

絵画への視線から興味要因を特定するアテンション機構 Attention that identifies interest factors from the gaze on paintings.

泉 青葉[†] 島川 博光[†] 原田 史子[†]
Aoba Izumi Hiromitus Shimakawa Fumiko Harada

1. はじめに

価値や効用が高いものであっても、世間一般に浸透していないものは数多くある。これは製品だけの問題ではなく、芸術品や人材などについても問題である。より広く認知されるには効果的にその価値や効用を伝える必要がある。

効果的な宣伝の方法に用いられる技術として協調フィルタというものがある。これはユーザの検索内容や視聴内容などのデータをもとに、適切な製品を表示す手法であり、これは今までも多くの研究が行われている[1][2][3]。しかし、これらはユーザの傾向をもとに推薦するため、世間に浸透していない、新規の製品の興味を誘発するような提案は難しい。よって、ユーザの傾向ではなくユーザ個々が持つ特有の興味を推定するような取り組みが必要である。

本研究ではユーザの興味を推定する手法を、トランスフォーマーのアテンション機構で実現する。ユーザの視線データと特性データを合わせ、ひとつの訓練データとし、モデルを訓練する。モデル内のパラメータを分析することで、ユーザのどのような特性がどのような点に対して作用するかを考察する。また、本研究では実際にユーザの興味を推定できているかの検証に絵画を用いる。本研究の結果を用いることで、ユーザの興味を把握でき、ユーザの興味から適した絵画を推薦できる。

2. ユーザ特性による興味の推定

2.1 絵画と価値

ユーザの興味が適切に推定できているかを検証するには以下要件を満たすものが必要である。

- (1) 画像データとして扱える
- (2) 種類が豊富である
- (3) 興味対象の候補が複数個所ある
- (4) 被験者による評価が容易
- (5) 被験者にとって身近でない

よって本研究では絵画を使って検証する。

2.2 視線データと興味

人間は目的によって視線が流動的に変化する。これは注意の焦点を動的に切り替えているためであり、眼球運動は直接興味を意味しない。しかし、人間がとある対象を長い時間眺めている場合、その対象に興味に向いていることを示している[4]。よって興味を推定する場合、視線が対象に向いている時間をもとに推薦に用いることは有用である。

また、人間の目から興味を推定する別のアプローチとして、瞳孔径を測定するという手法がある。目の瞳孔径は人間の心理状態を反映する。感情推定を行う際に、瞳孔径を

もとに分析を行うと、人間の瞳孔径と興味の間に関連があることが示されている[5]。よって瞳孔径をもとに推薦をすることは、ユーザの興味を直結した製品を推薦すると同義である。

以上より、目から得られる情報は興味の推定に大きく貢献することが分かる。これら情報を収集する技術にアイトラッキングというものがある。アイトラッキングとは PC などのデバイス上でどこを見ているかを座標として得る技術である。またアイトラッキングを利用した眼球運動のデータ取得は、実験の際の被験者の負担が少ない。これにより自然な状態で実験を行え、実験時と実用時のユーザの差を小さくすることができる。よって本研究ではアイトラッキングを行えるセンサの一つである Tobii を使用し、目の視線の変化及び瞳孔径を取得し、興味の推定に用いる。

2.3 ユーザ特性と興味

人間が画像内のどこを注視するかを推定したものを顕著性マップといい、深層学習などを用いた予測などが行われている。

よく用いられるものに CNN というものがある。CNN は深層学習の一種であり、入力に画像を用いる場合、画像内の特徴を顕在化して予測を行うことができる。CNN による視線予測では、アイトラッキングによって得た視線データを、一定時間当たりの平均とし、画像データに対する正解ラベルとして用いる[6]。これにより、顕著性マップを、画像データから予測ができ、どのような点が人間の注意を引くかを明らかにできる。

CNN 以外にも Transformer を使用し視線を予測する研究も存在する。Transformer とは、アテンション機構を用いて、与えられたベクトル内の影響をもとに予測をするモデルである[7]。アテンション機構により入力ベクトルの位置情報及び時系列間の影響をもとに推定が行えるため、動画の視線データを与え、視線の動きを予測するような研究がある[8]。

しかし、これらは本研究の目的であるユーザ固有の特性による視線への影響には触れていない。ユーザの特性を顕著性マップに反映するには入力ベクトルに対象のベクトルとは別に、ユーザ特性を表すベクトルが必要であり、それらの影響をもとに推定を行うべきである。アテンション機構はそれらの要件を満たし、ユーザ特性による顕著性マップの作成に最適である。よって本研究ではソース・ターゲット注意による Transformer を使用し、人間の視線を予測する。

[†]立命館大学 Ritsumeikan University

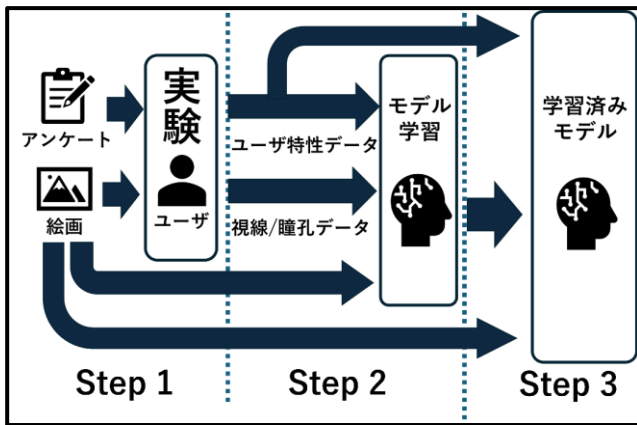


図1「手法概要図」

3. 提案手法

本研究ではユーザ特性と興味の間を調査する。図1は手法概要を表した図である。本研究は3段階に分けられ、それぞれ「ユーザごとの視線データの取得」「視線の予測モデルの作成」「視線予測からユーザの興味を推定」である。

3.1 Step1:ユーザごとの視線データの取得

本研究ではユーザの興味を推定に視線と瞳孔径という生体データの他、ユーザ個々の特性を使用する。よって、Tobiiを用いた実験で視線と瞳孔径のデータを得る。実験用に絵画データを複数枚用意し、被験者がそれらを鑑賞しているところをTobiiでデータを収集する。この際、被験者は絵画を規定時間自由に鑑賞するよう指示し、被験者の興味による視線の変化を促す。次に被験者ごとに得たデータをもとに、画像内の各ピクセルの視聴時間を計算し、絵画データそれぞれでヒートマップを作成する。このヒートマップは被験者が規定時間内でどのくらいそのピクセルを眺めていたかを意味し、興味を推定に必要である。

次にユーザの特性を得るための実験を行う。このユーザ特性はユーザ個々の持つ価値観や性格が数値として現れるのが好ましい。よって本研究では被験者にアンケートを実施し、それらを被験者の特性として扱う。

3.2 Step2:視線の予測モデルの作成

ユーザの持つ特性は異なり、ユーザ固有の興味を推定するには専用のモデルを作成する必要がある。よって本研究では被験者ごとにTransformerモデルを作成し、その被験者から得られたユーザ特性データと視線/瞳孔径データをもとに訓練を行う。入力ベクトルに被験者のユーザ特性データと実験に使用した絵画データを使用し、正解ラベルに視線/瞳孔径データを用いる。これにより、絵画に対して被験者固有の興味を予測できるように学習を進めることができる。

3.3 視線予測からユーザの興味を推定

実験により得たデータで学習したモデルを使って、未学習の絵画に対して興味を予測を行う。入力ベクトルに被験者のユーザ特性データと実験に使用していない絵画を入れると、学習によって得たユーザの特性をもとに興味のヒートマップが出力される。

しかし、現段階では出力されたヒートマップが興味を示しているかはわからない。よって本研究では未学習の絵画データに対しても視線/瞳孔径データを取得し、それと比較することで検証を行う。

検証の結果、十分な精度が出た場合、本モデル内のパラメータは十分に被験者の興味を表していると考えられる。よって入力ベクトルに用いたユーザ特性データを操作することにより、特定の特性による視線/瞳孔径の変化を出力することができる。この特性ごとの出力とパラメータを分析することでユーザ固有の特性による視線/瞳孔の影響を考察でき、本研究の目的を達成することができる。

4. おわりに

本研究ではトランスフォーマーのアテンション機構の特性を利用し、ユーザの特性による興味要因の特定の手法を提案した。ユーザの興味は視線と瞳孔径に表れ、それをトランスフォーマーで予測することで興味を推定する。実験によってユーザ特性データと視線/瞳孔データを収集。実験データでモデル学習し、興味による視線/瞳孔を被験者ごとに予測する。この学習済みモデルを使用することで未学習の絵画に対して予測を可能にする。

本研究は協調フィルタのようなユーザの傾向ではなく、個々の有する特有の趣向を反映した興味を推定することができる。よって本研究を使用することで、ユーザにとって最適なものを推薦することが可能になる。また、本研究では興味の種類による分析ができ、ユーザ特性の特定の要因による推薦も可能になる。

参考文献

- [1] Carlos A. Gomez-Urbe and Neil Hunt. "The Netflix Recommender System: Algorithms, Business Value, and Innovation.", *ACM Trans. Manage. Inf. Syst.* Vol 6, No 4, Article 13 (January 2016), pp19. <https://doi.org/10.1145/2843948>
- [2] Bamshad Mobasher, Robert Cooley, and Jaideep Srivastava. "Automatic personalization based on Web usage mining. *Commun.*" *ACM*, 43, 8, (Aug. 2000), pp142-151. <https://doi.org/10.1145/345124.345169>
- [3] Liangjie Hong and Brian D. Davison. "Empirical study of topic modeling in Twitter. In *Proceedings of the First Workshop on Social Media Analytics (SOMA '10)*". Association for Computing Machinery, New York, NY, USA, pp80-88. <https://doi.org/10.1145/1964858.1964870>
- [4] Hirayama, Takatsugu and Dodane, Jean-Baptiste and Kawashima, Hiroaki and Matsuyama, Takashi, "Estimates of User Interest Using Timing Structures between Proactive Content-Display Updates and Eye Movements", *IEICE Transactions*, Vol 93-D, pp1470-1478, (2010),doi: 10.1587/transinf.E93.D.1470.
- [5] Eckhard H. Hess and James M. Polt, "Pupil Size as Related to Interest Value of Visual Stimuli", *Science*, Vol 132, No 3423,pp349-350 (1960), doi 10.1126/science.132.3423.349.
- [6] Florian Strohm, Mihai Băce, and Andreas Bulling. 2024. Learning User Embeddings from Human Gaze for Personalised Saliency Prediction. *Proc. ACM Hum.-Comput. Interact.* 8, ETRA, Article 229 (May 2024), 16 pages. <https://doi.org/10.1145/3655603>
- [7] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*. Curran Associates Inc., Red Hook, NY, USA, 6000-6010.
- [8] Melanie Heck, Janick Edinger, Jonathan Bünemann, and Christian Becker. 2021. The Subconscious Director: Dynamically Personalizing Videos Using Gaze Data. In *Proceedings of the 26th International Conference on Intelligent User Interfaces (IUI '21)*. Association for Computing Machinery, New York, NY, USA, 98-108. <https://doi.org/10.1145/3397481.3450679>