

(2016年10月19日ご講演)

FRONTEO の人工知能 KIBIT とそのビジネス活用

株式会社 FRONTEO

取締役最高技術責任者 行動情報科学研究所所長 武田秀樹委員

私は、この15年ほど自然言語処理を使ったアプリケーション開発をやってきた。以前は特許分析を中心に、技術価値を自然言語で書かれた技術文書の中からどのように自動的に見いだすのか、資産価値としての特許技術をどのように算出するか、また、そのような技術価値を、技術を保持している企業の財務諸表とどのように結び付けて解析するのかをやっていた。FRONTEO もリーガルテクノロジーという大量のテキストデータを扱う分野で事業を行っている会社である。入社後は、メールやビジネス文書の中でどのように証拠の発見を行うかという技術の開発を手掛けるところからスタートした。

会社概要

まず会社の紹介をしたい。7月に社名を UBIC から FRONTEO に変更した。また、我々は、日本では東証、アメリカでは NASDAQ に上場している。設立以来、国際訴訟における証拠の発見という分野でずっとサービス展開をしてきた。その中で自然言語処理、機械学習を使ったエンジンの研究開発に取り組み、そこで開発したものを使ってビジネスの領域を拡大している。

最近ではビジネスインテリジェンス、マーケティング、ヘルスケアといった分野にも力を入れており、従来のリーガル分野と併せて活動領域を広げつつある。

主要取引先としては、もともと国際訴訟支援をやっていたということもあり、海外を中心にビジネスをされている製造業のほか、法執行機関、また法律事務所や、最近では金融機関が我々のお客様になっている。

我々は KIBIT というテキストの解析を行うエンジンを展開している。このエンジンの名前は人間の心の機微という”KIBI”と、情報の bit というところから、人間の心の機微を理解したいという意気込みを込めて名付けている。

我々が長く手掛けてきた、国際訴訟や不正の調査で証拠を発見する業務は、失敗の許されない分野の一つである。例えば国際訴訟における証拠発見であれば、期限までに弁護士同士が交渉をし、合意したデータを相手方に提出する、というレギュレーションがアメリカにはあるが、そこで必要なデータが仮に出なかったとすれば、何か隠しているのだろうということになり、訴訟上不利になってしまう。また、出す必要がないデータを出してしまうのは、相手方に見せる必要がないものを見せることになってしまい、ビジネス上の不利になりかねない。きちんと定められた範囲の情報をきちんと期限内に相手方に出すことが重要なわけである。このような分野での証拠発見を通じて技術を作ってきた。

AI を取り巻く状況認識

次に我々が考える AI を取り巻く状況認識を改めて説明させてもらいたい。技術としての AI はデータに基づく最適化プロセスである。その最適化プロセスを学習と呼ぶわけであるが、これはデータとそのデータを通じて予測したい結果をどのように出力するのか、最適化の機構を備えた情報処理の総体、一般的なものを指しており、広い概念である。ただ、現在の情報処理技術の中で AI 技術は、真の意味で人間の知能をコンピューターで再現するところまで進んでいない、しかし、特定の分野の中で人間の判断を効率的に行う情報処理が、社会実装として進んできているということが基本的な認識としてある。そのような状況の中で最近では AI をかなり広い事業領域、目的で使っている傾向があると考えている。

今の AI を取り巻く背景の中でやはり重要だと考えているのは、2000 年代前半以降、クラウドという言葉に象徴されるように、ハードウェアの性能が非常に上がり、コストが下

がり、インターネットを通じてシステムリソースを安価に簡易に利用できる環境が整ってきたこと、またインターネットの普及に伴ってデータ量が非常に多くなってきたことである。今までの皆さんの発表の中でも同様のご指摘があったが、我々も、ここが重要なのではないかと考えている。もちろん AI の応用技術が洗練されたという側面もあるが。また、データが増えることによって、ビッグデータの中から必要な情報を取り出したいというニーズが顕在化されてきたことが今のブームを盛り上げている背景になっていると認識している。

－Google Deep Learning

AI 技術の現状を見てみたい。例えば、Deep Learning の事例として有名な Google の猫の話であるが、この解析を行うには、1,000 万件の画像を 1,000 台のコンピューターに入力して 3 日処理をする必要があった。今 AI が注目を浴びているが、ビッグデータの中で情報を見つけたいというニーズがある反面、コンピューテーション能力や学習のためのデータ量を必要とする技術であるという現実もあると考えている。

－AI サービスは増加の一途、しかし…

FIRSTMARK のマップで AI に関連する企業の技術を見てみると、様々な形の技術がある。そして、実際にこれら AI の技術をユーザー企業が使おうとした場合には、何らかのシステムインテグレーションが必要になると考えられる。また、処理するために非常に多くのシステムリソースが要求され、且つ、データの量も必要とされるので、その点でもインテグレーションの際に手間が掛かってしまう。

－AI の導入コスト、実装期間

こういった現状を鑑みると、AI を実際にユーザー企業が使おうとした際の課題として大きいのは、導入時のコストである。いろいろな AI の種類があるが、一般的に数千万から数億円規模の導入コストが掛かると見積もることができる。そのシステムを導入する際にも、

例えばその技術の特殊性から Deep Learning だったら Deep Learning 自体の設計を専門家が行う必要がある。Deep Learning にしても、何でもデータを入れたら勝手に最適解ができてくるような技術ではないので、ノードの設計や、パラメーターチューニングなど、様々なインテグレーションを専門家が行う必要があるため、コストが掛かる。

一 大企業を中心に拡大する AI 活用

経産省資料「攻めの IT 投資経営銘柄 2016」で、AI に対して積極的な取り組みをしている代表的な企業を見てみると、みずほフィナンシャルグループ、日産自動車、三菱商事、JR 東日本、ブリヂストンなど、非常に大きな企業が多い。売り上げのレベルで見てもかなり大きい。こういった企業が中心になって AI の取り組みを行っているのが現状かと思う。

こういった現状から更に進んで、中小のユーザー企業にも AI 技術をもっと普及させて、AI 技術が一般的に使われる状況になることを目指した場合、まだまだそこに壁があると思う。コスト、膨大なデータ、導入の手間といった課題を超えないと、本当に AI と人間が融合して課題を解決していくような社会の実現にたどり着くには、長い時間が掛かると思う。

AI、どう使う？

さて、ここまで我々が持っている課題意識を共有させてもらったが、次に、AI と言ってもいろいろな種類のものがあるので、我々がどういう分野で、どのような課題を解決するために AI を使おうと思っているのかを紹介する。

一 生産性の向上

我々が取り組むべき対象課題と考える一つ目は、ホワイトカラーやエキスパートの生産性向上である。エキスパートとして、最初は弁護士を対象にしていたわけであるが、そこからヘルスケアの分野では医師や看護師、情報分析の分野であればアナリストとターゲットを広げている。OECD 加盟国の労働生産性ランキングを見ると、人口 1 億人以上の国は

日本とアメリカしかないが、アメリカが4位、日本は21位で、日本は生産性が高いとは言えない状況がある。この中でどうやったら生産性を上げられるのかが、取り組むべき大きな課題と考えている。

一健康な社会

もう一つ大きな課題は、健康に関する課題だと思っている。日本は2060年には2.5人に1人が高齢者という超高齢化社会に突入する。その中でいかに健康な社会を実現するのか、例えば労働力の問題も、このような健康な社会と密接に関係している。健康でずっと働ける社会が実現できれば、労働力の確保にもつながる。そのために早い段階で病気を見つけて治療する必要があるし、悪くなる前に未然に防ぐ、未病のような対応も有効である。

一ダウンサイジング

大きくこの2つの領域を課題としてAIの活用に取り組んでいるが、これらの課題に取り組む際に先に述べた課題を解決しつつ、取り組む必要がある。そこで意識してやっていることがダウンサイジングである。AIをやるには現在コストも時間も手間も非常に掛かる。このような状況の中、大きなシステムを必ずしも必要としない、ユーザー企業が手軽に使えるようなAIのシステムを作って流通させることで、恩恵をさらに広く行き渡らせる取り組みが必要であると考えている。

一材料（パーツ）だけでは…

先ほどAIのシステムを導入するためには、まだまだ高コストのインテグレーションが必要だという話をしたが、インテグレーションが必要ということは、AIの技術がまだパーツになっている状態だと思う。実際にユーザー企業が、現場でAIが搭載されたアプリケーションを使おうと思った場合には、そのビジネスドメインの中でどのような業務をどのようなレベルで代替・サポートするのかをきちんと分かった上で、最終的なアプリケーションの姿に落ちている必要があると思う。

ーKIBIT 搭載のアプリケーションー

そのような課題意識を持って我々はAIを搭載されたアプリケーションを作ることに取り組んでいる。リーガルの分野であれば、eディスカバリ（後述）の業務ソフトウェアを作って、その中で弁護士の判断力を学習するような機能を取り入れる。また、特許の解析であれば特許を解析するためのワークフローを備えたアプリケーションを作る。このように、業務にすぐに使えるソフトウェアをどんどんアプリケーションとして作って、それを提供することによってAIの恩恵をすぐに使える状態にすることを意識している。

ーいま注目される「Deep Learning」とデータの現実

今、アルファ碁などに代表されるようなDeep Learningが社会的に注目されているが、現実の課題はゲームだけでなく多岐にわたる。その課題を解決する際に、ニューラルネットワーク系のアルゴリズムであれば非常に多くのデータが必要とされるが、ユーザー企業がそのような大量のデータを持っていないケースが多々ある。また、データの絶対量だけではなく、ラベリングされた教師にできるデータを必ずしも多く持っていないというケースもある。それほどたくさんデータを持っていない、またAI教育用の教師データを作る手間をあまり掛けたくないといった課題は、お客様との相談の中でも多く出てくるテーマである。

また、我々が主に対象にしているモダリティーはテキストデータであるが、このモダリティーは、画像や音声に比べてDeep Learningに向いていないことがある。テキストデータは非常に多義性があるモダリティーであるし、曖昧性がある。文脈の中でどう扱われるのかによって同じ言葉でも意味の解釈が変わってくる難しさがある。大規模なデータに対して離散的な学習を行い、そのまま実際のビジネスシーンに使えるかという点、現実問題なかなか難しいところがある。

一実用的なアルゴリズム

そこで、少ないデータであっても効果を出すことができる仕組みを作れないかということとで独自に開発したのが、我々が **Landscaping** と呼んでいるアルゴリズムである。この特徴は、まずテキストに特化しているということが挙げられる。それから、少量の学習で効果が上がるものを目指した。例えば訴訟の現場で使おうと思った場合、弁護士がまずお手本として実際の判断を少しやってみせる。そして、その判断を学習して、対象となる数百万件、数千万件というデータに対して判断をするわけである。このような使用をする場合、最初にデータが少なくても学習ができないと結局たくさんのデータを人間が示さなければならぬことになり、コストの削減、効率化につながらない。従って、少ない手本の中で行ったその人の判断の特徴をうまく学習して、その背後にある大量のデータに対して先行して判断がかけられるということが重要だったわけである。そのような理由から、非常に大量のデータを要求するアルゴリズムではなく、少ないデータの中で人間の判断をより多く取り入れる形で学習を行えるアルゴリズムの開発を行った。ただし、その人間が間違っただけの教え方をすると判断も間違ってしまうという一面もあるので、きちんとした特徴を与える必要があるということが前提になってくる。

また、学習の際にはビッグデータが必要であると言われるが、そもそも教師データに求められるのは、学習させたい判断に対して、きちんとその判断を機械で再現できるに足りる特徴をデータの中から拾えるかどうか重要であって、最終的には学習の質の問題が重要である。ビッグデータが必要とされるのは、質の問題を担保するために、ビッグデータであれば、巨大であるがゆえに、その中に再現したい判断を学習するに足りる情報が入っているだろうという期待があるわけである。それが我々のビッグデータの有効性の解釈であるが、人間が効率的に正解データ、質の良い学習データを与えてあげることによっても、もっと少ないデータであっても判断ができるということもあると考えている。もちろん判断のバリエーションを担保するだけのデータ量は必要であるが、我々はどちらかということと人間の判断と協調させることによって、少ないデータでも効力を出せる方向を目指した研究開発を行っている。

FRONTEO における人工知能開発の背景とビジネス活用

こういった特徴を持ったエンジンを使って、テキストデータの解析という切り口で、さまざまなシーンでビジネス展開をしているが、次に我々はどのようなシーンで自分たちのエンジンを使ったビジネスをやっているのかを紹介していきたいと思う。

一国際訴訟支援 eDiscovery

まず、我々が長く手掛けてきた国際訴訟支援、e ディスカバリと呼ばれる分野の事例である。この e ディスカバリの分野というのは、非常に短い期間の中で、大量のデータの中から証拠となるデータを抽出して相手に正確に渡す必要がある。日本ではあまりなじみが無い制度であるが、アメリカでは多くの弁護士がこのような訴訟の際の証拠を見つける業務に従事している。このような業務に対して訴訟の際ユーザー企業が払っている額は、かなり大きなものになる。

では、その e ディスカバリの中で、どのような工程に特にコストが掛かっているのかというと、レビューと言われる工程である。これは平たく言うと弁護士が証拠を閲覧する作業である。大きなプロジェクトであれば、大きな執務室の中で、閲覧ソフトの画面を皆で開き、数百人の弁護士たちが証拠を長い時間見ていくというシーンも見られる。

守本正宏委員（FRONTEO 代表取締役社長）より補足

メールやプレゼンテーション資料など、基本的には全部デジタル化されたドキュメントを目で見る。

武田委員

従来の技術である検索もいまだに利用されている。この分野では、検索のキーワードの数が 1,000、2,000 という規模と大きい。この検索で絞り切れない文書をたくさん人間が見るというわけである。

では、なぜ検索で絞りきれないのか。具体的な例として、談合に関連した実際のメールと何の問題もないメールの2通を比較し、理由を明らかにしてみる。

メール①

「今日の夜、予定ありますか？久々に飲みにも行けないかと思ひまして、駅前の居酒屋に8時くらいでどうですか？」

メール②

「もし良ければ、今日にでも飲みに行きませんか？前回から時間も経っていますし、またお話できればと思います。良い個室の居酒屋を見つけたので、そこにしましょう。〇〇さんも誘った方がいいですかね。」

メール②が不正示唆メールである。これは「前回から」という部分で、定期的にずっと会っているのだなと分かる。また「個室の居酒屋」ということで、人に聞かれたくない話をしているのだなということも分かる。さらに非常に多くの固有名詞、登場人物が上がっているため、その登場人物間に特有のコンテキストがある、といった具合である。このような観点を重ねていくと、かなり証拠が絞れていく。ところが、これを検索で見つけようとすると、このような飲み会が談合の証拠の一つだと分かっている弁護士が、「飲み」というキーワードを最初に思い浮かべて検索する。また、他のキーワードも考えていき、1,000、2,000 と考える。しかし、前述したような微妙な特徴の違いは、検索だけでは拾い切れず、多くのメールが検索結果として上がってきってしまうので、時間を掛けて読み解いていく必要があるというわけである。結果的に、かなりコストが掛かってしまう。

守本委員

例えば「個室」とか「時間」とか、そういうものをキーワードとして選んでしまうと、「個室」というキーワードもほかのメールで当然使っていることがあるので、どうしてもよい「個室」が入ったデータが出てくる。リーガルテクノロジー（リーガルテック）というアメリカを中心に発展している産業があり、我々以外にも1,000社以上ある。他の企業もしのぎを削って文書の絞り込みをやろうとするが、結局どれほど頑張っても絞り切れない。いろいろな組み合わせがありすぎる。その結果、最終的に人が見るしかなくなる。

武田委員

絞り込めない状態で、前述のような調査をすると、訴訟の当事者である企業が負担するコストは大きなものになってしまう。

守本委員

先ほどの 2 つの文書の違いであるが、その違いを一瞬で分かってしまう人がいる。その人はなぜ分かるのだろうか、という観点からそのメカニズムを明らかにしようとしても、その人は、なぜ分かるかをあまり言わないし、もっといえば分からない。何故かは分からないが、経験から来る勘で分かるらしい。その勘で分かるものをうまく技術的に再現できないかというアプローチで使ったのが、我々の手法になる。

武田委員

リーガルの分野にこのような課題があり、我々は実際に弁護士が判断したデータを多く取り扱うサービスを行ってきた。そのような状況の中で、弁護士の判断が、どういったらある程度の汎用性を持って再現できるのだろうかということに取り組んでアルゴリズムを作ってきたという経緯がある。

ー Predictive Coding を使った事例

業務の中で、どのように KIBIT を利用しているのか、もう少しお話をしたい。リーガルテックの中で、機械を使ってレビューを行うことを、Predictive Coding と呼ぶが、この Predictive Coding に KIBIT を使っている。Predictive Coding をどのようにレビューの行程の中で使っているかであるが、1 次レビューの行程を機械にサポートさせるということをやっている。正確に言うと、0.5 次レビューのようなものを少数の人間がやって、そこでコンテキストを教えた上で機械がまだレビューされていないデータに対しての判断を施す。そしてその後、また人間が確認するというをやっている。これは、まだ機械学習の技術が浸透していく過渡期ということもあると思うが、機械だけに全部任せて、レギュレー

ションとして問題がない、ということには、まだなっていない。しかし、アメリカでは司法省が、大量のデータの場合はこういった機械学習の技術を使って判断することをやってもよい、と技術の利用を後押しするようなコメントを出している。ただ、そのような先進的な姿勢を持つアメリカですら、人間もやはり入れたほうがよいというところは残している。だから、どうしても人間がボトルネックになる現状はまだある。それでもコストは 5 分の 1 ぐらいになる事例もある。もともとかなり大きな金額が掛かっているということもあるが、このコスト削減の効果は非常に大きい。

守本委員

案件ごとにトレーニングをするが、データ量が少なくても学習できるので、短い時間ですぐ使えるようになる。明日証拠を出さなければいけない、というような世界で作った技術で、1 カ月学ばなければいけないようなものではない。

武田委員

中には共通して他のケースで使えるモデルもある。同じ種類の不正であれば使える場合もある。

一訴訟支援における活用「Review Heat Map」

もう一つ、この技術で定量化されるようになったのが、弁護士の判断の質である。我々のソフトの中に Review Heat Map と呼んでいる機能がある。これはどのようなことをやっているかという、どの文書に、弁護士がどのような評価をしているのか確認し、その結果を KIBIT によるスコアと比較することで、作業品質の確認を行う。縦軸に弁護士の名前が並び、横軸には KIBIT を使ってスコアリングしたスコアが出る。スコアが高いということは、その案件の中で証拠となる可能性が高いことを指している。右のほうに行くにつれて、Heat Map の中でも赤い色になっていくので、大体の方が右側のほうに行くにつれて赤い文章が並ぶようになっている。ところが、右側の方が赤くなっていない色分布になって

いる方は、スコアが高い文書を証拠ではないとみなしているということである。また、この反対の色分布になっている場合は、スコアの低い文書を証拠であるとみなしてしまっていることを指している。出さなければいけない文書を出していない可能性や、逆に、出さなくてもよい文書を出してしまっている可能性につながっている。

この機能を作ってから、レビューアーとして契約する弁護士と、最初の段階でこういったアセスメントをするようになった。それで、何回指摘してもこの判断のずれが是正されない方は、最初の段階で契約を解除する場合もある。こういったエキスパートの業務の質の判断は、以前は行うことが難しかった。

ーリーガル事業における AI 適用事例

以上のリーガル事業の部分をまとめると、eディスカバリの中で非常にコストが掛かる部分を、我々のエンジンを使ってコスト削減しているという事例となる。また、こういったノウハウを使って、法執行機関にこのソフトウェアを導入していただいている。そこで捜査員の方々が集めてきたドキュメントに対してこの機能を用いて捜査するなど、具体的に細かいところは申し上げられないが、使っていただいている。

また、第三者委員会の調査。企業で何か問題が起こった場合に第三者委員会を作って、自社の事案に対して調査するようなことをよくやるが、その中でもこのようなソフトウェアを使ってかなり短い時間で証拠を見つけることができる。実際にあった事案だと、20人ぐらいのチームで2週間ほどかけて見つけれなかったものを、この学習を使うことによって1日で発見できた例もある。また言語解析の部分だけではなく、媒介中心性というロジックを使い、人のネットワークを可視化するツールと組み合わせて使うことによって、最初に想定していなかった共犯者を見つけるということも可能である。この機能は、ある不正に関連するやりとりをしている人のネットワークだけ浮かび上がらせる。その中でどいう人がキーパーソンになっているのかをネットワーク化して見ると、今までチェックしていなかったが、実はこの人が黒幕なのではないかとか、不正の協力者が実はほかの会社にいた等の分析ができる。

守本委員

メールで人の関係性を示すのに、今まではメールのやりとりの本数が多いとか、そちらで判断しがちだったが、まず我々の AI、KIBIT で関係するメールだけ抽出して、そのメールをやりとりしている人だけを出すので、本当に重要なやりとりしている人だけが見つかるので、今まで知らなかった人も浮かび上がってくる。

－電子メール監査システム

こういった訴訟支援の領域で成果が上がってくると、では、事案が起こらないほうがよいのではないかと、事前に止められるのだったらそれに越したことはないのではないかとという発想に当然なっていく。そこで取り組んだのが監査のシステムである。もともと平時にこういったシステムを導入することによって、事案が起こること自体を軽減できないか、そのリスクを下げられないかに取り組んだ。

・ EMAIL AUDITOR の概要

これが EMAIL AUDITOR と呼んでいる電子メール監査システムのソフトウェアである。企業がこのようなシステムを導入する動機として、いろいろな背景があるが、ポイントはやはり訴訟における調査のときと同じ構造である。1 回何らかの訴訟に巻き込まれた会社で非常に意識のあるところは、人間の監査をあらかじめ始めているところも多い。ところが、実際に見ようと思うと大量のメールがヒットしてしまう。結局メールの中身を見尽くせないで、また同じようなことが起こってしまう。このような悪循環を繰り返してしまうわけである。そこで、判断をこの KIBIT に代替させて、毎日やりとりされる多くのメールのすべてに対して判断が行き渡る状態を作る。これでリスクを大きく下げられる。

もう一つは、人間がアクションにつなげられる解析をしてやる必要がある。例えば、情報漏洩の不正を見つけようとしたときの考え方であるが、ICT 的な発想の範囲で解決策を考えると、情報漏洩という状況は情報が外部に漏れたときだけ、と考える。ところが、人間に着目していくと、もっと違った絵が浮かび上がってくる。

情報は基本的に使われないと意味がないので、情報を使う権限を持っている人が本当に外に持ち出そうと思ったときに、止めることはなかなか難しい。だから、そういった状況に至る前に何とか止められないかと考えた。

情報漏洩のステップは犯罪学的なアプローチを使って3つの段階に分けられる。1つは醸成という段階。これは会社に対する不満がたまってきた、もう辞めよう、独立しようとか、次の会社に行こうとするという工程を指している。次に、それを決めてしまったら、外に持っていくためのデータを集め始める。これが準備という工程である。自分の社内のネットワークを使って、そう言えばあの顧客データの最新版はどうなっていたかとか、製品データの最新版を送ってもらえないかとか、そうした情報を、人を通じて物色するようなことが結構見られる。こういった形で情報を集めるのが準備段階で、実際に外に持ち出すと、これが実行という最終フェーズに突入する。ところが、実行フェーズになってくると、なかなか止めるのが難しい。だから、醸成準備の段階でなるべく手を打って、実行に移す可能性を少なくする。

先ほどのカルテル、談合というような例であれば、必ずしもその担当者に悪意があってやっているとは限らない。営業活動を熱心にやっていて、その中でレギュレーションを踏んでしまったようなケースもある。だから、このアプローチを使って初期の段階で、例えば競合相手と深いコミュニケーションを取り始めたときに、それは危険だからやめておいたほうがよいと法務・コンプライアンス部門で教育することで実際に防いでいるケースがお客様の中でもある。こういった人間のアクションにつながるような解析を KIBIT と組み合わせることによって、アプリケーションとして提供している。

守本委員

ある企業は、これを使って退職者の情報漏洩を防いでいる。日本の企業はほとんどカルテルを対象として監査している。実際これを入れた会社は、カルテルで制裁金を払った会社であるが、我々のシステムを入れた結果、米国司法省から、すごい取り組みだと認められた。

—PATENT EXPLORER

我々はリーガルの分野から徐々に活用領域を広げてきている。特許調査の分野では PATENT EXPLORER というソフトを提供している。この中でも KIBIT が動いている。

年々世界で特許は増えているが、特許調査の分野では、リサーチャーの育成に苦慮している。特許検索自体が特殊技術であるからである。従ってその技術の中身を知っていれば、例えば研究者の方も含め、なるべく手軽に「見つけたい特許を見つけられる」ソフトウェアが待たれていた。そこで、この KIBIT を使い、見つけたい特許の素案の文書を人間が作って入れてやると、その 1 件の文書を基に特徴を解析して、事前に出願されているものの中から近いものを見つけるという解析を行っている。

使い方は簡単である。例えば不要な特許、見つけたい特許の記述を最初に「あり」「なし」のチェックをするだけである。あとは学習のプロセスが走り、大量の特許データの中から見つけたい特許である可能性が高いに特許に、高いスコアが付いて上がってくる。

この分野も、機械学習が有効な分野だと思っている。従来であれば概念検索という自然言語処理を使った解析が使われてきたが、概念検索の場合、漏れがない検索をしようと思うと、ノイズが非常に多くなり、ノイズを少なくしようと思うと漏れてしまうというトレードオフの関係だった。それをなるべく漏らさないで、しかもノイズも少なくするようなやり方を実現している。

いわゆる学术论文とか、文献も全部対象にできると思うが、今我々は、主に特許文書で使っている。今後は論文調査も含めてやっていきたい。

このシステムは、トヨタの関連会社である、トヨタテクニカルディベロップメント (TTDC) と一緒に共同開発をしたものである。テクノロジーのほうは我々が提供させて頂き、業務のノウハウや現場への適応に関する部分は TTDC に担当頂いた。こういう形でコラボレーションを行っている。

守本委員

特許分析のエキスパートの暗黙知を我々が取り入れて作っている。

ービジネスインテリジェンス導入事例

次に、三菱重工業に導入いただいている例であるが、これはアナリストの知見を学んで資料を集めるところを効率化した例になる。

三菱重工業は、例えばエネルギー、産業機械、環境など、いろいろな分野を手がけているが、こういうたくさんの分野を俯瞰して分析できるアナリストは、貴重な存在である。経営層に向けてのレポートをマーケット部門がやっているが、その中で、例えばロイターや日経のテクノロジー系専門誌、そのほかのニュースソースといった大量のデータをまずはアナリストの方々が集めてきて絞り、その絞った物から必要かどうかを判断してレポートをまとめることを平常やっている。このレポート作りに非常に時間が掛かり、その中で、資料の収集についても多くの時間を要しているという課題があった。そこで KIBIT を入れることによって、生産性を向上させた。資料を集める時間は大体半分に削減されている。

ー精神疾患の定量的判断を可能にする医療機器の開発

次に、我々のもう一つの大きな課題である健康な社会の実現についてだが、この分野は同時にエキスパート、医師や看護師といった方たちの効率を上げたいというところにも関わってくる。このヘルスケアの取り組みを幾つか紹介させてもらいたいと思う。

1 つ目が、精神疾患の定量的判断に関する、「PROMPT」と呼んでいるプロジェクトである。これは AMED のグラントを受ける形で研究させてもらっているが、慶應義塾大学と今一緒にやっている。また、アドバンスト・メディアなど機械学習の技術を持った会社とのチームでやっている。この中で FRONTEO が担当しているのは、言語解析の部分である。

研究対象としては、気分障害（うつ）、また認知症もターゲットの中に入っているが、これらを定量的に自動診断しようというプロジェクトである。医師と患者の間に我々のデ

バイスを置いて、顔の表情や問診中の音声などを取得する。例えば音声であればトーンのようなものもあるし、ボイス to テキストにして我々の言語解析のところでその特徴を見る。こういったデータをクラウド上にいったん送って、クラウド上でいろいろなモダリティーに対する解析をし、最後にそれを統合して、この人はこのぐらいうつに近いというようなことを最終的に判断する。こういうデバイスを今造ろうとしている。

この課題になぜ取り組んでいるのかという背景だけ少し話したい。今精神疾患の診断は、様々な研究がされているが、決定的なバイオマーカーがないので、精神科医の経験的な判断に頼らざるを得ず、客観性を増やさなければならないという課題意識が持たれている。このシステムで、解析が実現した際には精神疾患における診断の客観性がかなり上がると考えている。

一 症状悪化の早期発見

次は、LITALICO と一緒に取り組んでいる障害者の就労支援に関するソリューションである。LITALICO は東証マザーズ（※2018年4月現在は東証一部）の上場会社で、障害者の就労支援を大きくやっている。LITALICO ワークスというサービスがあり、精神障害を患った方の就職支援を全国に事業所を持ってやっているが、その中で注意が必要なケースがある。就労移行のプロセスでストレスが発生し、その過程で症状が悪化してしまう可能性がある。早い段階でそういう兆候を見つけてケアすることが重要である。

そこで着目したのが、支援記録という情報である。全国の事業所から支援記録というレポートが毎日送られて、ライフネット支援室という中央の部署で確認している。1日数千通送られてくるので、全てのドキュメントに同じ質で判断することは困難である。かつこの人は危ないのではないかと予兆を発見する力には支援者によって個人差がある。熟練の方であればできるが、若手の教育も簡単ではなく、時間も掛かってしまう。これを KIBIT で何かできないかということで導入した。

最初に行った実験であるが、340件ぐらいの支援記録を対象にして学習させた。そこで症状が悪化したものと、そうではないものをランダムに混ぜたデータを用意して評価として

使ったが、上位 30 件の中に、悪化したケースをほぼ抽出することができた。4 カ月で本格運用を開始して、実際に使ってもらっている。

そして、実際に運用していく中で、AI に負けたくない、支援している方々のモチベーションが上がったという反応も見られた。また、若手の方が KIBIT の判断を見て、「あ、これは確かに見つけなければいけなかった」と、そこで学ぶ。つまり人工知能からの結果を見て、人が学ぶという側面もある。

まとめ

我々の課題意識としては、たくさんの方々にこういった機械学習や、自然言語処理のメリットを享受してもらい、その裾野をなるべく広げていきたいということがある。その中で現実の課題をどんどん解決していく。今 AI はブームとなっているが、このブームの中で社会実装をいかに作っていくのが本当に大きな課題だと思うので、これを一つでも増やしていきたいと思っている。

AI が人間の職を奪うとか、そういう話になりがちであるが、実際アプリケーションを提供している中で、モチベーションを持っていただいたり、逆にそこから判断を学んだりということを見ていると、やはり AI と人間が協調していく形、AI 融合社会が目指すべき姿なのではないかという感覚を我々は持っている。

最後に、このプロジェクトは日本の企業に対して提言を出していくのが目的なので、普段感じていることとも絡めて述べたいが、やはり人工知能を試したいユーザーはどんどん増えていっていると感じている。そのような状況の中で、ユーザー企業のデータ解析に対するリテラシーが大きく上がってきていると感じている。例えば機械学習における判断結果についての受容として、以前は 9 割判断が当たっているとすれば、1 割も外れるのか、というのが、10 年ほど前に特許の解析をやったときの多くの方の反応だった。しかし最近では 9 割当たることをどう生かすのか、と考えるユーザーが増えてきたと感じている。こう

した意識の変化は隔世の感がある。ユーザー企業はとにかく導入して、実ビジネスの中でどう活用するかを見きわめていくことを、どんどん進めていっていただきたい。

また、何でもデータを入れさえすれば判断ができるようになるものではないと思っているので、費用対効果も考えた上で、実際何に対して問題の解決をしようとするのか、どのようにこの技術を使っていくのかを設定していくことがやはり重要ではないかと考えている。

守本委員

いろいろな人工知能を作るうえでデータを読ませるために、アメリカの会社ではコストを掛けて多くの人がやはりラベリングをやっている。日本ではなかなかそれができないが、訴訟という特別な場所にいる我々のような中小企業がなぜできたか。実は訴訟費用の半分以上、証拠を発見する上での80%以上のコストが、人がラベリングすることに支払われている。エキスパートの弁護士がラベリングをし、そのデータをたくさんためておくという業務に携わっていたので、このような人工知能の開発につながったが、もう一つの特徴としては、日本人の発想でアプローチしたことによって、非常に少量のデータでも学習を可能にすることができた。学習するのに大量の情報を使う必要がない。データはあってもよいが、多く入れれば良くなるというわけでもない。うまく効率的に人の暗黙知を取り込んで、結果的に少量でも判断できると。そういった日本らしい人工知能を我々が開発し、今使われているという意味では、活用範囲をさらに広く使えるのではないかと考えている。