国立大学法人 大阪大学
〒565-0871 大阪府吹田市山田丘 1-1
TEL: 06-6877-5111 (代)
www.osaka-u.ac.jp

Press Release

研究成果

平成 30 年 11 月 5 日

分野： 工学系　　キーワード： 人工知能、画像認識、プライバシー保護

画像認識に伴う内心プライバシー情報の漏洩をブロックする仕組みを世界初開発

―人工知能技術の安心安全な実応用に向けて―

【研究成果のポイント】

- 写真に反映される持ち主の趣味･嗜好等の「内心プライバシー情報」が人工知能に基づく画像認識サービスの利用により漏洩することを回避可能な新方式を開発。
- これまで内心に関するプライバシー上の懸念は十分に周知されておらず、その解決法も議論されていなかった。
- 「安心安全」を重視した人工知能技術の研究開発ならびに実応用に大きく寄与すると期待。

❖ 概要

大阪大学大学院工学研究科の馬場口登教授、中村和晃助教らの研究グループは、人工知能※1 による画像認識サービス※2 において利用者の趣味･嗜好といった内心（以下、「内心プライバシー情報」と呼びます）が漏洩し得ることを明らかにし、この問題を解決するための新たな仕組みを世界で初めて開発しました（図 1）。

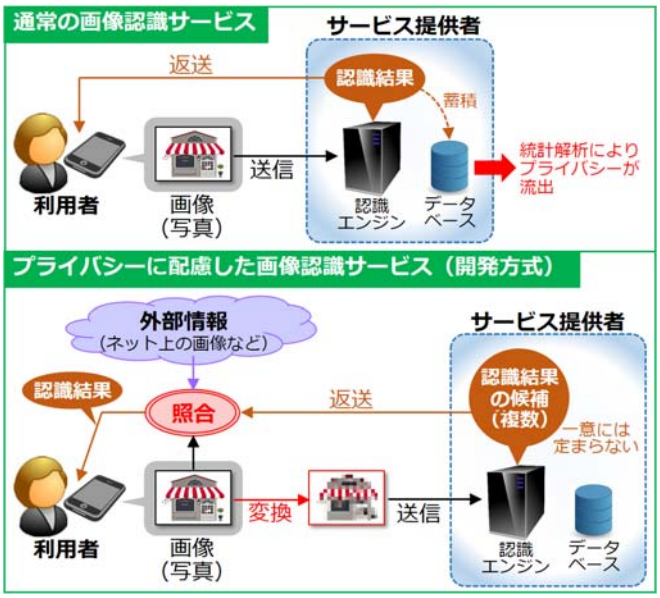

図 1　プライバシーに配慮した画像認識方式の概要

昨今、人工知能技術の発達により、インターネット上では非常に高性能な画像認識サービスが提供されるようになりました。しかし、このようなサービスでは、**サービス提供側が画像認識の結果を全て把握できる立場にあるため、人々が気軽に撮影した様々な写真に対する画像認識結果を蓄積し統計解析を行うことにより、個人の内心プライバシー情報がサービス提供側に流れる**事態が起こり得ます。例えば、写真中の建物の名称や品物の銘柄などが人工知能により認識･解析されれば、「写真の持ち主はこの時間この場所にいた」「写真の持ち主はこういったモノを好む傾向がある」などの情報がサービス提供側に漏洩する結果となります。

近年の人工知能技術の発展は目覚ましく、その高度化手法や実応用が活発に研究されていますが、上記のようなプライバシー上の懸念は十分に周知されておらず、その解決法も議論されてきませんでした。今回、本研究グループは、**この問題を明示的に指摘し、更にこれを解決可能な新方式を確立しました。本方式では、送信する写真に予め画像変換処理を施すことによりサービス提供側には認識結果の候補を絞ることしかできないようにする一方、利用者側では返送された「候補」を変換前の元の写真と照合することにより正しい認識結果が得られます**。

本研究成果により、画像や映像、音声といったマルチメディアの解析を目的とした人工知能技術の研究開発においてはプライバシーの視点も重要であることが社会に広く周知され、その結果、同分野の最先端技術が将来より安心安全な形で実社会に導入されることに大きく貢献するものと期待されます。

本研究成果は、米国電気電子学会（Institute of Electrical and Electronic Engineering; IEEE）発行の論文誌「IEEE Transactions on Information Forensics and Security」にて、2018 年 10 月 24 日（水）（日本時間）にオンライン速報版で公開されました。

大阪大学
OSAKA UNIVERSITY
国立大学法人 大阪大学
〒565-0871 大阪府吹田市山田丘 1-1
TEL: 06-6877-5111 (代)
www.osaka-u.ac.jp

❖ 研究の背景

画像認識は、画像(写真)中のどの位置にどのようなモノが写っているかを認識・解析する技術の総称で、その対象は人、動植物、建物、乗り物、道具など多岐に渡ります。人間には簡単ですが、コンピュータには非常に難しい処理であり、実用化は遠い状況が続きました。しかし、2012 年以降、ディープラーニングの登場によりその性能は急速に進歩し、現在では非常に高性能な画像認識エンジンを有する Web サービスがインターネット上で複数登場しています。一方で、画像認識サービスの高精度化はプライバシー上の問題を引き起こす要因ともなり得ます。例えば、ある利用者の所有する写真に対し画像認識処理を適用すると、写真中の人物の属性(性別、年齢など)や場所(地名や建物の名称など)、品物の銘柄といった情報が認識結果として得られます。サービスの提供側はこれらの情報を知り得る立場にあるため、これを統計解析することにより、その利用者がどのような属性の人々と行動を共にすることが多いか, どのようなモノや場所を好む傾向があるか、さらには過去実際に訪れた場所の履歴など、様々な情報を把握できることになります。これらは、利用者自身の趣味・嗜好といった内心の状態を反映しており、プライバシーに関わる情報と言えます。(図 2)。

図 2 内心プライバシーが流出し得る画像認識サービスの例
(テーマパーク等における施設案内サービス)

この「内心プライバシー情報の漏洩」という問題は、これまであまり検討されてきませんでした。この背景には画像認識の難しさがあります。画像認識の性能が十分でない場合、サービス提供側は誤った認識結果に対して統計解析を行うことになるため、内心プライバシー情報の流出は起こりません。しかし近年、認識性能の向上により上記の問題が表面化しつつあります。今回、本研究グループはこの点を世界で初めて指摘しました。

❖ 本研究成果の内容

上述の問題は、画像認識の過程において、その結果をサービス提供側が完全に特定できてしまう点に根本的な要因があります。そこで本研究グループは、サービス提供側が認識結果を一つに特定できないにもかかわらず利用者側には正しい結果とそれに基づくサービス情報が伝わるような画像認識の仕組みを開発しました。具体的には、利用者が対象となる写真をサービス提供側に送る際、特定の画像変換処理を自動的に適用します。この変換処理はサービス提供側の画像認識性能を低下させる働きを持ちます。すなわち、どれが正しい認識結果であるのかをサービス提供側には見分けにくくし、その候補を絞ることしかできない状況に導きます。これにより、提供側は認識結果の候補を複数挙げるに止まり、それらの候補が利用者に返送されます。一方、利用者は、変換前の元の写真を有していますので、これを返送された「候補」と(外部情報も利用しつつ)照合することにより、最終的に認識結果を一つに特定します。以上の仕組みにおいて、サービス提供側の画像認識性能を低下させすぎると、認識結果の候補が多岐にわたり、利用者による最終処理を高速・高精度に実行することが困難となります。逆に、提供側の画像認識性能の低下幅が小さいと、内心プライバシー情報の流出を防ぎきれません。このバランスを考慮し、認識性能の低下幅を適切に制御できる画像変換手法を考案した点が本研究成果の重要なポイントです。

大阪大学
OSAKA UNIVERSITY
国立大学法人 大阪大学
〒565-0871 大阪府吹田市山田丘 1-1
TEL: 06-6877-5111 (代)
www.osaka-u.ac.jp

❖ 本研究成果が社会に与える影響（本研究成果の意義）

本研究成果により、画像や映像、音声といったマルチメディアデータの認識技術を研究・開発する際には認識結果に含まれるプライバシー関連情報に配慮する必要があることが一般社会や研究者コミュニティに広く周知され、その結果、人工知能を用いた多様な技術が将来より安心安全な形で実社会に導入されるようになるものと期待されます。また、「プライバシー上の問題を生じさせることなくマルチメディア認識を実現する人工知能の設計」という新たな研究領域を創出し、人工知能分野の発展に大きく寄与することが期待されます。

❖ 特記事項

本研究成果は、2018 年 10 月 24 日（水）（日本時間）に、米国電気電子学会（Institute of Electrical and Electronic Engineering; IEEE）が発行する論文誌「IEEE Transactions on Information Forensics and Security」にてオンライン速報版で公開されました。

タイトル：“Encryption-Free Framework of Privacy-Preserving Image Recognition for Photo-Based Information Services”
著者名： Kazuaki Nakamura, Naoko Nitta, and Noboru Babaguchi
DOI： 10.1109/TIFS.2018.2876752

なお、本研究は、日本学術振興会 科学研究費助成事業 基盤研究(S)「メディアクローン攻撃を防御するコミュニケーション系」（課題番号：JPH06302、研究代表者：馬場口登）ならびに同 基盤研究(C)「情報セキュリティレベルの高いサーバ・クライアント型メディア認識機構の開発」（課題番号：17K00235、研究代表者：中村和晃）の一環として行われました。

❖ 用語説明

※1 人工知能

人間が行っているような知的処理をコンピュータ上で実現する技術やシステムの総称。近年では特に機械学習に基づくものを指すことが多い。画像認識の場合では、写っているモノの名称と紐付けられた画像（写真）を大量に用意し、これをデータとして取り込んで機械学習を行うことにより、様々なモノを認識するための基準をコンピュータ自身に構築させることが可能となる。このようにして開発された画像認識エンジンを本稿では「人工知能による画像認識」と呼んでいる。

※2 画像認識サービス

送られてきた写真に対し画像認識を適用し認識結果を返送するサービス、またはそれを応用した情報提示サービスのこと。後者の例として、施設や店舗の外観を撮影することによりその施設に関する情報（バーゲン品や口コミ情報など）がその場で得られる施設案内サービスや、店頭で製品を撮影することによりその製品に関する詳細情報がその場で得られる製品紹介サービスなどが考えられている。