

特許出願からみた匿名化関連技術の技術動向 —平成 29 年度特許出願技術動向調査より—

金沢 史明^{†1} 岸野 徹^{†1}

概要：特許庁では、平成 29 年度特許出願技術動向調査において「匿名化技術」を調査テーマとして取り上げ、統計調査を実施した。本調査では、匿名化技術とその周辺技術に関連する国内外の特許出願約 17,000 件、論文約 19,000 件を読み込み、その内容について、用途、課題、要素技術等の観点により作成した技術区分に峻別し、統計的な分析を行った。本調査で得られた分析結果を報告する。

キーワード：匿名化、特許出願、技術動向

Technology Trends on Anonymization and Related Technologies Based on Patent Application Analysis: FY2017 Patent Application Technical Trend Surveys

Fumiaki Kanazawa^{†1} Toru Kishino^{†1}

Abstract: We surveyed technology trends on anonymization and related technologies based on patent applications and journal articles. We read domestic and international 17,000 patent documents and 19,000 journal articles and classified them into various technological classifications we created. We report on results of this survey.

Keywords: Anonymization, Patent Applications, Technology Trends

1. はじめに

匿名化技術とは、特定の個人を識別可能なデータ（個人情報[Ⓐ]）から、特定の個人を識別することができないように、データを加工する技術である。

近年、SNS の流行やクラウドコンピューティングの発達に伴い、インターネット上のストレージには、パーソナルデータ[Ⓑ]を含む膨大なデータが蓄積されている。ビッグデータのように収集・蓄積されたデータに対して匿名化技術等を適用し生成された匿名化情報は、その利用価値を残したまま個人を特定困難にするものであるため、個人情報に適用される取り扱い上の制約が緩和され、第三者提供あるいは目的外利用が可能となる。これにより、各種データの流通・利活用が促進され、更には、新たなサービスの創造に資することが期待されている。

また、改正個人情報保護法では、匿名加工情報の規定の新設、不当な差別や偏見につながるような配慮個人情報の規定の新設、個人識別符号として、身体的特徴に基づく変換符号（顔特徴情報、指紋特徴情報）と、対象者ごとに個人を特定するために付与される符号（旅券番号、マイナ

ンバー）の 2 つの categories を明確化するなど、第三者へのデータ提供に関する配慮が強化された。

一方で、これらのデータには機密性を保護するための対策が講じられているものが多いとはいえ、これらの対策が必ずしも万全であるとはいえず、パーソナルデータの漏えい問題がマスコミなどで報じられる事例も後を絶たないのが現状である。特に我が国においては、公的機関や大企業で発生する個人情報流出事件が報道として大きく取り上げられる傾向にあり、プライバシー漏えいへの不安からパーソナルデータの流通に不安を感じる傾向が欧米に比べて強いといわれている。例えば、平成 29 年版情報通信白書[1] (pp. 78-79) によれば、パーソナルデータの提供について「とても不安を感じる」の割合は、日本においては比較的高く (24.7%)、「やや不安を感じる (59.4%)」と合わせると 8 割超であるのに対し、米国・英国・独国の 3 か国は 6 割程度で、明確な差があり (図 1)、匿名化技術に対する潜在的なニーズは高いといえる。

^{†1} 特許庁

Japan Patent Office

a 個人情報保護法（第 2 条第 1 項）では、『個人情報』とは、生存する個人に関する情報であつて、次の各号のいずれかに該当するものをいう。
一 当該情報に含まれる氏名、生年月日その他の記述等により特定の個人

を識別することができるもの 二 個人識別符号が含まれるもの」と定義されている。

b 個人情報に加え、個人情報との境界が曖昧なものを含む、個人と関係性が見出される広範囲の情報（平成 29 年版情報通信白書[1], pp. 53-54）。

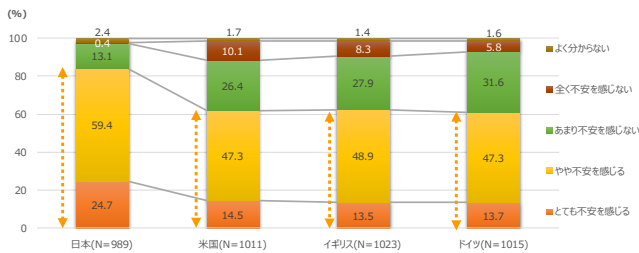


図 1 パーソナルデータの提供に対する不安感^c

特許庁では、市場創出に関する技術分野、国の政策として推進すべき分野を中心に、今後の進展が予想される技術テーマを選定し、特許出願技術動向調査を実施している。平成 29 年度においては、そのような技術テーマの一つとして「匿名化技術」を選定し、個人情報が含まれるデータの利活用のための匿名化技術とその周辺技術に関する特許や論文の動向を調査し、技術革新の状況、技術競争力の状況と今後の展望について検討した。

本稿では、平成 29 年度特許出願技術動向調査の 1 テーマである「匿名化技術」の調査結果について、その一部を紹介する。

なお、本稿は、調査結果をもとに、筆者らが独自に考察した内容を一部含むものである点に留意いただきたい。

2. 調査概要

本調査の調査対象を表現した技術俯瞰図を図 2 に示す。本調査では、匿名化技術と、秘匿計算・秘密計算、差分プライバシー、パーソナルデータサービス・パーソナルデータストア (PDS) や情報銀行など、匿名化技術とは異なった手法でプライバシーを保護する周辺技術とを含めて調査対象とした (以下、匿名化技術と周辺技術とをまとめて「匿名化関連技術」と呼ぶ)。

調査範囲、調査手法については、以下のとおりである。

(1) 調査範囲

- ・特許文献^d
 - 対象期間：2006 年～2015 年 (優先権主張年ベース)
 - 対象国：日本、米国、欧州、中国、韓国、PCT 出願
 - 使用したデータベース：Sharesearch (日立製作所提供)
- ・論文
 - 対象期間：2006 年～2016 年 (発表年ベース)
 - 使用したデータベース：Scopus (エルゼビア社提供)

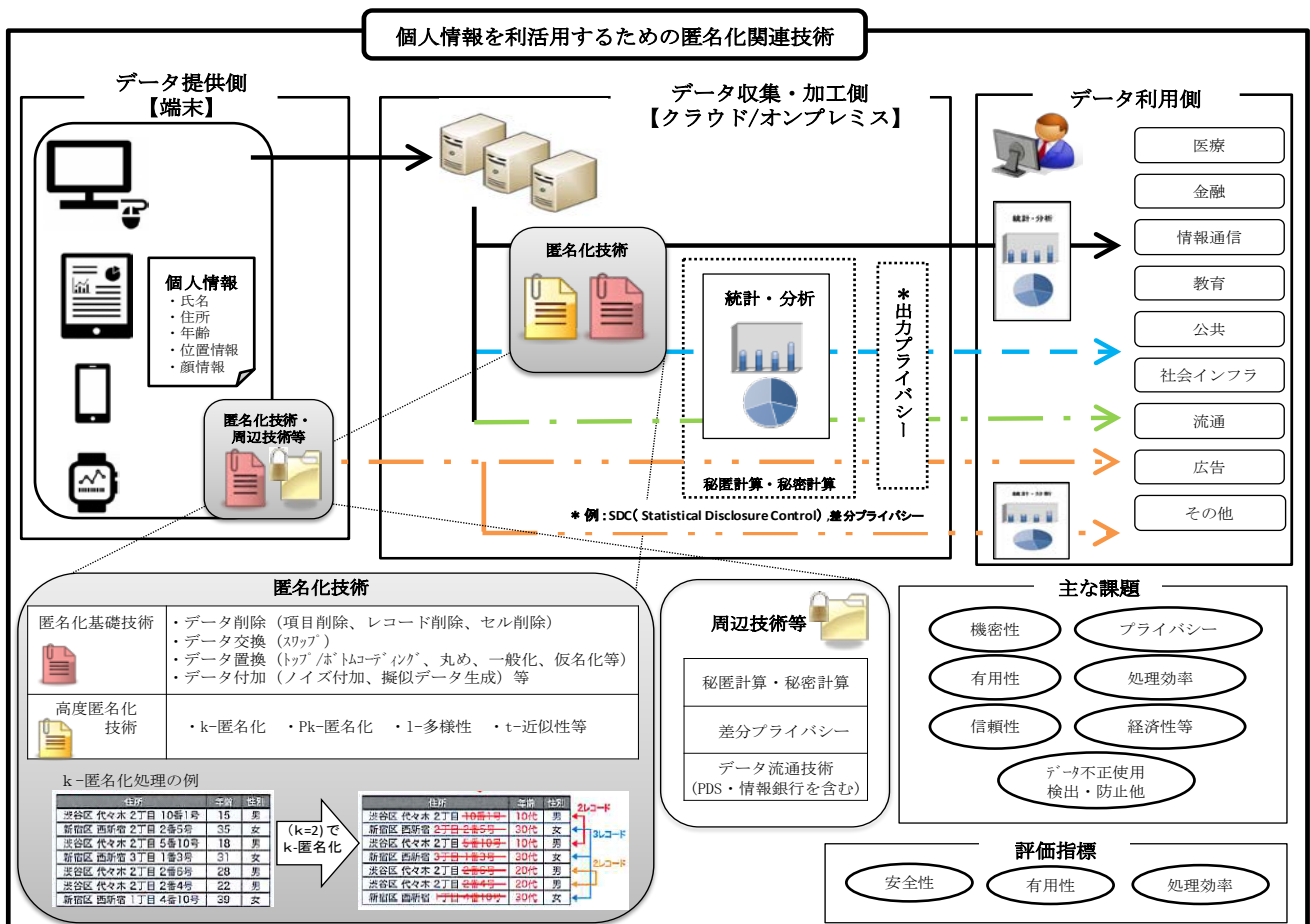


図 2 技術俯瞰図

c 平成 29 年度情報通信白書[1] (p. 79, 特に図表 2-2-4-4) を基に作成。
d 各国特許庁によって発行される特許、出願公開等の文献や、国際出願

(PCT 出願) の国際公開の文献。

(2) 調査手法

・特許文献

特許分類 (IPC^e, FI^f, CPC^g)、キーワード等を使うことによって、データベースで本調査の対象となる特許文献の母集合を得た後、母集合に含まれる特許文献を1件ずつ読み込み、あらかじめ本調査のために独自に設定した技術区分に分類した。

・論文

キーワードを使うことによって、データベースで本調査の対象となる論文の母集合を得た後、母集合に含まれる論文を1件ずつ読み込み、あらかじめ本調査のために独自に設定した技術区分 (特許文献の分類に用いたものと同じ) に分類した。

3. 市場環境

国内におけるパーソナル情報市場の分野別の市場規模を図3に示す。金融・保険業が6,036億円と最も多く、次に、小売業4,487億円、宿泊・旅行業4,368億円、通信3,675億円と続いている。

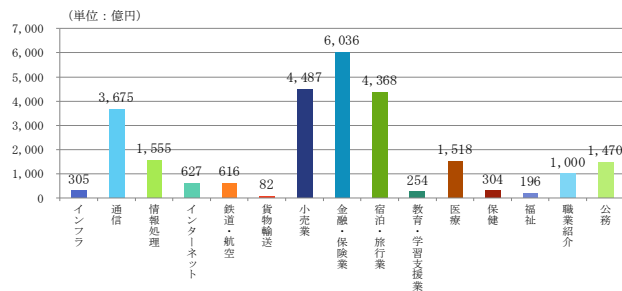


図3 国内のパーソナル情報市場の全体推計^h

4. 特許出願動向

4.1 全体動向

図4,5は、それぞれ、匿名化関連技術全体のファミリー件数ⁱ、出願人国籍 (地域) 別の比率と推移を示したものである。米国籍の出願人による出願が最も多く、全体の44.9%を占めている。次いで日本国籍 (22.4%)、中国籍 (17.6%)、欧州国籍 (9.2%)、韓国籍 (4.4%) となっている。また、2011~2012年頃から、特に米国籍と中国籍の出願人による出願が増加傾向にあり、それに伴い全体の件数も大きく伸びている。

図6は、日米欧中韓の5カ国・地域での出願人国籍 (地域) 別出願件数収支を示したものである。米国籍の出願人が積極的に他国への出願を行っている傾向が見られ、

欧州や韓国への出願のうち、約半分が米国籍出願人による出願となっている。また、日本や中国においても、3割強の出願は米国籍出願人によるものとなっている。一方で、米国籍以外の出願人は、他国への出願が相対的に少なく、自国への出願を中心に行っている傾向が見られる。

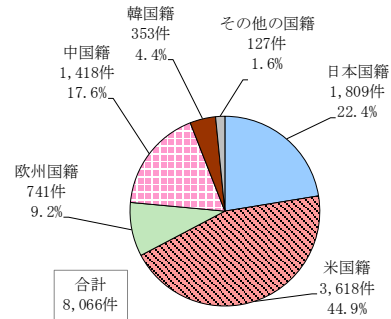
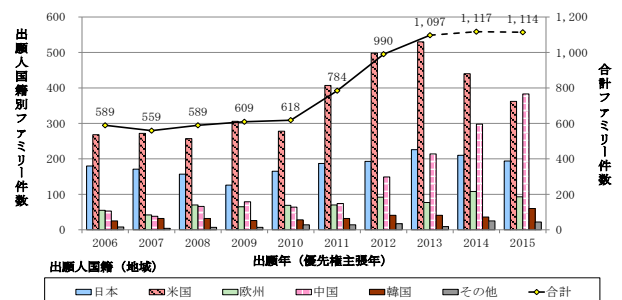


図4 出願人国籍 (地域) 別ファミリー件数の比率



注: 2014年以降はデータベース収録の遅れ、PCT出願の各国移行のずれ等で、全データを反映していない可能性がある

図5 出願人国籍 (地域) 別ファミリー件数の推移

4.2 出願人件数ランキング

表1は、日米欧中韓の5カ国に出願された、匿名化関連技術全体のファミリー件数の上位企業を示している。全体動向で見たとおり、1位にIBM、3位にグーグルなど、米国籍の企業が目立つが、2位のNTT、4位の日本電気など、日本企業も多数 (上位20社中9社) ランクインしている。

4.3 技術区分ごとの動向

4.3.1 データ形式 (構造化データ^j, 非構造化データ^k)

データ形式別一出願人国籍 (地域) 別のファミリー件数を図7に示す。日本は、非構造化データよりも構造化データの方が多く、日本以外では、すべて非構造化データの方が多くなっている。

用して、データの整理や検索ができる。コンピュータがビジネスで利用され始めたころは、全データ量のほとんどが構造化データだったが、現在では非構造化データが大半を占めるようになっている。

k 文書データ、電子メール、写真、動画など、定型的に扱えないデータ。コンピュータの利用範囲の広がりに伴い、近年、非構造化データをビジネスで活用するために、非構造化データを高効率かつ高速に管理、分析する処理技術が求められている。

e International Patent Classification (国際特許分類)

f File Index (日本国特許庁の特許分類であり、IPCを細分化したもの)

g Cooperative Patent Classification (欧州特許庁と米国特許商標庁とに共通の特許分類であり、IPCを細分化したもの)

h パーソナル情報保護とIT技術に関する調査 調査報告書[2] (pp. 171-209, 特に図表II.5.2-59)を基に作成。

i 調査対象国における優先権主張の基礎出願数をカウントしたもの。

j 顧客情報や在庫情報など、定型的に扱えるデータ。データベースを利

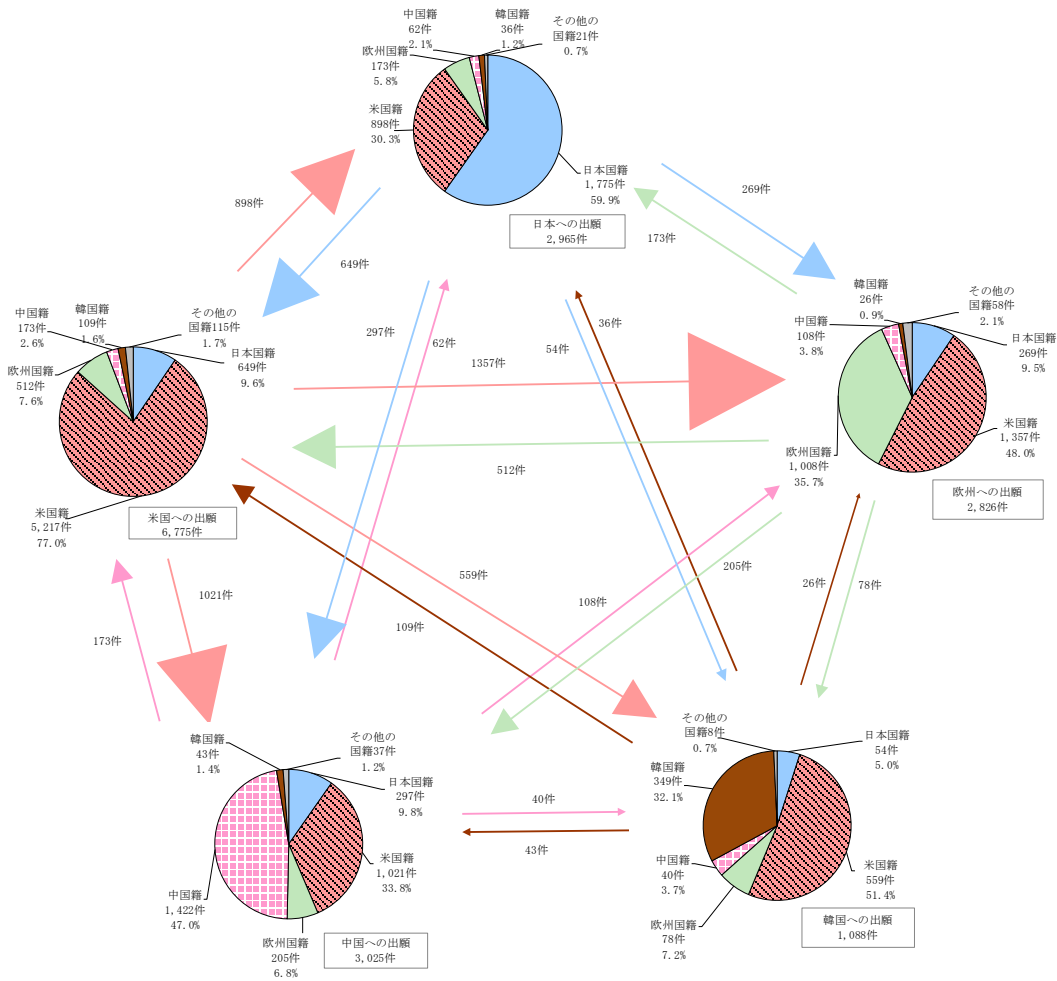


図 6 出願先国（地域）別—出願人国籍（地域）別出願件数収支
 （日米欧中韓への出願，出願年（優先権主張年）：2006年-2015年）

表 1 出願人別ファミリー件数上位ランキング
 （日米欧中韓への出願）
 （出願年（優先権主張年）：2006-2015年）

順位	出願人	ファミリー件数
1	IBM (米国)	258
2	NTT	238
3	グーグル (米国)	221
4	日本電気	219
5	富士通	182
6	マイクロソフト (米国)	180
7	日立	170
8	ノキア (フィンランド)	132
9	東芝	105
10	インテル (米国)	87
10	三菱電機	87
12	KDDI	79
13	ファーウェイ (中国)	75
14	フィリップス (オランダ)	73
15	キヤノン	70
16	クアルコム (米国)	69
17	サムスン (韓国)	68
18	富士フイルム	65
19	中興通信 (中国)	62
20	フェイスブック (米国)	60
20	ヒューレッドパッカード (米国)	60

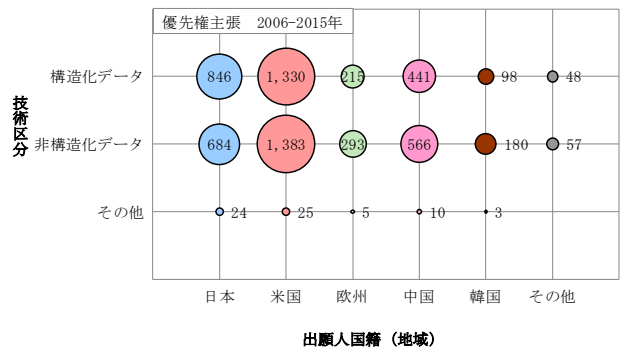


図 7 データ形式別—出願人国籍（地域）別
 ファミリー件数

また、構造化データ、非構造化データに関する出願人国籍（地域）別ファミリー件数の推移を図 8, 9 にそれぞれ示す。構造化データに関する出願は、日米欧中韓の合計が増加傾向にあるのと同様に、日本国籍出願人による出願も増加傾向にあるが、非構造化データに関する出願は、日米欧中韓全体（特に米国，中国）で増加傾向にあるのに対し、

日本国籍出願人による出願では増加傾向が見られない。

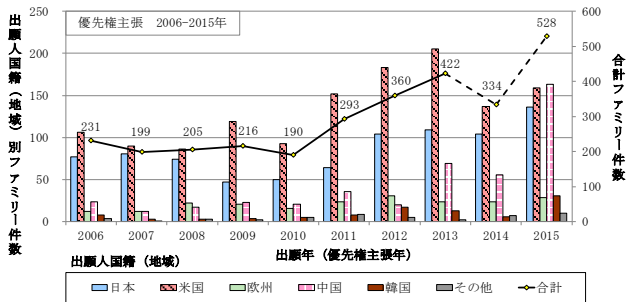


図 8 構造化データに関する出願人国籍（地域）別ファミリー件数の推移

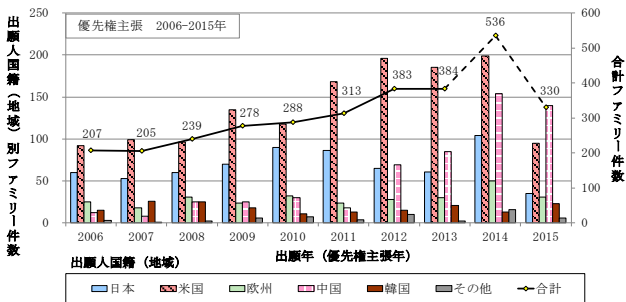


図 9 非構造化データに関する出願人国籍（地域）別ファミリー件数の推移

4.3.2 事業分野

事業分野別一出願人国籍（地域）別ファミリー件数を図 10 に示す。各国籍の出願人とも「情報通信」が最も多くなっている。また、米国籍、日本国籍の出願人は、「医療・ゲノム・健康・福祉」、「金融」、「社会インフラ」、「流通」分野など、情報通信分野以外にも満遍なく出願しており、例えば「医療・ゲノム・健康・福祉」と「流通」においては、欧州国籍、中国籍、韓国籍に比べ件数が多い。

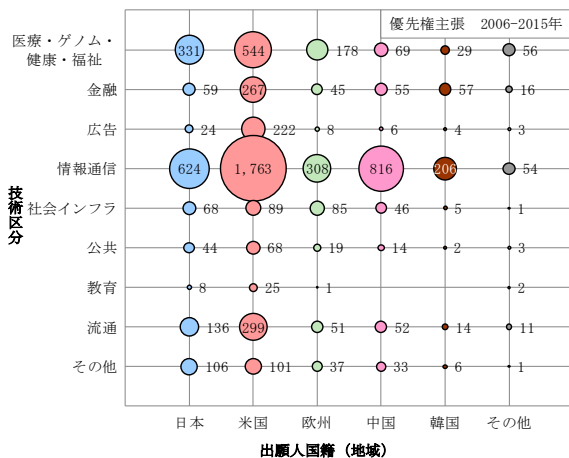


図 10 事業分野別一出願人国籍別ファミリー件数

図 11 から図 15 にかけて、事業分野別ファミリー件数推移の抜粋を示す。

「社会インフラ」と「健康管理・予防医学」においては、出願件数が全体的に増加しており、また日本国籍の出願割合が少なくないことから、今後高い成長が期待される。

一方、「金融」、「流通」、「医療」は、パーソナル情報市場の規模（図 3）が大きい事業分野と密接に関連しているところ、米国等の出願状況を見ると、我が国においても業界横断的な協力や産学連携などを通じた研究開発の促進によって、将来的な収益力アップが可能と考えられる。

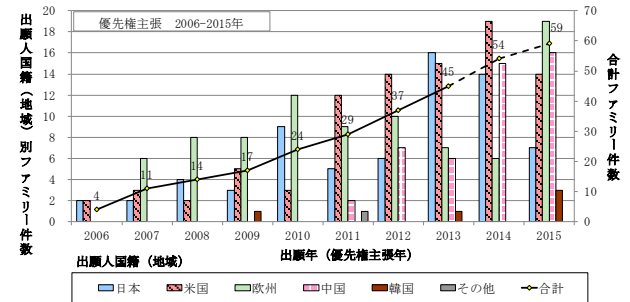


図 11 事業分野（社会インフラ）に関する出願人国籍（地域）別ファミリー件数の推移

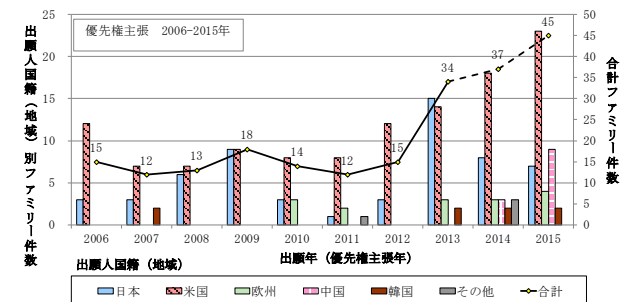


図 12 事業分野（健康管理・予防医学）に関する出願人国籍（地域）別ファミリー件数の推移

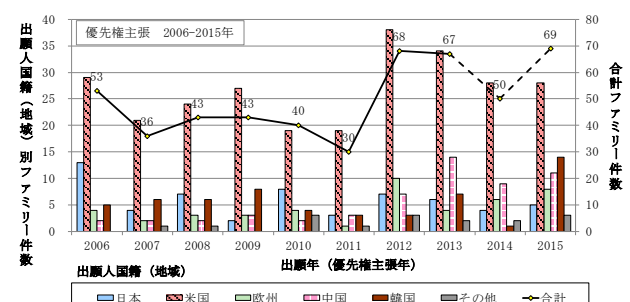
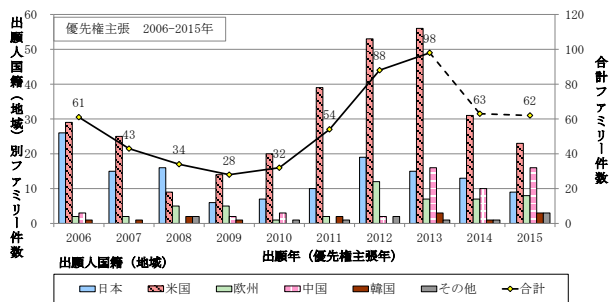
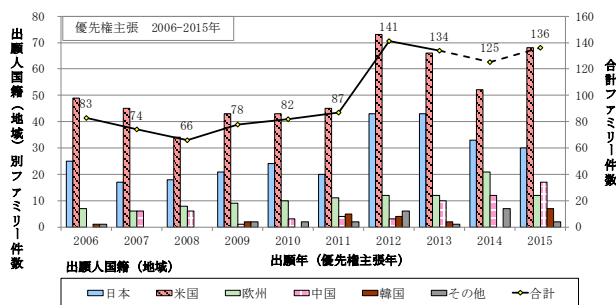


図 13 事業分野（金融）に関する出願人国籍（地域）別ファミリー件数の推移



注：2014年以降はデータベース収録の遅れ、PCT出願の各国移行のずれ等で、全データを反映していない可能性がある。

図 14 事業分野（流通）に関する出願人国籍（地域）別ファミリー件数の推移



注：2014年以降はデータベース収録の遅れ、PCT出願の各国移行のずれ等で、全データを反映していない可能性がある。

図 15 事業分野（医療）に関する出願人国籍（地域）別ファミリー件数の推移

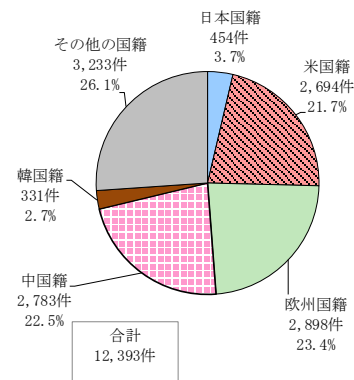


図 16 研究者所属機関国籍別論文発表件数の比率

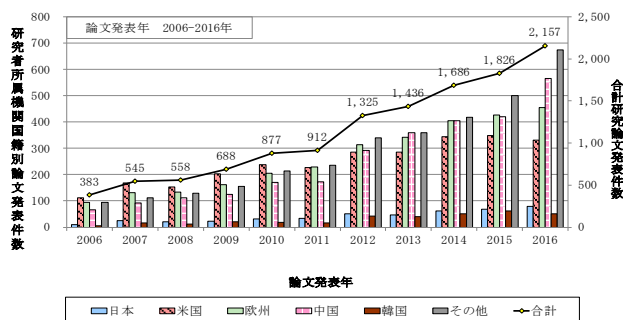


図 17 研究者所属機関国籍別論文発表件数の推移

5. 研究開発動向

5.1 全体動向

図 16, 17 は、匿名化関連技術に関する研究者所属機関¹⁾国籍別の研究論文²⁾発表件数の比率及び推移を示したものである。件数比率は、欧州国籍 (23.4%) が最も多く、次いで中国籍 (22.5%)、米国籍 (21.7%) の順で拮抗している。日本国籍 (3.7%)、韓国籍 (2.7%) はこれらと比較すると少なくなっている。なお、「その他の国籍」の内訳は、インド (867 件, 7.0%)、カナダ (369 件, 3.0%)、オーストラリア (327 件, 2.6%) の順となっている。件数推移を見ると、全体的に増加傾向にある。国籍別には、2010 年までは米国の研究論文発表件数が多く、2011 年に欧州が、2012 年に中国が米国を抜き、2013 年以降その差は広がっている。

5.2 評価指標の標準化

匿名化関連技術にはさまざまな手法が存在し (図 2 技術俯瞰図)、その性能を比較するためには評価指標が必要となる。例えば、匿名化したデータから個人を再識別したり、個人の属性を推定したりすることの困難度や、匿名化による情報損失の程度を評価することが検討されている。

特に、我が国では、情報処理学会コンピュータセキュリティ研究会が中心となって毎年開催されている PWS CUP (Privacy Workshop Cup) という匿名化及び再識別技術のコンテストを通じて、匿名化技術の評価指標の研究が進められている。

世界的に見ると、プライバシー保護に関しては、ISO (国際標準化機構)、IEC (国際電気標準会議) などで国際標準化に向けた議論が進められているが、匿名化技術の評価指標についての標準化作業はまだこれからである。

我が国の、匿名化技術の評価指標の研究に関する先進的な取り組みは、社会のニーズに適応した実用化と、国際的な標準化作業への貢献が期待される。

6. 提言

調査結果を踏まえ、提言をまとめた。以下に概要を示す。

なお、特許庁 Web サイトにて報告書 (概要) を公開している³⁾ので、併せて参照されたい。⁴⁾

1 論文の第一著者の所属機関

2 Scopus (エルゼビア社提供) に収録された論文 (電子化された英文抄

録を有する論文)。

3 <http://www.jpo.go.jp/shiryou/gidou-houkoku.htm>

6.1 匿名化関連技術動向の観点（提言 1）

- (1) 我が国においても、クラウド上には多様なデータが蓄積され、パーソナルデータを含む非構造化データを活用する場面が増えているところ、目的外利用や第三者提供の需要を勘案しつつ、パーソナルデータの流出による被害を低減するため、構造化データを対象とした匿名化関連技術に加えて、非構造化データを対象とした匿名化関連技術の研究開発にも注力すべきである。
- (2) パーソナルデータを含むデータの解析処理において、人工知能（Artificial Intelligence: AI）技術の利用が益々拡大することが予想されるところ、匿名化関連技術の適用によって解決可能な課題と、匿名化関連技術だけでは解決できない課題があることを踏まえつつ、AIの特性に対応した匿名化関連技術の研究開発が必要であり、例えば以下のような研究課題に取り組むべきである。
 - 匿名化する前のデータをAIに入力した場合の出力結果が、匿名化した後のデータをAIに入力した場合の出力結果とほぼ変わらない性質を満たす匿名化関連技術には、どのような種類があるか、どのような性質が求められるかを特定すること。
 - 将来、AI技術の発達により、匿名化した後のデータから、個人が再識別されたり、個人の属性が推定されたりするリスクが高まる恐れがあるが、このようなリスクを妥当なコストで回避あるいは低減する技術を開発すること。

6.2 事業分野動向の観点（提言 2）

事業分野ごとに、匿名化の対象となるデータの形式や、求められる評価指標（安全性、有用性、処理効率等）は異なる。したがって、事業分野ごとのアプリケーションの特性を調査し、アプリケーション側の要求に適合した匿名化データを得る技術を開発する必要がある。

特に、重点事業領域として以下の事業分野が挙げられる。

- 高い成長と日本企業の優位性が期待される事業分野（「社会インフラ」、「健康管理・予防医学」など）
- 業界横断・産学連携や継続的技術改良で将来高収益が期待される事業分野（「金融」、「流通」、「医療」など）

6.3 評価指標と国際標準化動向の観点（提言 3）

- (1) 事業分野ごとに、匿名化技術の適用例を蓄積し、ビジネスに役立つ評価指標を開発することが必要である。
- (2) 学術的な研究が先行している評価指標は、事業分野特化型指標のプロトタイプとして、あるいは、幅広い分野で汎用的に適用可能な指標として、引き続き開発を進めることが求められる。
- (3) 産官学の連携により、実用的な評価指標を早期に確立し、普及啓発活動を推進することで、ビジネスの活性化につなげることが肝要である。

- (4) 評価指標の開発における我が国の先進的取組を国際標準化につなげるべくイニシアチブをとり、フロントランナーとしてのアドバンテージを獲得するべきである。

7. おわりに

本稿では、平成 29 年度特許出願技術動向調査の 1 テーマである匿名化技術の調査結果の一部を報告した。

本調査が、匿名化関連技術に関わる企業や大学等における研究開発テーマや技術開発の方向性を検討する際の参考の一助となれば幸いである。

謝辞 本調査のアドバイザーボードにおいて、大変有益な議論をしていただいた、菊池浩明委員長、佐久間淳委員、高橋克巳委員、中川裕志委員、美馬正司委員の皆様に、深く感謝申し上げます。

参考文献

- [1] 総務省、平成 29 年版情報通信白書、2017。
- [2] 独立行政法人情報処理推進機構、パーソナル情報保護と IT 技術に関する調査 調査報告書、2012。