



# 人工知能による 専門家の判断のサポート



—現状における人工知能のビジネス応用の実際—

武田秀樹 ((株)UBIC 行動情報科学研究所)

## 人工知能活用の背景

### ①人工知能が専門家の判断をサポート

法律、医療などさまざまな領域において、弁護士などの専門家は、業務上、専門知識に基づく高度な情報探索を行っている。たとえば、法律の分野では、訴訟に関連する証拠の発見、知的財産分野では、技術文献の探索などがそれにあたる。

本稿では、これら専門家の情報探索をサポートする人工知能をテーマに、人工知能のビジネスへの応用の一例として米国関連訴訟における eDiscovery と呼ばれる証拠開示業務を中心に、医療・ウェルネス分野や知的財産業務への適応についても紹介したい。

## 人工知能を適応可能な課題

### ①人工知能が必要とされる背景

ITの革新、普及により、現在はあらゆる場面において情報が巨大化している、いわゆるビッグデータの出現である。企業活動に関しても、社内に蓄積される情報、業務で扱う情報は例外なく巨大化しており、従来の手法では業務の遂行が困難なケースが現れている。その例として挙げられるのが、弁護士や医師、コンサルタントといった各分野の専門家が情報を見て、彼らの知見により判断を下すことが必要とされる領域である。対象となるデータがビッグデータ化することで、合理的な時間内での判断が困難になっているということはもちろん、そもそも巨大化したデータをどのように処理すればよいのかなど、業務環境の変化への対応を迫られている。

これらの領域で、効率的な業務の遂行を可能にする手法の1つとして挙げられるのが、人工知能の活用である。専門家は、専門家自身が経験の中から培った知見を用いて、目的に応じた判断を下している。そこで、その専門家の知見を人工知能に学習させることができれば、微妙な情報の違いを人工知能が専門家の観点で識別し、大量データの分析、判断の自動化が可能となる。

また人工知能は当然のことながら疲労や、集中力の低下といった人間では逃れることのできない生理的現象とは無関係のため、継続的に正確な判断を行うことが可能である。つまり、人間ではまず不可能なビッグデータの分析も、人工知能を活用することで、ただ分析が可能になるというだけでなく、効率的かつ高い精度で行うことが可能になる。

たとえば、過去に蓄積された医療データから有用な情報を抽出することによって、医師が最適なケアを判断するためのサポートを行う医療応用システム(Watson)を、IBMが開発していることは周知のとおりである。このように、人間をはるかに超える速度で大量のデータをミスなく延々と精査できるというコンピュータの特性を、「専門家の知見を活用する」という方向に活用する動きとして重要なものに、1980～90年代にかけて開発が進んだエキスパートシステムがある。このエキスパートシステムでは、専門家の知見をある種の形式知化させることで、専門家の判断をコンピュータに代替させていたが、そこには限界があった。当然のことながら、すべての知見、知識を形式知化することは困難なためである。特に専門家の判断に重要な要素は、感覚といった説

明の難しい、いわゆる暗黙知であるケースも少なくない。この暗黙知の活用こそ、エキスパートシステムでは超えられなかった、人工知能による汎用的な判断を実現する上での鍵と言える。もちろんエキスパートシステムが実現した形式知の活用も無視することはできない。現実としては、目的に応じて、暗黙知と形式知の活用のバランスをはかり、アプリケーション化することが肝要と考える。

すでに、人工知能を導入することで大幅なコスト削減、作業の精度向上を実現している分野がある。国際訴訟支援における証拠開示 (Discovery) への対応である。Discovery とはアメリカ連邦民事訴訟規則で規定された手続きの1つで、裁判の審理が始まる前に被告、原告の双方が訴訟に関する証拠を提出しあい、論点の整理を行うというものである。多くの訴訟は、この段階で賠償金の金額や、訴えの取り下げが決まり、和解となる。それゆえに Discovery は審理同様に重要な手続きの1つで、この作業の成否が訴訟に及ぼす影響は計り知れない。またアメリカ独自の手続きではあるが、グローバルでビジネスを行う企業にとって、米国での訴訟リスクは無視することができないため、この手続きへの理解は必須と言える。実際に多くの日本企業が Discovery 作業において苦労しているという現実もある。

Discovery の中心的な作業として、弁護士を中心に行う証拠の選定がある。これまでは紙資料を対象に行っていた作業を、ビジネスのデジタル化が進んだ現在、ローカルの PC やサーバに蓄積された電子メールや各種ビジネス文書といった大量の非構造化データに対して行っている。そのため、証拠開示 (Discovery) は電子証拠開示 (eDiscovery) と呼ばれるようになった。

## 国際訴訟における人工知能の活用について

### ●技術動向

eDiscovery の工程を規定した標準ワークフローとして EDRM (Electric Discovery Reference Model) があ

る。その中の1つに、大量の電子データが訴訟に関連するかどうかを判断する閲覧 (Review) という工程がある。この Review において Predictive Coding<sup>☆1</sup> と呼ばれる機械学習応用技術が多く用いられている。従来の Review では、企業内に蓄積された何十万、何百万といった数のファイルを弁護士が1通1通目視し、訴訟に関係するかどうかを確認していた。大きな事案では何十億円という費用が発生する eDiscovery の中で、70～80% がこの Review にかかる費用と言われており、その費用の削減、時間の短縮により訴訟を有利に進めたいというニーズから、eDiscovery では機械学習を応用した技術の開発が進んできた。

Predictive Coding のメカニズム (図-1) は、一般に教師あり学習と呼ばれる種類のものである。具体例を挙げると、以下の3ステップにより説明される。

1. 調査：弁護士などの専門家がサンプルを仕分けする、教師データの作成段階。
2. 学習：人工知能が教師データから専門家の判断基準を学ぶ入力段階。
3. スコア付け：学んだ判断軸に則り、人工知能が未知のデータに対してスコア付けを行う出力段階。

UBIC の人工知能を例に挙げると、スコア付けアルゴリズムでは、スコア付けの基礎量として情報間の相互依存性を定量的に評価する「伝達情報量」を採用している。このアルゴリズムには教師データとして、訴訟に関係のある (Relevant) データだけでなく、訴訟に関係のない (Not Relevant) データも学習の対象とするという特徴がある。この学習方法は再現率 (Recall Rate) を高めるのに適しており、証拠となるデータを網羅的に特定する必要がある訴訟事案において、その有効性が確認されている。また、伝達情報量の特性から報酬の高低のみから出力ポリシーを最適化することが可能なため、本アルゴリズムは広義の強化学習と呼ぶこともできる。加えてその基本アルゴリズムのほかに、

☆1 Predictive Coding：仕分け対象となるデータ (ドキュメント、電子メールなど) のうち、一部のサンプル・データを人間が仕分けを行い、その結果に基づきコンピュータが残りのドキュメントを仕分けするという手法。

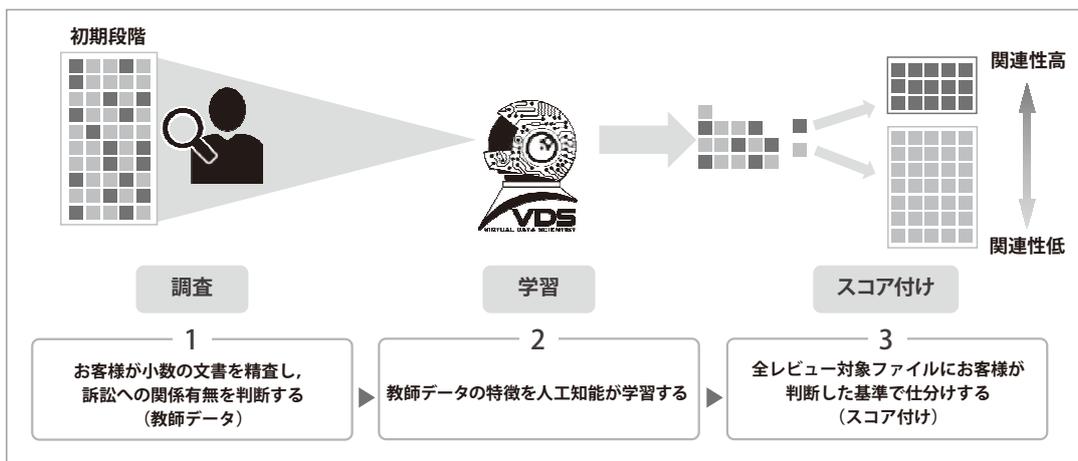


図-1 Predictive Codingの仕組み

1. スコア付け判定の自動的かつ段階的な最適化を行うウェイト・リファインメント。
2. 業務の中でスコア結果に対するフィードバックを繰り返し学び精度を高めるプログレッシブ・アナライザー。

などの技術を用いることで、eDiscoveryでの情報抽出において求められる、網羅性の最大化を目指している。

以上のように、eDiscoveryの現場では大量のデータを高速に解析する必要があることから、人工知能の開発、実用化が進められてきた。今回取り上げたのはUBICの事例であるが、その他Equivio, Recommind, Content Analyst Companyといった企業も機械学習応用技術を実用化している。Recommindが発行している『A RECOMMIND WHITE PAPER』<sup>1)</sup>に記載されているように、彼らの中にはベイズ法やサポートベクタマシンなどのアルゴリズムを採用している企業もあり、今後もさまざまな方式の人工知能が国際訴訟の現場において磨き上げられることが予想される。中には用いているアルゴリズムを開示していない企業もあるため、十分に情報を揃えた上での詳細な比較は行えないものの、今後も人工知能は実ビジネスの世界での問題解決に役立てられると考えられる。

### ●司法側の動向

司法の現場でも、徐々にではあるがPredictive Codingをはじめとした機械学習応用技術への理解

が進んでいる。これは主に年々増加するデータ量に応じて人力でのデータ処理、特にReview作業が行えなくなってきていることが原因に挙げられる。

具体的には、2012年にAndrew Peck判事が初めて訴訟におけるPredictive Codingの使用を許可し、その中でPredictive Codingを「膨大なデータ量を扱う事案においては非常に有効な手段」とであると述べている。また続く2013年には米国司法省から、Predictive Codingの有効性を述べ、その活用を推奨する『IP, Antitrust and Looking Back on the Last Four Years\_20130208』<sup>2)</sup>や『HEARING ON OVERSIGHT OF THE ANTITRUST ENFORCEMENT AGENCIES\_20131115』<sup>3)</sup>のような文書も発行されており、これ以降はUBICをはじめとする企業がその技術を実案件において有効活用している事実からも、訴訟の現場において機械学習応用技術はなくてはならない問題解決の手段として認識されていると考えることができる。

### ●日本の訴訟における国際競争力

米国訴訟において、日本企業は大きなディスアドバンテージを背負っていると言える。冒頭でも述べたとおり、そもそもeDiscoveryとは、米国連邦民事訴訟規則で規定された、証拠開示に関する手続のことだが、この制度は訴えた側と訴えられた側が直接証拠を開示しあうという制度である。証拠をお互いに開示しあうため非常に公平性の高い、両者に平等な制度であるはずなのである。しかしながら、

## 米国訴訟案件における実績比較

実際の案件で、UBICのPredictive Coding技術によるレビューは、従来手法によるレビューと比較し、3倍のスピード、1/5の費用で完結できることを実証した。

	総ドキュメント数	総作業時間	費用 (USD)
UBICによるレビュー	174,041	1,941	200,000
従来手法によるレビュー	91,609	3,191	550,000

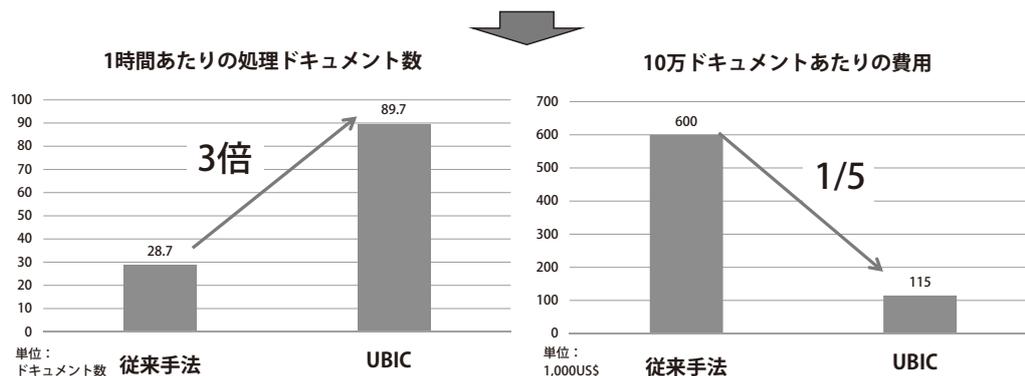


図-2 実際の訴訟案件における、Predictive Coding技術によるレビューと従来手法によるレビューの比較

1. 本国ではないアメリカの規定に則って手続を進めなくてはならない。
2. 訴訟の現場であるアメリカでは用いられることのないアジア言語で記載された文書を多く有している。

このような背景を持つ日本企業は、この eDiscovery への対応のために多くの時間とコストをかけなければならない。さらにアジア言語の解析が可能なベンダがほとんどいないといった問題も合わさり、日本企業の多くは訴訟に対応する体力が尽きてしまったり、もしくは eDiscovery への不適切な対応により、不利な条件での和解を余儀なくされたりすることで、多額の課徴金や制裁金を支払うケースが多々発生している。このような現状を背景に、日本を中心とするアジア企業においては、特に強く機械学習応用技術への期待が高まっていると言える。

## 国際訴訟での人工知能の活用事例

### ●活用事例

この章では、実際の活用事例を通し、人工知能の導入効果を紹介したい。すでに述べた通り、人工知能は、国際訴訟の eDiscovery におけるレビューの工

程で活用され、作業時間の大幅な削減と品質の向上に貢献している。実際の活用方法だが、人工知能は弁護士が「関連あり」「関連なし」で仕分けた一部の文書群を教師データとして学習し、膨大な文書にスコア付けを行う。結果として、「関連あり」と仕分けされた文書の特徴と近い特徴を持った文書に高スコアが付与され、弁護士は、高スコアの文書を優先的にレビューすることで効率的かつ効果的に法廷に提出すべき文書に辿り着くことができる。たとえば、実際に UBIC の人工知能が活用された案件では、人工知能を使用しない従来手法による作業と比較し、レビュースピードで3倍以上、コストは5分の1となり、人工知能の有用性を証明している (図-2)。

また、この人工知能は、カルテルや情報漏えいなどの予防的処置として、Eメール監査システムの中でも活用されている。従来のEメール監査では、監査担当者がキーワード検索などを行いながら、日々、蓄積するEメールを1件ずつ監査していた。キーワード検索による監査は、監査対象を広げるためにキーワードの種類を増加させるとゴミヒットも増加し、監査対象を絞り込むためにキーワードの種類を限定すると取りこぼしが生じるといった非効率

かつ非効果的なものであった。しかし、人工知能を活用することで、少量の教師データでスコア付けを実施し、高スコアのEメールのみを監査することで作業が完結するようになった。そのため、監査担当者は、キーワードの設定や目視による監査作業に忙殺されることなく、本来、エキスパートが集中すべき仕事に集中できるといったメリットを享受することとなった。たとえば、人工知能監査を取り入れたA社では、2カ月間かけて行っていた監査工程を4日間に短縮することで社員の負担を大幅に減らすことに成功している(図-3)。

### ●方式

UBICを始めとして、米国における国際訴訟向けのプラットフォームの提供形態は、クラウドが主流である。その理由は、以下の通りである。

1. その国の法がおよぶ地理上のデータセンタで処理を行う必要がある。そのためUBICは、日本以外にも、米国、韓国、台湾、EU(イギリス)のデータセンタにクラウド環境を保有している。
2. クライアントの機密情報を取り扱うため、クライアントの環境はネットワークレベルで分かれている必要がある。UBICは、独自のクラウド基盤であるIntelligence Cloudを構築し、クライアントの重要な機密情報をセキュアなネットワーク環境で運用している。
3. 日本・アジア各国の法務・知財担当者、米国弁護士がデータを閲覧する必要があるため、外部からのアクセスが必要である。たとえば、UBICの場合、外部のインターネット環境から2重ログイン(シンクライアントCitrix+UBIC解析プラットフォームLit i View)で管理している。

## 人工知能によるEメール監査の事例

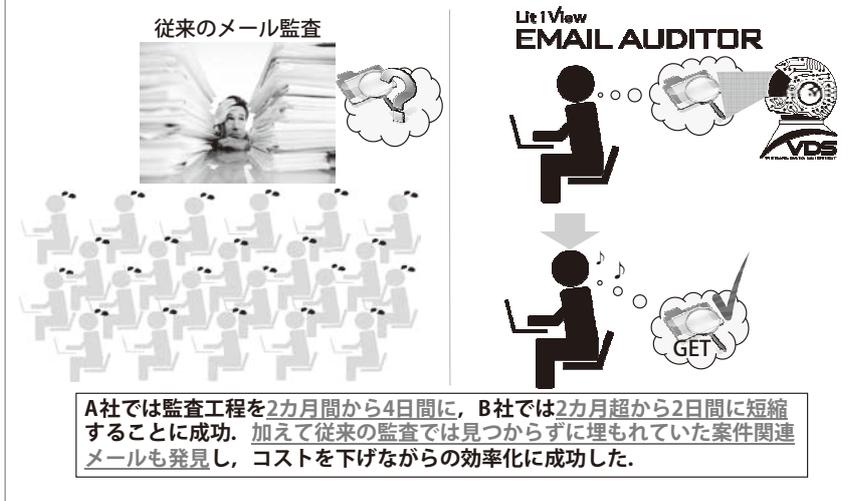


図-3 従来手法によるメール監査と人工知能を活用したメール監査の比較

## その他領域での人工知能の応用拡大

人工知能による専門家の判断サポートは、他の領域でも進みつつある。ここでは、医療・ウェルネス分野、知的財産分野への取り組みを紹介する。

### ●医療・ウェルネス分野

近年患者の高齢化が進む中、病院内での入院患者の転倒転落事故をいかに低減するかが大きな課題となっている。転倒転落事故の原因となる要素は多岐に渡る。たとえば睡眠薬の投与、運動機能の低下、認知症、せん妄状態を含む意識障害・注意力の低下などである。従来から病院も、さまざまな転倒転落防止策を行っている。具体的には、ベッドに備え付けられたセンサによる転倒転落の把握、拘束具、また、転倒転落の可能性が高い患者を把握するための、看護スタッフによるアセスメントスケールの運用である。身体の状態やこのような対応策は一定の効果을上げながらも、看護側に大きな間接工数を強いるという課題がなお残されている。また、その工数の増大ゆえに、現実的には入院時や手術時など、限定的なタイミングでしかアセスメントができず、よりタイムリーに、効率よく、かつ効果の高い転倒転落防止対策の存在が待たれている。

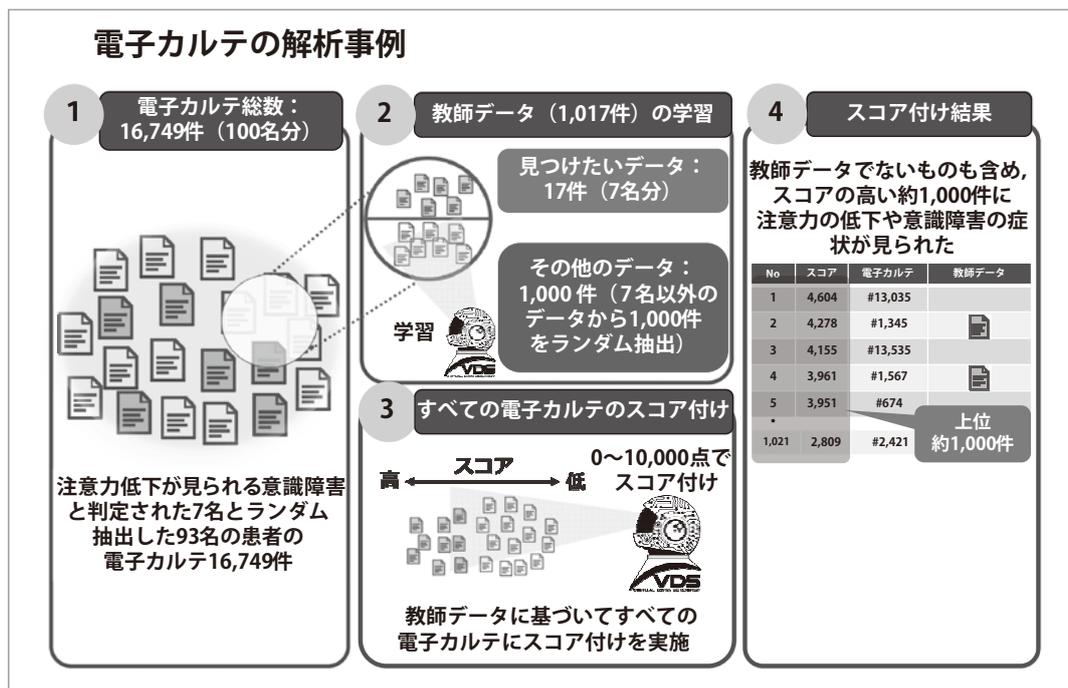


図-4 転倒転落防止を目的とした、電子カルテの解析事例

NTT 東日本関東病院と UBIC が行った、転倒転落防止を目的とした電子カルテの解析実験は、このような状況の打開に挑戦したものである。先に、アセスメントのタイミングが入院時や手術時に限定されているという課題を挙げたが、それでは、もっと高い頻度で看護状況が更新されている既存の情報源はあるだろうか？ ある。それは電子カルテに記載された看護記録である。看護記録は、看護側の観察が毎日更新されている。この電子カルテ上の記載について、転倒転落の原因となる要因を抽出可能かどうか、探索的実験 (図-4) を実施した。実験は、無作為に抽出された、約 17,000 件の電子カルテ (1 件が、1 名の患者の 1 日分のカルテとなるよう加工) の中で、いわゆる SOAP<sup>☆2</sup> と呼ばれる電子カルテ記載様式の中の S (Subjective: 主観的情報) と O (Objective: 客観的情報) に該当するデータに対して、看護師が評価を行い、注意力の低下やせん妄状態が見られるカルテを 17 件抽出した。また、抽出したくないデータのサンプルとしてランダムに抽出

☆2 SOAP: 電子カルテ記載様式の 1 つ。「S (Subjective)」患者の話や病歴など主観的なデータ。「O (Objective)」診察や検査から得られた情報。「A (Assessment)」S と O を評価したもの。「P (Plan)」S、O、A をもとにした治療方針、これら 4 つの要素で構成される。

した 1,000 件の電子カルテ情報を併せた 1,017 件の電子カルテ情報を教師データとして、人工知能に学習させ、元の母集団約 17,000 件のデータに対して、スコア付けを行った。その結果、スコアの高い上位 1,000 件、とりわけ上位 200 件弱のカルテには、その多くに注意力低下やせん妄の症状が見られた。今後、予測ロジックの確定と、大規模データによる実証実験を行い、最終的にはスマートフォンのようなデジタルデバイスから、看護スタッフが転倒転落の可能性の高い患者を把握することができるアプリケーションの開発を行う予定である。アプリケーションによって把握された転倒転落可能性が高い患者を集中的にケアすることにより、より現実的で対応可能なケアの実現を目指すことが可能となる。

#### ●知的財産分野

世界の特許出願件数は、2012 年には 234.7 万件を超え<sup>4)</sup>、情報は増加の途をたどっている。この事実は、先行技術調査や無効審判調査のような知財業務で調査対象となる特許情報の数も膨大になっていることを同時に表している。従来からの特許調査のやり方は、技術分類やキーワードでのブーリアン検索が主な手段として使われているが、このような検索

手法では、事前に技術分野やキーワードのバリエーションを予測して検索式を構築する必要があるため、高度な専門性と経験が要求される。

トヨタテクニカルディベロップメントとUBICが行った、人工知能による特許文献抽出実験は、知的財産業務における特許文献調査について、大量の文献の中から人工知能によって、より効率的で簡易な特許文献の探索を目指した実験である。実験は、出願前に行われる先行技術文献の調査を想定して行われた。実験は複数の特許について行われているが、1つの例を紹介すると、約8,000件の特許文献の集合の中から、どのくらい早くエンジン制御技術に関する無効資料が抽出できるかを試みた。その結果、最大250倍の効率上の優位性が見られた。現在より多くの特許文献の抽出を行う運用実験を行い、検証が進められている。

## 人工知能のビジネス応用の 現状とこれから

以上の事例のように、人工知能による専門家の判断のサポートは、ビッグデータ化した情報の中からでも必要な情報を探索し、専門家がそれぞれの業務を滞りなく実行することを可能とする。それに加えて、業務の高速化、品質の向上にも貢献することができる。

人工知能を活用することは、企業に対してどの

ような価値を生むだろうか？たとえば、国際訴訟分野において人工知能を活用するということは、eDiscoveryのコストを下げ、日本企業のように、言語や法制度の違いで元々不利な条件を持っている国の企業でも、欧米の企業と訴訟で対等に戦えるための条件を整えることができる。また、転倒転落防止や、特許文献検索業務でも、エキスパートの生産性向上に寄与することが可能である。

人工知能を使ったソリューションの広がりには、数多くの可能性を秘めている。元々日本は人工知能の研究者も豊富な土壌がある。産官学連携なども通じて、新しいソリューションを作り、日本発の人工知能ソリューションが多くの貢献を世界に行っていくことを祈って筆を置きたい。

### 参考文献

- 1) A RECOMMIND WHITE PAPER, <http://www.aiim.org/pdfdocuments/36879.pdf>
- 2) IP, Antitrust and Looking Back on the Last Four Years, <http://www.justice.gov/atr/public/speeches/292573.pdf>
- 3) HEARING ON OVERSIGHT OF THE ANTITRUST ENFORCEMENT AGENCIES, <http://docs.house.gov/meetings/JU/JU05/20131115/101486/HHRG-113-JU05-Wstate-BearW-20131115.pdf>
- 4) 特許行政年次報告書 2014年版, <https://www.jpo.go.jp/shiryoutoushin/nenji/nenpou2014/honpen/1-1.pdf>  
(2015年8月1日受付)

武田秀樹 (正会員) [hideki\\_takeda@ubic.co.jp](mailto:hideki_takeda@ubic.co.jp)

UBIC 行動情報科学研究所所長。1996年早稲田大学卒業、専攻は哲学。2013年UBICのCTOに就任。Lit i Viewのプラットフォームの開発、人工知能「Virtual Data Scientist」開発の指揮を執る。