起こすクリック行動は、ポジション [5] [6]、信頼性 [7]、品質 [7]、プレゼンテーション [8]、配信手段 [9]などのバイアスに左右される。近年では、これらのバイアスを考慮し、ユーザの主観的な知覚や、コンテンツの品質自体を考慮した行動分析が行われている。

ユーザの主観的な知覚として、コンテンツの有用性に着目した研究がある。Belkinらは、ユーザが検索結果に対して感じる有用性は、ユーザのコンテキストに依存すると論じた [14]。Maoらは、検索結果の有用性がユーザの満足度に相関することを

示した [15]。Jiangらは、適合度、有用性、真新しさ、理解のしやすさ、信頼性、労力という多次元の尺度で検索結果を評価した [16]。結果として、多次元の尺度を用いる評価は、有用性のみを用いて評価する場合よりも、より正確にユーザ体験を評価できることを示した。また、ユーザがコンテンツをクリックした後に評価する場合と、ユーザがセッションを終えた後に評価する場合では、ユーザのコンテキストの違いに起因して評価結果が異なることを示した [16]。さらに、ニュース記事の品質がユーザ体験に影響を与えるという報告もある [10] [17]。Luらは、タイトルでユーザの目を引くような品質の低い記事は、クリックされやすいことを示し、低品質記事を閲覧することが、その後のユーザの記事閲覧行動に悪影響を及ぼすことを示した [17]。品質の低い記事がユーザ行動に対して与える影響についても調査を行っている [10]。

これらの研究のスコープは、ユーザが知覚したコンテンツの品質がその後のコンテンツの消費に対して与える影響の調査であった。対して本研究は、ユーザに対して掲出したコンテンツの品質が、異なるドメインである広告の消費活動に対して与える影響を分析することに注力する。

2.3 記事品質の予測

ソーシャルネットワークサービスの普及により情報の拡散力が高まったことで、拡散される情報の品質がより多くの人に影響を及ぼすようになった [18]。そのため、コンテンツの品質を予測する研究が盛んに行われている。

低品質な記事を予測する問題として、クリックペイトやフェイクニュースを予測する研究がある。Chakrabortyらは、クリックペイトの予測を活用してブラウザ上でクリックペイト記事に対する警告を発するシステムを開発した [19]。Indurthiらは、クリックペイトを二値分類するのではなく、クリックペイトの強さを回帰問題として扱った [20]。Biyaniらは、クリックペイトに関する定義を行い、ウェブコンテンツにおけるURLや、タイトルと本文の類似度などのコンテンツ情報がクリックペイトの予測に寄与することを示した [21]。Potthastらは、ソーシャルメディアにおけるクリックペイトのデータセットの構築を行い、クリックペイトの予測を255もの特徴量を用いた予測モデルを構築した [22]。Shuらは、フェイクニュースに関する包括的なレビューを行い、ソーシャルネットワークサービスにおけるユーザエンゲージメント情報を活用することが、フェイクニュースの予測に寄与することを示した [23]。Zhouらは、フェイクニュースに関する研究課



図2 典型的なランキング形式のニュースシステムであるGunosyのサービスイメージ。

題をまとめ、自然言語処理や情報検索など様々な分野におけるつながりを含めた研究の方向性を示した [24].

高品質な記事を予測する問題として、Luらは、記事に対する暗黙的フィードバックから記事の品質を予測できるかを論じた [17]. Jiangらは、検索結果である文書に対する有用性、真新しさ、理解のしやすさ、信頼性を暗黙的フィードバックから予測する手法を示した [16].

これらの研究は、記事コンテンツの特徴や、ユーザの記事に対する暗黙的フィードバックを活用し、ソーシャルネットワークのエンゲージメントを付加情報として活用し、記事品質を予測する研究であった。対して、本研究では比較的安価に手に入れることができる、ニュースフィードにおける広告のフィードバックを付加情報として活用し、記事品質の予測性能を向上させることを目指す。

3 準備

この章では、本論文で扱う記事品質とニュース閲覧の設定について述べる。

3.1 品質の定義

既存研究と同様にニュースの記事品質については、下記の4つの項目を考慮する [17].

信憑性 信憑性が高いとは、そのニュースが本物であるか、信頼性が高いことを意味する。逆に、内容が想像上のもや誇張されたものであれば、ニュースの信憑性は低い。

価値 ニュースの一部に下品、暴力的、残虐的、ポルノ的な要素を含む場合、価値は低い。

表現 文章が客観的かつ正確であり、情報が豊富だが冗長ではない場合の表現は、表現の品質が高い。

見出し ニュースの見出しが不完全または偽物、表現が大げさ、下品または、記事の本文と矛盾している場合、タイトルの品質が低い。

これらの項目を考慮し、本研究では、高品質、低品質に記事を分類する。

3.2 品質の判定

品質の判定は、専門家による判定と、単語の一致による機械的判定を用いる。まず低品質の記事については、事前に下品、暴力的、残虐的な単語をリストアップし、タイトルと本文に1つでも一致する単語が含まれているニュース記事を低品質とした。高品質の記事については、文章の客観性や正確性、適切な情報量を判定する必要があるため、専門家による判定を行った。本研究では、これらの高品質、低品質に分類された記事を扱う。

3.3 ニュース閲覧の設定

Googleニュース、Yahoo!ニュースやGunosyなどのニュースフィードを利用してウェブの情報を取得するユーザは、特にスマートフォン端末を利用するケースで顕著に増えている。ユーザはニュースフィードサービスに訪問した際に、ユーザの興味に基づくニュース記事やニュース動画などのコンテンツを推薦されると同時に、コンテンツの間にネイティブ広告と呼ばれる広告を推薦される。典型的なインターフェースは、図2に示されているように、タイトルと画像付きのニュース記事がランキング形式で縦に整列されている。

ニュースフィードサービスは、ユーザがランキング形式のフィードをスクロールする度に、新たなニュース記事や広告を一定の量まで表示し続ける。ユーザがランキングにおける記事のタイトルや画像をクリックすると、システムはクリックされたニュース記事の本文や画像の詳細コンテンツを表示する。コンテンツの閲覧は任意のタイミングで中断することができ、中断した場合は、元のランキングにおけるクリックした位置に戻る。このように、ユーザはアプリケーションに訪問したセッションにおいて複数の記事を閲覧することができる。また、これらの一連の動作において、ユーザは任意のタイミングでシステムの利用を止めることができる。

4 ログベースの分析

この章では、分析に利用するデータセットと指標について記述する。また、記事の品質と消費行動の関係性、記事の品質と広告消費行動の関係性について調査する。

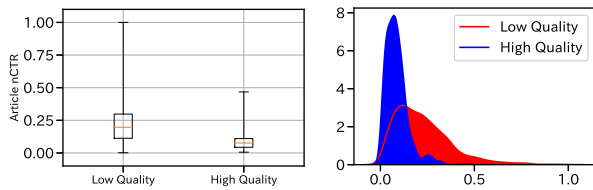


図3 記事品質とArticleCTR

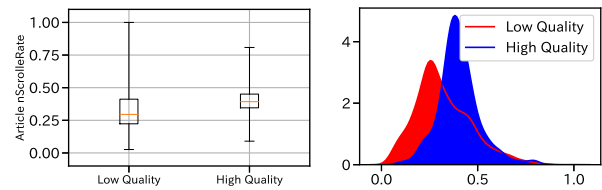


図4 記事品質と記事スクロール率

4.1 ユーザフィードバックの取得

我々は日本のニュースサービスであるGunosyにおいて、30日分のユーザフィードバックを収集した。このユーザフィードバックは、システムがユーザに記事を表示したログや、ユーザが記事をクリックしたログに加えて、広告の表示とクリック、さらには広告のコンバージョンデータが含まれる。ここで、コンバージョンとは広告をクリックした後に、広告に紹介されている他のアプリケーションをダウンロードする行為や、広告商品の詳細情報を請求する行為を指す。

記事はランキング形式でユーザに表示されており、ランキングにおける順位ごとにデータは集計可能である。この章の分析においては、ランキングの上位3件に表示されるニュース記事と、上位4件目に表示された広告を対象とする。ビジネスの観点から詳細な数は報告することができないが、収集したログのユーザのセッション数は合計1,000万を超えている。

4.2 インタラクション指標

記事と広告の消費に関する指標には、クリックをベースにした指標とコンバージョンの指標を用いる。

Article Click-Through Rate (ArticleCTR) ユーザが記事をクリックした数を、記事が表示された回数で割った値。

ArticleScrollRate ユーザが記事を読んだ長さを記事の全体の長さで割った値。

ArticleSessionLength ユーザが記事に滞在した時間の長さ。

Ad Click-Through Rate (AdCTR) ユーザが広告をクリックした数を、広告が表示された回数で割った値。

Ad Conversion Rate (AdCVR) 広告の詳細ページにおいて、コンバージョンした回数をクリックした回数で割った値。

なお、各指標をそのまま用いることは、ビジネス上のリスクがあるため各指標の正規化を行った。例えば、Article CTRについては、全ての記事のCTRの中で最も大きい値で各記事のCTRを割った値であるArticle normalized-CTR (Article nCTR)を本論文では掲載する。また同様にAd CVRもAd nCVRを掲載する。

4.3 記事品質とArticleCTRの関係

図3に記事品質とArticleCTRの関係を示す。左の図は、記事品質ごとに分けたArticle nCTRの箱ひげ図であり、右の図は記事品質ごとに分けたArticle nCTRのカーネル密度推定を用いた確率密度関数である。

記事の品質が低い場合のクリック確率の中央値は、高品質の記事よりも高く、また分布の広がりも大きいことが見て取れる。これは、既存研究の品質に応じたクリック率の分析 [17]と同様に、

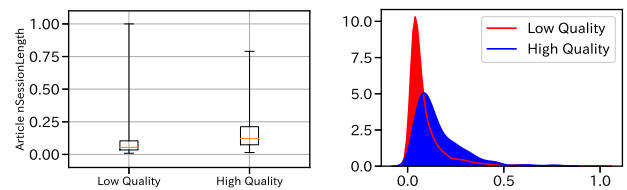


図5 記事品質と記事滞在時間

低品質な記事はタイトルでユーザの注目を惹く傾向があることに起因してクリック率が高くなっていると考えられる。

4.4 記事品質と記事スクロール率の関係

図4に記事品質と記事スクロール率の関係を示す。左の図は、記事品質ごとに分けたスクロール率の箱ひげ図であり、右の図は記事品質ごとに分けたスクロール率のカーネル密度推定を用いた確率密度関数である。

全体として高品質な記事は、低品質の記事に比べて、スクロール率が高いことが見て取れる。これは、高品質な記事は、タイトルを見るだけでニュースの概要が把握できた上で、詳細な情報を確認しに来たユーザが多いことに起因すると考えられる。一方で、低品質な記事は、ユーザーはタイトルに惹かれて記事をクリックし、記事の中身が期待に反したコンテンツであると分かった時点で記事詳細面から離脱していると考えられる。

4.5 記事品質と記事滞在時間の関係

図5に記事品質と記事滞在時間の関係を示す。左の図は、記事品質ごとに分けた記事滞在時間の箱ひげ図であり、右の図は記事品質ごとに分けた記事滞在時間のカーネル密度推定を用いた確率密度関数である。

全体として高品質な記事は、低品質の記事に比べて、滞在時間が長いことが見て取れる。これは、高品質な記事は、低品質な記事に比べて文章量が多いことに起因すると考えられる。一方で、低品質な記事は、記事の文章量が少ない、またはスクロール率の結果が示すとおり、最後まで閲覧されることが少ないと考えられる。

以上の、記事の品質が記事の閲覧行動に与える影響は、既存研究でも一部報告されている内容と同様であった [10] [17]。そのため、我々が本論文で利用しているニュースアプリケーションのデータセットは、他のニュースアプリケーションにおけるデータセットと同様の特性を持っていると考えられる。

4.5.1 記事品質と広告CTR、広告CVRの関係

図6に記事品質と広告CTRと広告CVRの関係を示す。これらの結果は、広告が、低品質または高品質のいずれの記事と一緒に表示されているかを用いて算出した結果である。左の図は、記事品

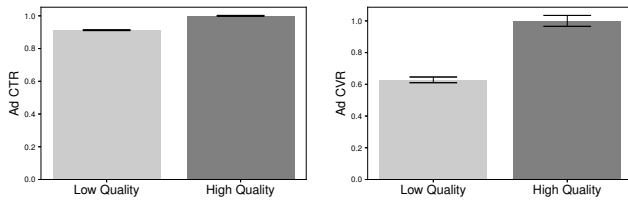


図6 記事品質と広告CTR, 広告CVR

質ごとに分けた広告CTRであり、右の図は記事品質ごとに分けた広告CVRである。

高品質な記事が、広告と共にユーザに掲出されている場合、低品質な記事が掲出されている場合と比べ、広告のCTRは高いことが分かる。また同様に、高品質の記事の場合、低品質な場合に比べて、広告のCVRが顕著に高いことが分かる。

これは、低品質な記事は、高品質記事に比べてユーザの視線を集めることから、広告のクリックも奪っていることが広告CTRの低下につながっていると考えられる。一方で、広告CVRの結果は、広告をクリックした上で実際に広告商品に対するコンバージョンが増えていることを意味する。これは、高品質な記事をユーザに掲出することが、ユーザの信頼性を高め、結果として広告商品の購入といった深いユーザ行動につながっているのではないかと考えられる。しかしながら、これらのログベースの分析には、様々な交絡要素が存在するため、記事の品質がユーザ行動に与える影響を正確に見積もることは難しい。そこで、次の章では、実際にランダム化比較試験を行うことで、記事の品質と広告の消費についての因果関係を確かめる。

5 オンライン実験

この章では、A/Bテストを用いたオンライン実験について報告する。前の章までの分析において、記事の品質は、記事や広告の消費と相関があることが確認できた。そこで、A/Bテストによるランダム化比較試験によって、記事の品質と広告の消費についての因果関係を確かめることが本章の目的である。特に、高品質の記事がユーザの広告消費行動に与える影響を調査する。

5.1 設定

ログベースの分析と同様に、オンラインニュースサービスであるGunosyにおいて実験を行った。A/Bテストにおけるユーザの分割としては、ユーザがサービスを開いた際に最も最初に表示される画面上に、高品質な記事を優先的に表示するユーザをtreatment群に、意図的に高品質な記事を表示しないユーザをcontrol群に分けた。ここで、高品質な記事は、オフライン分析と同様に、専門家による人手を介して判定された記事である。高品質な記事を優先的に表示するとは、ランキングの上位3つの記事を高品質記事に限定することを指す。意図的に高品質を表示しないとは、サービス上の既存アルゴリズムの結果をそのまま出力することを指す。高品質な記事を固定するタイミングは、朝8時から夜21時までの間において最低5時間に一度とした。

実験対象は、サービスにおけるユーザの一部を対象にした。A/Bテストの実施期間は、ログベースの分析を行った期間とは重

表1 広告指標に関するオンライン実験結果

	AdCTR	AdCVR	AdCV/DAU	Sales/DAU
iOS Uplift	+1.71%	+2.37%	+5.34%	+3.61%
Android Uplift	+3.17%	+4.27%	+7.63%	+5.59%

複のない1ヶ月間である。

5.2 結果

得られた結果を表1に示す。DAUとは、Daily Active Userの略で、サービスにおける日毎のアクティブユーザの数である。AdCV/DAU、Sales/DAUの指標に関しては、日毎に集計した値を全期間で平均した値を記している。集計結果は、iOSデバイスを使うユーザとAndroidデバイスを使うユーザに分けて記した。これは、過去の実験から、使用する端末のOSに応じて、実験結果が異なるケースがいくつか知られていたためである。

全体的な結果として、高品質な記事を優先するtreatmentユーザ群の広告数値が改善していることが見て取れる。特に、コンバージョンに関する指標が大きく上昇している。広告クリックに関する指標も上昇しており、最終的な売上に関する指標も向上していることが見て取れる。過去のA/Bテストの知見と同様に、iOSユーザとAndroidユーザによる結果の違いも見られ、Androidユーザのほうが広告指標の上昇が大きかった。iOSデバイスとAndroidデバイスを使うユーザー間ではデモグラフィック情報も異なり [25]、デモグラフィック情報の違いが広告効果の指標の違いを生んでいるのではないかと考えられる [26]。以上の結果から、記事の品質は、広告の消費に影響を及ぼすことが分かった。

6 広告のフィードバックを活用した記事品質の予測

この章では、記事の暗黙的フィードバックに加えて、広告の暗黙的フィードバックを活用することで、記事品質の予測性能が向上するかを検証する。

これまでの分析からニュースの品質はユーザの広告の消費に影響を及ぼすと結論づけた。本研究では、内部の専門家のアノテーションによって、ニュースの品質の一部をラベル付けしていた。この方法は信頼性は高いが、コストが高く、実際のアプリケーションに適用するのは難しい。このように、ニュースの品質をどのようにして判別するかは重要な研究課題である。

品質を付与するもう一つの方法は、品質を自動的に予測する方法である。従来の予測モデルは、ニュースのコンテンツの情報や、ニュースの暗黙的フィードバックを活用していた。本研究では、ニュース自体から得られる情報に加えて、さらに広告の暗黙的フィードバックを活用することで、予測精度を向上させることを目指す。

6.1 学習

学習には、下記の特徴量を用いた。

ArticleCTR 記事のClick Through Rate.

MediaName 記事の提供元であるメディア名。

表2 記事品質の予測精度

	Precision		Recall		F1-Score		Support
	w/o ad	with ad	w/o ad	with ad	w/o ad	with ad	
Low Quality	0.908	0.924	0.948	0.963	0.928	0.943	1482
High Quality	0.534	0.671	0.380	0.483	0.443	0.561	226
Accuracy					0.873	0.899	1708
Macro Average	0.721	0.797	0.664	0.723	0.686	0.752	1708
Weighted Average	0.859	0.890	0.873	0.899	0.864	0.892	1708

ScrollRate 記事のスクロール率。

SessionLength 記事の閲覧時間。

MediaCVR メディアごとに集計したCVR。

ArticleCVR 記事ごとに集計したCVR。

MediaCVRとArticleCVRが、広告フィードバックを用いた特徴量であり、それ以外が記事自体のフィードバック情報である。MediaCVRは、ある記事が表示されたときに、同一セッションで広告に対しておこされたコンバージョンをメディアごとに集計した値である。同様にArticleCVRは、コンバージョンの記事ごとに集計した値である。

学習のモデルには、ツリーベースの分類モデルであるLightGBM^{*1}を用いた。学習のモデルに利用したパラメータとしては、高品質か否かの二値分類を指定し、目的関数としてはbinary loglossを用いた。その他のパラメータはデフォルト値を用いた。学習データは、train:validation:test=3:1:1になるように分割し、学習と評価を行った。

6.2 結果

モデルの学習結果を表2に記す。表において、広告の素性を用いない場合をw/o adで表し、用いる場合はwith adで表している。

全体として、広告の素性を用いる場合の方が、用いない場合よりも分類性能が向上していることが見て取れる。特に、広告の素性を用いることは、高品質記事の分類に大きく寄与していることが分かる。高品質記事は、低品質記事よりもアノテーションの数が少なかったため、学習データは不均衡となっていた。そのため、ラベルの数で重みを付けたweighted averageよりもmacro averageにより顕著に学習結果の差が現れている。

この結果は、ユーザが接触した記事の品質と、広告消費量の相関分析の結果とも整合する。広告のコンバージョンに関するフィードバックは、ユーザが体感する品質や信頼性の一部を予測するために有用であると考えられる。

7 まとめと今後の課題

近年の多くのウェブサービスは、ユーザを惹きつけるサービスの主要なコンテンツと、収益のための広告を混ぜてをユーザに掲出する。そのため、ユーザ行動におけるコンテンツと広告の相互作用を解明することが重要な研究課題である。本研究では、特にニュースサービスにおける記事の品質が、ユーザの広告消費行動

に与える影響について調査した。

記事の品質の分類には、人手と機械的判別の両方を組み合わせて半自動的に高品質と低品質の分類を行った。コンテンツが広告の消費に与える影響の調査としては、まずオンラインニュースサービスにおけるログベースの相関分析を行った。結果として、記事の品質と、広告の消費傾向には相関があることが確認できた。次に、記事の品質と広告の消費に関する因果関係を調査するために、A/Bテストを用いた大規模なオンライン実験を行った。A/Bテストにおける分割は、高品質な記事を優先的に表示させるユーザ群と、従来の推薦アルゴリズムの出力をそのまま表示するユーザ群に分けて実験を行った。結果として、高品質な記事を掲出したユーザ群の広告の消費指標は全体的に上昇した。特にAndroidユーザにおいては広告のクリック率が3.17%上昇、コンバージョン率が4.27%上昇、最終的な売上が5.59%向上した。最後に、広告の消費情報を活用することで、記事品質の分類性能が向上することを確認した。特に、高品質記事の分類の性能が向上することによって、macro f1-scoreについては9.6%の向上があった。

今後の課題として、高品質記事を優先して表示したユーザ群のユーザ体験を定性的な面で分析する予定である。また、ユーザ体験と収益の双方を同時に最適化する際に、記事品質による影響を考慮したロジックを組み込むことで、実際にユーザ体験と収益が向上するかを確認することが今後の課題である。

謝辞

この研究は、株式会社Gunosyの、Media Machine Learningチーム、Data Reliability Engineering (DRE) チーム、Business Intelligence (BI)チーム、コンテンツ運用チームの協力のもと遂行されたものです。また、筑波大学の加藤研究室の皆様には、温かいご指導ご鞭撻を賜りました。この場を借りて深く御礼申し上げます。

参考文献

- [1] Jiahui Liu, Peter Dolan, and Elin Rønby Pedersen. Personalized news recommendation based on click behavior. In *Proceedings of the 15th international conference on Intelligent user interfaces*, pp. 31–40, 2010.
- [2] Georges E Dupret and Benjamin Piwowarski. A user browsing model to predict search engine click data from past observations. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 331–338, 2008.
- [3] Yifan Hu, Yehuda Koren, and Chris Volinsky. Collaborative filtering for implicit feedback datasets. In *2008 Eighth IEEE International*

*1 <https://github.com/microsoft/LightGBM>

- Conference on Data Mining*, pp. 263–272. Ieee, 2008.
- [4] Rong Pan, Yunhong Zhou, Bin Cao, Nathan N Liu, Rajan Lukose, Martin Scholz, and Qiang Yang. One-class collaborative filtering. In *2008 Eighth IEEE International Conference on Data Mining*, pp. 502–511. IEEE, 2008.
 - [5] Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, and Geri Gay. Accurately interpreting clickthrough data as implicit feedback. In *ACM SIGIR Forum*, Vol. 51, pp. 4–11. Acm New York, NY, USA, 2017.
 - [6] Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, Filip Radlinski, and Geri Gay. Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search. *ACM Transactions on Information Systems (TOIS)*, Vol. 25, No. 2, pp. 7–es, 2007.
 - [7] Yisong Yue, Rajan Patel, and Hein Roehrig. Beyond position bias: Examining result attractiveness as a source of presentation bias in clickthrough data. In *Proceedings of the 19th international conference on World wide web*, pp. 1011–1018, 2010.
 - [8] Chao Wang, Yiqun Liu, Min Zhang, Shaoping Ma, Meihong Zheng, Jing Qian, and Kuo Zhang. Incorporating vertical results into search click models. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pp. 503–512, 2013.
 - [9] Jimmy Lin, Salman Mohammed, Royal Sequiera, and Luchen Tan. Update delivery mechanisms for prospective information needs: An analysis of attention in mobile users. In *The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 785–794, 2018.
 - [10] Hongyu Lu, Min Zhang, Weizhi Ma, Ce Wang, Feng xia, Yiqun Liu, Leyu Lin, and Shaoping Ma. Effects of user negative experience in mobile news streaming. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR'19, p. 705–714, 2019.
 - [11] Xiangyu Zhao, Xudong Zheng, Xiwang Yang, Xiaobing Liu, and Jiliang Tang. Jointly learning to recommend and advertise. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '20, p. 3319–3327, 2020.
 - [12] Jinyun Yan, Zhiyuan Xu, Birjodh Tiwana, and Shaunak Chatterjee. Ads allocation in feed via constrained optimization. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '20, p. 3386–3394, 2020.
 - [13] Weixun Wang, Junqi Jin, Jianye Hao, Chunjie Chen, Chuan Yu, Weinan Zhang, Jun Wang, Xiaotian Hao, Yixi Wang, Han Li, Jian Xu, and Kun Gai. Learning adaptive display exposure for real-time advertising. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, CIKM '19, p. 2595–2603, 2019.
 - [14] Nicholas J Belkin, Michael Cole, and Jingjing Liu. A model for evaluation of interactive information retrieval. In *Proceedings of the SIGIR 2009 Workshop on the Future of IR Evaluation*, pp. 7–8, 2009.
 - [15] Jiaxin Mao, Yiqun Liu, Ke Zhou, Jian-Yun Nie, Jingtao Song, Min Zhang, Shaoping Ma, Jiashen Sun, and Hengliang Luo. When does relevance mean usefulness and user satisfaction in web search? In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pp. 463–472, 2016.
 - [16] Jiepu Jiang, Daqing He, and James Allan. Comparing in situ and multidimensional relevance judgments. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 405–414, 2017.
 - [17] Hongyu Lu, Min Zhang, Weizhi Ma, Yunqiu Shao, Yiqun Liu, and Shaoping Ma. Quality effects on user preferences and behaviors in mobile news streaming. In *The World Wide Web Conference*, WWW '19, p. 1187–1197, 2019.
 - [18] Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. *Journal of economic perspectives*, Vol. 31, No. 2, pp. 211–36, 2017.
 - [19] Abhijnan Chakraborty, Bhargavi Paranjape, Sourya Kakarla, and Niloy Ganguly. Stop clickbait: Detecting and preventing clickbaits in online news media. *ASONAM '16*, p. 9–16, 2016.
 - [20] Vijayasaradhi Indurthi, Bakhtiyar Syed, Manish Gupta, and Vasudeva Varma. Predicting clickbait strength in online social media. In *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 4835–4846, 2020.
 - [21] Prakhar Biyani, Kostas Tsioutsoulis, and John Blackmer. "8 amazing secrets for getting more clicks": Detecting clickbaits in news streams using article informality. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI'16, p. 94–100, 2016.
 - [22] Martin Potthast, Sebastian Köpsel, Benno Stein, and Matthias Hagen. Clickbait detection. In *European Conference on Information Retrieval*, pp. 810–817. Springer, 2016.
 - [23] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter*, Vol. 19, No. 1, pp. 22–36, 2017.
 - [24] Xinyi Zhou, Reza Zafarani, Kai Shu, and Huan Liu. Fake news: Fundamental theories, detection strategies and challenges. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, WSDM '19, p. 836–837, 2019.
 - [25] Somit Gupta, Ronny Kohavi, Dumitru Tang, Ya Xu, Reid Andersen, Eytan Bakshy, Niall Cardin, Sumita Chandran, Nanyu Chen, Dominic Coey, Mike Curtis, Alex Deng, Weitao Duan, Peter Forbes, Brian Frasca, Tommy Guy, Guido W. Imbens, Guillaume Saint Jacques, Pranav Kantawala, Ilya Katsev, Moshe Katzwer, Mikael Konutgan, Elena Kunakova, Mínyong Lee, MJ Lee, Joseph Liu, James McQueen, Amir Najmi, Brent Smith, Vivek Trehan, Lukas Vermeer, Toby Walker, Jeffrey Wong, and Igor Yashkov. Top challenges from the first practical online controlled experiments summit. *SIGKDD Explor. Newsl.*, p. 20–35, 2019.
 - [26] Liang Wang, Kuang-chih Lee, and Quan Lu. Improving advertisement recommendation by enriching user browser cookie attributes. In *Proceedings of the 25th ACM International Conference on Information and Knowledge Management*, CIKM '16, p. 2401–2404, 2016.