

Web アクセス履歴データにおける本人一致率の評価

Evaluation of Personal Match Rate in Web Browsing History Data

加藤 拓弥*
First Author

満保 雅浩†
Second Author

キーワード Web アクセス履歴データ、jaccard 係数、本人一致率、bookmark

あらまし

Web マーケティングなどビジネスに Web アクセス履歴データが活用されている場面が多々ある。Web アクセス履歴データはパーソナルデータに含まれる。使用するパーソナルデータが個人との関係性が弱いものであったとしても、分析を進めていくにつれ特定の個人を識別できるようになる可能性がある。法整備は進んでいる反面、保護技術については現在も開発が進んでいる。ゆえにパーソナルデータのプライバシー保護は必須であり、また保護技術の開発は急務である。

本論文では Web アクセス履歴データに着目し、そのプライバシーリスクを本人一致率という観点から評価を行う。Web アクセス履歴データの期間は 15ヶ月分を使用する。そのうち最新の 1ヶ月分を元にして、残り 14ヶ月分と比較し、本人一致率を導出する。

手法として、2つの期間において、ユーザー ID とアクセスした URL を収集し、アクセスした URL 集合をそれぞれ作成する。ID とその集合間の類似度から本人一致率を求める。集合間の類似度には jaccard 係数を用いる。

はじめに、アクセスした URL 集合を用いた本人一致率を求める。またこれに加え、bookmark、timestamp、URL のカテゴリーなどを追加した場合の本人一致率を求める。ここで示す bookmark とは、ユーザーが bookmark 機能を用いてアクセスした URL のことを指す。また、URL のカテゴリーとはその Web サイトが何について記載されているのか大まかな括りのことである。例えば天

気について記載されている Web サイトの場合、カテゴリーは天気となる。この分類には NetStar の分類を用いる。

これらを通して、アクセスした URL 集合を用いた本人一致率と他の観点を追加した場合の本人一致率を比較し、bookmark を加えた場合の本人一致率が優位であることを示す。

参考文献

- [1] WarpDrive プロジェクト, ”<https://warpdrive-project.jp/index.html>”
- [2] NetSTAR カテゴリー一覧, ”<https://www.netstar-inc.com/products/category/>”

* 金沢大学, 石川県金沢市角間町, Kanazawa University, Kakumamachi, Kanazawa City, Ishikawa Prefecture

† 金沢大学, 石川県金沢市角間町, Kanazawa University, Kakumamachi, Kanazawa City, Ishikawa Prefecture