

ATIS 第 404 回例会報告

2018 年 9 月 19 日、国立オリンピック記念青少年総合センターにおいて ATIS 第 404 回例会が行われました。

まず、代表幹事から、会員の入退会状況、例会活動、スケジュール紹介、第 1 回幹事会報告、その他の報告が行われました。7 月に特別会員として(株)カネカ、賛助会員として(株)サン・フレア、(株)パソナナレッジパートナーの 3 社が入会しました。

続いて、長年、ATIS 例会において講演をして頂いた元特許庁の後谷陽一様に感謝表彰を行いました。

引き続き、2 件の講演を開催しました。1 件目の題目は、「最新の機械翻訳技術と今後の展望」、2 件目は「これからの特許調査解析の展望と求められる人材」でした。翻訳関係者やこれから調査担当者を目指す方々にとっては、大変興味深い内容でした。

講演 1 「最新の機械翻訳技術と今後の展望」

講演者 NICT フェロー 隅田 英一郎 様

初めに、機械翻訳の概要についての紹介がありました。NICT での機械翻訳の研究開発は、SMT（統計的機械翻訳）から着手し、その後 NMT（ニューラル機械翻訳）にシフトしました。現在の NMT はとても流暢な翻訳が可能になっており、ネイティブが作成した文章と勘違いするレベルになっています。しかし、そのためには、コーパスサイズとして 100 万文以上の大量のデータで学習させることが必要で、コーパスサイズが大きいほど翻訳精度が高くなります。NMT の進歩の度合いは早く、2017 年までは SMT の精度が良かったのですが 2018 年には NMT の精度が SMT を追いつき、2019 年には NMT の精度が SMT よりも格段によくなると見込まれています。特に、特許翻訳については、対訳データ量が超大規模になるので、NMT を選択することが適しています。その際注意することは、データの量と言語の構造から言語間の翻訳精度に差が発生することで、言語としては日/韓が最も精度よく、続いて中/英、日/中、さらに日/英の順で、日/独の選択は若干劣ります。NMT の特許翻訳への活用可能性は高く、外国語明細書作成支援の使用例もあります。NMT は、適用分野ごとに学習することで流暢かつ正確な訳文を得ることができるなど高い実力を持っていますが、一方、訳抜け、湧き出し、高性能な GPU が学習に必要ななどの課題も残っています。訳抜けは学習データに含まれる単語が無いと訳されないこと、湧き出しは不要な単語を学習することで原文に無い訳をすること、例えば原文に「発表する」しかなくても学習データに「木曜日に発表する」が多数含まれていると本来「木曜日」は不要にもかかわらず勝手に単語を追加して訳すことです。また、学習時の演算処理には高性能な GPU が必要になり、ハード費用が増加することも想定されます。しかし、訳抜けや湧き出しはソフト改善により徐々に減ってきており、ハード費用も今後低下していくことが見込まれます。NICT としては、より精度の高い翻訳を提供できるように、開発した「みんなの自動翻訳」を SMT から NMT に換装しました。

NMT を支える技術は、アルゴリズム、対訳データの準備、翻訳精度の評価、GPU の 4 つから構成されています。アルゴリズムはオープンソース化が進んで差が出にくい部分ですが、対訳データの準備や翻訳精度の評価は利用者の支援を受ければ受けるほど、精度が向上する差別化ポイントになります。従ってユーザーの協力をできるだけお願いしたいと考えています。NICT と総務省は、これに関係して翻訳データを集積する『翻訳バンク』の運用を開始して機械翻訳の精度向上を進めています。翻訳バンクは、色々な分野で使用されている対訳コーパスを一カ所に集めて高精度な自動翻訳機を開発するための辞書に活用することを目的としています。現在は、汎用の NMT のみですが徐々に専用分

野の NMT を供給してさらなる精度向上を計画しています。翻訳バンクの事例紹介として、製薬分野での複数の紹介がありました。

翻訳の評価については、内的基準（人間評価、自動評価）と外的基準（アプリケーション達成度による機械翻訳の評価）を考えており、それぞれ一長一短があり必要に応じて使い分けます。

NMT は特に分野適用に適しており、学習させた汎用 NMT に対して分野コーパスを使って更に追加学習させることで、その分野ごとに適した精度の高い翻訳が可能になります。現状で考える最良の翻訳は、RBMT（ルールベース）、SMT、NMT を併用して、流暢さ、忠実度、汎用性、必要なハードウェアなど、それぞれの特徴に合わせる機械翻訳です。NICT の NMT は Office のプラグインも用意しており、Word や Excel など普段使っているファイルを直接翻訳することが可能で、既に複数の企業にライセンスしています。



講演 2 「これからの特許調査解析の展望と求められる人材
-IPCC（工業所有権協力センター）の状況をもとに-

講演者 IPCC 常務理事 保倉 行雄 様

最初に、IPCC の概要説明がありました。IPCC は 1985 年に公益法人として設立され、2009 年に一般財団法人に移行し、工業所有権に関する調査（先行技術調査）と関連する基盤の整備（分類付与）を目的としています。人員は約 1800 名ほどで、検索事業（内外文献検索）、分類付与事業（一元付与、公開後解析）、公益目的事業（特許検索競技大会・大学知財活動助成）が主な事業です。検索事業では、2013 年は国内文献検索 80%、海外文献 1% でしたが、2018 年には国内 50%海外 30%と海外文献の検索割合が格段に増えています。特許出願状況では 2005 年から中国の出願が増加し 2009 年に日本を、2011 年には米国を抜いて世界トップの出願数となっています。分類付与業務は、特許が特許庁に出願されると IPCC で自動大分けして FI と F タームの分類を付与する事業を行っています。

検索事業のために、主に 50 代以降の人材を採用しています（採用時の平均年齢は 55 歳くらい）。求める人材としては、技術能力が高く、コミュニケーション能力や英語の読解力に優れている方です。事前研修、主幹指導、端末

操作研修、OJT、任用後研修、スキルアップ研修、特別研修、一般教養研究、管理職研修等、豊富なメニューがあり、人材育成に力を入れていますので、検索や知財経験が無くても問題ありません。

研究所の概要説明では、分類等コードの推定の研究、図/表の処理技術の実用化研究、外国文献検索支援機能の研究テーマの紹介がありました。分類等コードの推定の研究では、自動大分けシステムで分類したのちに F タームと FI コードを推定するものです。単語だけの一致判定ではなく文章に踏み込み類似性判定を行い、高精度の F タームと FI コードを推定します。図/表の処理技術の実用化研究では、イメージマッチング手法による図の検索手法及び表からテキストを抽出して検索する手法を研究しています。外国文献検索支援機能の研究では、英文検索式作成補助ツールを開発して、和文検索式から日英の関連語を抽出し所望の訳語選択し英文検索式を導く研究をしています。

特定登録特許調査事業は、特例法に基づく「特定登録調査機関」として民間向けの先行技術調査サービスを行っています。出願後かつ審査請求前の案件が対象で、審査前の適切な補正、外国出願の要否の判断、早期審査案件、オープンクローズ戦略としての公開前取下げ、等で活用いただけます。

最後は、特許検索競技大会、表彰（アドバンスコース）、フィードバックセミナーの紹介がありました。特許検索競技大会は、J-PlatPat や商用 DB を用いた実際の検索作業を含む問題で実務能力を競う大会で、初級コースに加えプロ向けのアドバンスコースがあり、2018 年も全体で約 400 名が参加しました。表彰は、アドバンスコースで優秀な成績を残した個人と団体に贈呈し、フィードバックセミナーは、大会問題の正解例や検索方法を解説するセミナーです。これらを各社の自己研鑽、人材育成に活用検討してください。



以上