

Visual Mining Studio、Visual R Platform、Deep Learner、Big Data ModuleはAlkanoとしてリニューアルしました。

アジア特許情報研究会 知財情報解析チーム 様

特許情報解析は
いよいよディープラーニングへ

知的財産の一つである特許・実用新案。その出願にあたって企業はグローバルに対応する必要に迫られている。アジア特許情報研究会様では国をまたぐ各種の特許情報を、企業の垣根を越えたメンバーで研究している。その中の知財情報解析チームでは、**Text Mining Studio** (以下**TMS**) や**Visual Mining Studio** (以下**VMS**)、さらに**Deep Learner**といった各種ツールによって解析作業を合理化、高精度化している。

Interview

TMS+VMSにより特許情報解析が一段と効率化

特許情報解析に**TMS**を導入したきっかけを教えてください。

安藤 特許出願にあたって調査すべき関連特許とはときには数万件にも及び、担当者の負担は尋常ではありません。それを最新の技術を活用していかに効率化、省力化、そして高精度化していくか。それが我々メンバーの研究目的です。はじめは効率化のためにテキストマイニングのツールを使い、どの特許を読んだらいいか、どの順番で読むべきかをテキストベースでランク付けすることを主に行っていました。近年は世界の特許・実用新案出願の約6割を中国が占め、中国語による文献も非常に多くなっているため、中国語のテキストマイニングなども研究しています。

西尾 **TMS**を使い始めたのは2015年からです。自然言語処理に非常に適していますね。特に文章を解析用に細かく分ける形態素解析技術が優秀で、特許に関連する専門用語をきちっと抽出してくれます。例えば「炭酸カルシウム」を「炭酸」「カルシウム」と分けて抽出してしまうと解析結果が大きく変わってしまうのですが、**TMS**では始めから「炭酸カルシウム」と1単語で抽出されます。そういった処理の切れ味が素晴らしいという印象です。文節や助詞、助動詞それぞれに狙い通りの調整もできます。さらに結果をExcelで開くことができる形式で出力される点も、後々使いやすくて助かっています。

TMSに次いで、**VMS**も導入されました。

平川 **TMS**を使ううちにだんだん我々の要求も上がりました。特許情報の中には、担当者が読みたい情報とそうでないものがあり、必要ないものはノイズでしかありません。そういうノイズを排除してさらに特許情報を研究する効率を上げたいという要求が出てきました。さらに、抽出した特許情報をどのように分類するかも課題に上がりました。特許情報を俯瞰して体系立てて分類するのはとても難しい作業なのです。そういう要望を数理システムに相談したところ**VMS**を薦められ、使うようになりました。

安藤 テキストマイニングで得たデータを**VMS**で解析することで、調査目的に応じて特許情報にランキングやプライオリティを付けてノイズを排除したり、合理的に分類することが可能になりました。場合によっては1万件以上ある特許情報の中から必要な情報を1,000件、100件と、だいぶ絞り込めるようになりました。**TMS**によるテキストマイニングだけでなく、そのデータを活用して**VMS**でデータマイニングへ、一種の機械学習を行うというワンランク上のレベルに到達したイメージです。この組み合わせで、さらに思うような結果を得ることができるようになりました。

西尾 **TMS**のデータを用意して、それとモデルを作るアルゴリズムとをグラフィカルに矢印で結び操作をすれば、あとは**VMS**が自動的に計算してくれます。計算速度は速いですし、インターフェースは使いやすい。統計解析の知識があれば、さらに自由に使いこなせるでしょう。ランキングや分類をするにはどのようにモデル構築するかにかかってくるので、そのモデルを構築できるスキルを早く身につけたいと思っています。そうすれば**VMS**上で再現できるようになり、あとはデータを投入すれば狙った結果が得られるようになります。



アジア特許情報研究会
知財情報解析チーム

安藤 俊幸 様 (中央)

花王株式会社

研究開発部門 知的財産部

西尾 潤 様 (左)

株式会社ユポ・コーポレーション

商品開発本部 市場開発部 次長

平川 雅彦 様 (右)

JFEテクノリサーチ株式会社

知的財産事業部 知財調査部

倉敷グループ グループ長

PROFILE

アジア特許情報研究会
知財情報解析チーム 様

アジア特許情報研究会は中国や台湾、韓国などの東アジア、およびASEAN諸国など、情報が不十分な地域の知財情報を研究する目的で2008年設立。メンバーは企業各社の知財関連のスタッフなど21社22人で構成されている(2018年7月現在)。この中で、知財情報解析チームはテキストマイニングや機械学習による効率的な特許調査法を研究、各所でその成果を発表している。

Deep Learnerのご感想をいただけますか。

平川 Deep Learnerは、VMSのアドオンとして数理システムから先行版の案内があったときに※、さっそく導入を計画しました。ディープラーニングは今最先端の解析技法で、そろそろ使ってみたいという技術者としての好奇心もありましたね(笑)。

安藤 解析したいデータを用意して、画面上のアイコンの手順通りにどんなディープラーニングのネットワークにするか、何層くらいにするかを選んで、細かいパラメータを設定、その際、TMSで解析したテキストをそのまま投入することも可能です。あとは実行ボタンを押せば学習がスタートしますが、そのときの進捗状況がグラフィカルに出てくるところが私は気に入っています。ディープラーニングのロスの計算が進み学習が進んでいる状況が画面に表示されるのですが、それがリアルタイムに減っていく。学習がどこまで進んだか、いつ終わりそうなのかが見て分かる。だから、いろいろなパラメータを入れて確かめたくなくて、ついつい夜ふかししてしまうこともあります(笑)。

平川 私はモデルを推奨してくれるところがいいと思いました。Deep Learnerではどのモデルにするか、データや目的に合わせて推奨してくれます。これまで特許情報の分類で、機械学習の一つである非線形サポートベクターマシンを使って解析していたのですが、それだと平均40%程度しか精度が出ない。それがディープラーニングの線形モデルを使うと、約80%も出た。これまではどのモデルにするか、それを確かめてみる作業に時間や手間がかかっていたのですが、推奨モデルを出してくれるのはありがたいです。

安藤 ディープラーニングってこんなに手軽だったのか!という衝撃をこのツールで受けました。ディープラーニングのモデルは表現能力があります。モデルの対応力や柔軟性が高いので、細かな特徴を拾って学習させるような精度の高いモデルを作ることができ、それによって解析精度のさらなる向上が期待できます。そういったディープラーニングの特長を容易に利用できる。Deep Learnerは素晴らしいツールです。

ツール全体のご感想や、あるいは今後のご希望をお聞かせください。

安藤 我々の特許情報研究は、TMSに始まってVMS、そしてDeep Learnerと、数理システムのツールで着実にステップアップしてきました。しかも数理システムのスタッフは、その時々に応じて生じる私たちの疑問やニーズにいていない、しかも迅速に答えてくれます。課題に対する解決手段を教えてくれたり、その際も分かりやすく可視化してくれたり、そういった一つひとつの対応に、解析に対する熱意や愛情が感じられます。常日頃からの対応に、非常に感謝しています。

西尾 TMS + VMS、この組み合わせで特許情報解析の敷居がものすごく下がりました。Deep Learnerによって、その解析結果がさらに高精度になることを期待しています。私はこれらツールを統合したパッケージができるといいと思っています。特許情報解析のパターンや目的に応じて、「このケースはこのモデリングを使ってこういうデータを投入すればいい」といった作業がワンストップでできるパッケージです。そうすれば解析にかかる時間をもっと合理化、短縮でき、その分の時間を特許情報の読み込みといった本来業務に充てることができます。

※Deep Learnerは2018年10月1日より販売しています。

アジア特許情報研究会では研究員を募集しています!

アジア特許情報研究会 事務局 伊藤 徹男 様

私たちには研究したいテーマがたくさんありますが、研究メンバーが足りていません。アジア・新興国の知財情報、知財情報解析に関心がある方は、ぜひ事務局までご連絡ください。
Mail:patentsearch2006@yahoo.co.jp



Deep Learnerによる解析の一連の流れ

