


対談



筑波大学大学院 ビジネスサイエンス系
津田 和彦 教授



(株)NTTデータ数理システム 取締役
小木 しのぶ

曖昧な日本語をきっちりマイニングできる その技術に驚かされました

筑波大学大学院で、15年前からテキストマイニングの講義を行っている津田 和彦教授。古くから自然言語処理の研究に携わり、プログラマーとしても有名ワープロソフトの開発に関わってきた経歴も持つ津田教授は、NTTデータ数理システムのテキストマイニングツール **Text Mining Studio** (以下、**TMS**) をどう評価されているのか。企画段階から**TMS**開発を担当したNTTデータ数理システム 小木 しのぶがお聞きしました。

民間企業でワープロソフトの言語処理機能を開発しました (津田)

小木 津田先生は民間企業でのご経験がおりですね。

津田 大学卒業後、電機メーカーに5年間、鉄鋼メーカーに7年間勤めたあと、大学教員になりました。鉄鋼メーカー時代にはワープロソフトの文章校正機能を開発しました。ワープロソフトを使っていると、テキスト入力の際に赤や緑の波線が出てくるときがありますよね。例えば『東京の文京区の大塚の何々の』と「の」を多用しすぎるなど間違った言葉づかいをしたときに指摘する機能です。その開発に携わりました。

小木 いつから自然言語処理研究に関わられたのですか。

津田 徳島大学で自然言語処理の研究室に所属していました。その研究室では日本語変換システムを開発しており、私はアルゴリズムの中心的なプログラムを組みました。

小木 そうだったのですか。

津田 当時はプログラマーという立場で、とにかくいいプログラムを作りたくてプロジェクトに関わりました。その後、電機メーカーを経由して自然言語処理のニーズがあった鉄鋼メーカーに移りました。有名な製品の開発に携われたことは、ひとつの成果だと思っています。

小木 いまは大学教授のほかに、社外取締役として企業運営にも関わっていらっしゃいますね。

津田 私が所属している筑波大学大学院 ビジネス科学研究科は社会人のための大学院で、ここに来る学生たちはビジネスの課題を持ってやってきます。それを学生たちと一っしょに考えていくことが私は好きです。ときには学生の会社の実データを見ながらテーマを考えたり企画立案したりしています。このような状況の中で学生や学生の上司が重役に昇進した際、「社外取締役になってもらえませんか」と声がかかることがあります。企業のガバナンス向上のために、社外取締役を置く風潮が高まったという背景もありますね。

小木 ビジネスに関して先生がお持ちの知見や幅広い知識が評価されたのでしょうか。そうやってアカデミックの世界にいらしゃってもビジネスの世界とのつながりをもって、両方の感覚を持たれているのは素晴らしいことですね。

PROFILE

津田 和彦 教授

1986年徳島大学工学部 情報工学科卒業、1994年同大学工学研究科システム工学修了(工学博士)。三菱電機株式会社 LSI研究所、住友金属工業株式会社システム研究開発部を経て、1998年筑波大学助教授・社会学系、2005年筑波大学大学院ビジネス科学研究科教授。主な研究分野に自然言語理解、情報検索、ソフトウェア工学、データ構造とアルゴリズム。

テキストマイニングは15年ほど前から大学で教えています (津田)

小木 先生は、経営学学位取得のプログラムの中でテキストマイニングの講座を開かれています、その講座の狙いをお聞かせいただけますか。

津田 15年ほど前、「テキストマイニング」という言葉が出たばかりの時期から講座を始めています。当時、SNSやブログなど個人が情報発信をする手段を得て、それが広まっていた。企業がその情報に価値を見出すようになると、マーケティングなどに活用できないかといった声が大学院に来る社会人学生から聞かれるようになったのです。我々は社会人大学院として教育プログラムを提供する必要があると考え、テキストマイニングの講義を始めました。

小木 学生さんや企業のニーズが発端だったのですか。

津田 5、6年前のAIブームのあたりから取り組む学生はさら

に増えています。修士課程の定員30名に加え、他専攻の学生からも希望があり、合計で40~50名が受講する人気の講座になっています。

小木 テキストマイニングも最近は身近になってきました。講座を開始された当初と比較して、授業の雰囲気は以前と変わってきていますか。

津田 テキストマイニングを通して知りたい目的がはっきりしてきました。以前は、テキストマイニングすれば何か分かるかもしれないという期待が強かったのですが、最近は「当社の製品がどの地域で好評なのか知りたい」など、テキストマイニングによって得たいものが具体的にになっています。目的がはっきりしているため、その後どう分析していけばいいか、方針や方法論も立てやすくなります。

TMS開発では、自然言語処理で悩みました (小木)

津田 そういえば、御社がリリースしている**TMS**は、小木さんが開発に携わったそうですね。

小木 はい。2000年に当社でリリースしたデータマイニングツール**Visual Mining Studio** (以下、**VMS**) を契機に、お客様からのマイニングに関するご相談が非常に増えました。そうした折、当社の創業者である山下 浩が次はテキストのマイニングが必要になると予見。ありがたいことに私に声がかかり、開発が始まりました。当時はほとんど資料がなく、やっと見つけた本を何回も何回も読み直してはどんな機能を持たせるべきか考え、試行錯誤してつくったものです。2005年に**TMS**をリリースすると、アンケート結果から定性的な情報を分析したいといったニーズが次第に増え、おかげさまでいまは実にさまざまなお客様に使っていただけている状況です。

津田 テキスト処理をし、その結果をマイニングするツールとするとあたって開発の苦労はありましたか。

小木 **TMS**に先立って**VMS**を開発していたので、マイニングの部分はそのエンジンを使う前提で開発が進められました。ただ、テキスト情報をマイニングの手法を使ってどのように役に立つ知識に昇華させるかということに苦労しました。日本語は膠着語*に分類され、そうではないほかの言語に比べて文章を単語分けする必要性があり、また、語順の前後も生じます。そのため、分かれすぎてしまった単語をまとめあげ、語順の前後を考慮しながら、意味を持つ知識の形にどのように仕上げるか、といった自然言語処理とマイニングの融合、まさにテキストマイニングの実現に悩みました。日本語に関して、先生はどのような点が難しいとお考えですか。

津田 おっしゃる通り、単語分けが必要なところですね。また、主語を飛ばすこと。それから口語で語尾を濁すところも特徴的です。文字にすると三点リーダー「…」です。「性

能は良いけど…」と書かれたら、大事な部分がソフトウェアでは読み取れない。褒めているのに、「…」がつくと不満があるとれますよね。直接的に表現せず、協調性を大事にするのは日本人の民族性だと言われたりもしますが、そういう性質が言語にも現れていると感じています。

小木 感性分析という手法が最近注目を集めていますよね。日本語こそ、そういった分析がビジネスの世界で役立つのではないのでしょうか。

津田 アメリカやドイツで感性分析を行っている企業は結構ありますが、日本語は難しそうですね。

小木 行間を読む、という領域になりますね。

津田 コンピュータは行間を読み取れませんので、その文章にないものをソフトウェアで補完しなければいけません。すでに文章中にあるものなら、どう解釈するかを多数決による論理で行えます。人工知能の自動学習も基本的には多数決論理です。多数決は学ぶことができますが、事象としてないものは学べません。行間を読むには、ルールベースで処理するしか方法はないんですね。

小木 行間に潜むルールを、人間がコンピュータに教えてあげるといっていいのでしょうか。

津田 そうですね。三点リーダーがつくと、それ以外に言いたいことがあると文章につけ加えるんです。

小木 先生のご研究では、そのようにテキストデータをモデリングしていくところと、ルールベースとを両方組み合わせて分析されているパターンが多いのでしょうか。

津田 行間を知りたいといったテーマですとルールベースの研究になります。一方、例えば顔文字などは多数決論理で分析できます。目や口の形を見れば、全何百、何千種類もある顔文字も、喜んでいる、怒っているといった感情を見分けることは多数決で可能です。

* 膠着語:ある単語に接頭辞や接尾辞のような形態素を付着させることで、単語の文の中の文法関係を示す言語。対照的な言語は屈折語。
ドイツの言語学者ヴィルヘルム・フォン・フンボルトによる自然言語の分類。

曖昧な意味もここまでマイニングできるんですね（津田）

小木 TMSの印象はいかがでしょうか。

津田 小木さんのご説明の通り、TMSはマイニングに強い会社が自然言語処理の世界に出てきて作り上げたツールですよね。自然言語処理からマイニングの世界に入ってきた私にとっては、曖昧な言語から出てきた意味や意思をどうマイニングするのか興味があったのですが、それに見事に応えてくれた製品だったので驚きました。テキストマイニングではほかにもツールがありますが、その中でずば抜けて分析結果の表現がうまいと感じています。

小木 ありがとうございます。当社の製品群全体がそうなのですが、ユーザー様から「こういうツールがあったらいいのに」「こういうことをやってみたい」といったお声をいただいて、それをできるかぎり製品に反映するよう努めており、そういうプロセスから生まれたひとつがTMSといえます。

津田 逆に、開発の立場から、TMSの“売り”はどのような点ですか。

小木 世の中にはフリーでも素晴らしいテキストマイニングツールがたくさんあります。その中で当社が有償としているひとつの意義は、皆様に継続的にご提供することを企業として保証していることです。それは有償ツールとして絶対を守る必要があると考えています。また、ユーザーの皆様への厚いサポートも継続しなければいけません。

津田 なるほど、そうですね。

小木 また技術的なこととして、TMSやVMSのほか数理最適化やシミュレーションといった当社ツールは、Visual Analytics Platform（以下、VAP）というプラットフォーム上でシームレスに連携してお使いいただけるようになっています。VAP上で前処理したデータをTMSやVMSにかけて、その結果から予測をしたり、

ほかのデータと組み合わせて数理最適化を進めたり、その結果から立案した計画はどの程度有効かといったシミュレーションまで、ひとつのプラットフォーム上で実行可能です。こうした分析系のデータサイエンスの世界を一気につなげることが当社のツール群の大きな特徴です。

NTTデータ数理システム 主要パッケージ製品リリース年表



※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。

言語と数字を融合して分析できるツールがあるといい（津田）

小木 当社に対する印象をお聞かせいただけますか。

津田 私がNTTデータ数理システムをはじめて知ったのは20年ぐらい前です。当時、統計処理など数学の世界では外国製ツールが多い中、数理システムは和製で本格的なツールをつくっていることで有名でした。その会社がいきなりTMSをリリースしたのでびっくり。数学の専門家は曖昧性のない数値データが専門で、曖昧性のあるテキストを扱うのは難しいだろうと思いついていたからです。でも実際にTMSを使ってみると、マイニングの部分は見事にチューニングされていてとても

よいものでした。さらに御社はツール開発だけでなく、コンサル的な仕事やデータ分析の受託もしている。そうしたサービスから気づきやアイデアを得て、それをまたツールに反映するといったサイクルもある。いままさにビジネスを展開している顧客のニーズがツールに反映されているわけです。時代にうまく乗っているなというイメージを持っています。今後もさらなる活躍を期待しています。

小木 ご期待に添えるよう精進してまいります。本日はありがとうございました。