

# 個人情報活用と自然言語処理

松井くにお（ニフティ株式会社）

matsui.kunio@nifty.co.jp

## 概要：

ビッグデータの活用が話題になっている。ガートナーが毎年発表するハイプサイクル(\*1)では、2013年には期待度のピークを迎え、「過度の期待」の対象となっている。こういった活用において、CCCのTポイントカード情報の「共同利用」や、JR Suica情報の売買は、現行の個人情報保護法の様々な問題点を明らかにしている。

そこで、2015年には、世界水準に合わせた個人情報保護法の改正が行われる予定で、ユーザから得られる個人情報を個人が特定、識別できない匿名化を行なうことによって、ユーザの同意を得なくてもそれらの情報を活用できるようになる。

現在の個人情報活用は、経済性が社会性より優先しており、得られた情報を組織内で閉じた形で情報推薦に利用されているものが多い。しかしながら前述のように仮名化だけを行なった Suica 情報を他社へ転売後も情報の持つ属性から再識別可能性を残しており、社会から批判の対象となった(\*2)。現代の企業の価値創造には、経済性と社会性の両立が不可欠である。

そこで、できる限り有効な情報活用と、ユーザが安心して情報を提供できるプラットフォームを提案したい。そこには、自然言語で書かれた大量の個人情報の様々な属性に対してその値を抽出できる技術と、個人を特定化できない情報群を作る概念体系を管理する技術によって、情報の匿名化を行なうことができる。本討論では、このような自然言語処理の新しい応用分野について、その可能性を議論したい。

(\*1):ハイプサイクル：<http://www.gartner.co.jp/press/html/pr20131015-01.html>

(\*2)鈴木正朝教授の意見：<http://itpro.nikkeibp.co.jp/article/COLUMN/20130827/500450/>