

論文の内容の要旨

論文題目 ウェブの情報統合に基づく人物属性の推定の方法論に関する研究

氏 名 大澤 昇平

本研究では、ウェブの情報に基づき、機械学習を用いて人物属性の推定を行う。人物属性とは性別や能力などの人物の特徴のことである。人物属性の推定は、採用活動やマーケティングなど対象とする人物のことを知る必要がある活動において有用である。

機械学習の枠組みにより人物属性の推定を行なう研究に関しては、Twitterから性別を推定する研究などが挙げられる。しかし、ウェブ上の情報は欠損が多く、単一の情報源からでは知ることのできない情報も多い。一方、機械学習の枠組みにより人物属性の推定を行う研究として、分散する複数の情報源を統合するという観点での研究については行われていない。こうした研究には、エンティティ・リンキングなどの情報統合の知見を活かすことができると考えられる。複数の情報源を用いることで属性推定の精度向上を行なうことが期待できる。

こうした研究が行われていない理由は、最近まで、ウェブ上に公開されている人物情報は企業が特定の目的のみに利用することを規定しており、そこから分析者が情報統合を目的としてデータを取り出すことは困難だったからである。しかし、近年、一部の企業やパブリックセクターが人物情報をウェブ上の二次利用を促進する仕組みを通してデータを提供する事例が増えてきており、これらから得られる情報を活用できることが期待できる。このような仕組みから得られる情報を統合し属性推定を行うために、本研究では、a) 属性推定に活用できる情報の収集法、b) 収集した情報を統合した素性生成の方法について研究を行う。

このような要素課題に関する研究を通し、機械学習により人物属性の推定の手法を構築する方法論を示す。

