

特集 「Web 技術, ビジネスモデルと AI」

人工知能技術のネットビジネスでの活用

Transferring AI Technology to Web Business

内山 幸樹
Koki Uchiyamaホットリンク株式会社
Hottolink Inc.
uchi@hottolink.co.jp, http://www.hottolink.co.jp**Keyword:** internet business, AI technology, applied research, commercialization.

1. はじめに

私は 2000 年に、「ほっとする」情報社会の実現、知識循環プラットフォームの実現を目指し、ホットリンク株式会社を創業した。今でいう Web 2.0 にあたるサービスを当時からつくってきており [内山 06], その過程でさまざまな学術機関との連携を行ってきた。近年でも、人工知能技術を活用しているベンチャー企業はいくつもあるが、比較的長い間、人工知能に関する技術を活用する試みを続けてきたと考えている。

本稿では、これまで著者がたどってきた経験を体験記的に書き記したうえで、それをまとめる形で、人工知能技術をネットビジネスに応用する際にヒントとなる知見を述べる。前半は、出来事を時系列で追っていく形になるが、それを通じて、後半の知見のベースとなる事実を共有できればと思う。なお、本稿で「企業」として言及している対象は、IBM やマイクロソフトなどの「成熟した大企業」ではなく、ネットサービスを提供する「新興企業」を想定している。

2. ホットリンクにおける人工知能技術の活用の歴史

2.1 東京大学石塚教授との出会い

1999 年、ソニーがネットだけで完結する販売方法の損害保険販売会社（現 ソニー損保）の設立を準備していた。その際、ホームページ上でのキャラクターエージェントのサービスの開発を著者は担当していた。キャラクターエージェントの質問に答えていくと、最適な保険オプションの選定・見積り提示、事故シミュレーションから（販売会社側から見た）より良い保険への説得をエージェントが行うというサービスであった。Ajax という標準的なフレームワークがまだなかった頃で、同様のしくみを JavaScript と DHTML で開発した。サーバ側で、保険の営業マンの知識を蓄積させたエキスパートシステムをつくり、それとブラウザ上のキャラクターエージェントが非同期で通信するしくみだった。

そのときに、クライアント側のプログラミングを担当していたアルバイトが、東京大学石塚 満教授のもとで研究を行っている学生であった。彼は、研究室で継続して研究されていた MPML (Multimodal Presentation Markup Language) 技術 [筒井 00] の知見をベースに、商用サービスの企画を実現するために必要となる独自の仕様追加を行った。ビジネスの最先端のニーズを吸い上げ、それを解決しながら実サービスまでつくり上げた。非常に優秀な学生であった。

「そんなおもしろいアルバイトをさせてくれる会社があるのだったら会ってみたい」という石塚教授の言葉で、お会いすることになる。当時、研究室では、コンシェルジュサービスに必要となる、インタフェース側の MPML 技術と、そのプレインにあたる検索技術・要約技術・Web マイニング技術の研究をされており、まさにホットリンクが目指す世界に必要な技術がちりばめられているように感じた。

2.2 研究室の技術活用

その直後、著者は今の会社を創業するのだが、創業当時、提供していたサービスは、ブラウザのツールバーからユーザの閲覧行動やブックマーク情報を読み出し、Web ページの内容の分析とユーザ嗜好の分析を行うことで、Web ページや人を検索・推薦するというものであった。現在のソーシャルブックマークがこれに近いが、当時必要だったのは、論文からの知見ではなく、実際のデータを見ながら自社のサービスに合わせた研究開発を行うことだった。しかし、バイトで関わる学生の本業は学業・研究なので、学生自身が企業の実サービスの構築をしながら研究をしてもらおうというところまで巻き込むことは難しかった。

その他、携帯電話の待受画面にキャラクタを配置し、RSS リーダ的機能や各種インターネットサービスを統合させる実験サービスの構築も早い段階から行っていた。また、携帯電話にデフォルトで組み込まれている 3D レンダリングエンジンのキャラクタを MPML 技術を利用して動くようなラッパーの開発、キャラクタメール、キャラクタニュースなどの携帯向けアバターコンテンツサ

ービスを開発した。残念ながら2000年当時、大きな事業として立ち上がることはなかったが、現在はNTTドコモが、iコンシェルという待受アプリでパーソナライズされた情報提供サービスを始めている。また、モバイルやGREEのような携帯SNSにおいてアバターサービスが流行しており、アバターを起点にパーソナルな情報サービスを提供する動きが始まっている。

最後に、石塚教授からは、ホットリンクの技術を理解していただいたうえで、必要となるシステムの発注をいただいた。実は、創業後間もないベンチャー企業にとって、発注していただけるというのは、涙が出るほどうれしいことだった。さらに、東京大学と取引があるという実績を外部に示せる。本当に助けていただいたという感謝の思いがある。

2・3 京都大学西田教授との出会い

京都大学西田豊明教授(当時は東京大学)は、今でいうWeb 2.0の中心的な考え方である一般ユーザの知識からコンテンツを生成するという考え方を学術的に捉え、実際にサービス化し実証するという取組みを行っていた。その一つの形が、画像付き掲示板から放送番組を自動生成するPublic Opinion Channelというサービスで[畦地 01]、今でいうニコニコ動画の原型のようなものだった。認知心理学的視点も含めた総合的な研究であり、さまざまな知見を学ぶ機会があった。当時、西田教授は、アイデアを実サービスとして実装し、検証し、さらに研究にフィードバックする、というサイクルをつくっており、企業が出る幕がないほどであった。

2003年後半は、まさに日本でBlog文化が芽生えるタイミングであり、一般ユーザがネット上に自分の知識を吐き出すという、POCの概念を実現する文化が浸透しはじめるときであった。そのとき、私はNTTデータと共同でDoblogというBlogポータルサイトの企画を行っていた。Blogというサービスは、POCのような知識循環を引き起こすための知識の入口だ、という明確な思いがあった。Doblogの開発初期には、投稿されたBlogの画像を自動でめぐりながら、画像の下に投稿されたBlogのテキストがテロップで流れる「シーン」という機能を実装したが、これはPOCの考え方をヒントにしたものであった。

POCでは、一部の限定されたユーザに協力してもらう形で研究成果の実証をしていたと記憶しているが、ネット文化が普及した今であれば、企業を巻き込み、共同で洗練された企画とインタフェースに仕立て上げ、インターネットで一般ユーザに公開しながら、研究と実証のサイクルを回すということを大規模にできるだろう。

2・4 Blog分析サービスの事業化

現在、ホットリンクでは、「電通バズリサーチ」と「クチコミ@係長」という二つのブランドで、クチコミ分析

のASPサービスを提供している。このサービスにはさまざまな研究機関での知見・技術が組み込まれており、その中のいくつかをご紹介します。

§1 大阪大学松村氏によるBlogからの性別判定技術

マーケティング担当者には、リサーチを行う対象のデモグラフィック属性を知るということは当然至極のことのように、Blog分析においても、ブロガーの属性を判定したい、というニーズがあった。そこで、2ch掲示板内の語の伝播研究で有名な大阪大学の松村真宏氏に依頼して、Blogの書込みからの著者の性別判定エンジンの開発を協力して行った。このときは、NTTデータと共同運用していたBlogポータル「Doblog」のユーザ5万人のプロフィールを利用して、25,000件のユーザプロフィールをもとに、サポートベクターマシン(SVM)で男女別の利用キーワードを学習させ、残り25,000件のデータで検証し、確信度が一定値以上のBlogに対して90%以上の精度を上げた。このエンジンは、大きな変更の必要もなく、現在のサービスに組み込まれ、ニュースでも多く取り上げられた。

このコラボレーションは、非常にスムーズに行われた。研究の前提となる学習用データをホットリンクが大量に保有し提供できたこと、研究者にやってもらいたいことが明確であったこと、松村氏が研究を実現させるための道具をすでに保有していたことなどが、理由にあげられると思う。

§2 総合研究大学院大学の竹田隆治氏による

スパムフィルタリング技術

ホットリンクのスパムBlogフィルタには、総合研究大学院大学博士課程の竹田隆治氏の技術を一部使わせていただいている。きっかけは、竹田氏から、ホットリンクの会社HP経由で技術をアピールしていただいたことだった。ほかのBlogポータルにも売込みをしていたが、いろいろな理由で採用されなかったという。

竹田氏とお会いして、半端でない量のBlog記事を自分の目で読み、スマートな方法のみでなく、地道な手作業も組み合わせる研究開発していることがわかった。そして、その成果をWebで公開しながら精度の検証をされているところに、企業の実サービスを構築する際の泥臭さをわかっている方と思えた。

スパム排除技術はたちごっこ側面がある。だから、研究されている技術をそのまま企業に渡しても、時間が経つと精度が落ちてしまうのは当然である。アルゴリズムというよりヒューリスティックな作業もかなり必要である。また、ホットリンクが保有している技術やノウハウと融合させれば、さらに精度が良くなることも予想できた。

そこで、竹田氏には、単に技術を提供してもらうという形ではなく、より良く進化させていく地道な共同作業と一緒にやってもらえるかをお願いし、了解いただいた。実サービスに組み込む際に、竹田氏の技術は、論文には

なっていないノウハウが非常に多く、メンテナンス性を高めるため、実際のソースコードとその解説書をもとに、すべてリファクタリングし組み込んだ。

この技術は、これまではホットリンクの Blog 分析サービスで利用されてきたが、現在、ほかの Blog ポータル事業者に対してスパムフィルタ機能を API サービスとして提供するという取組みが始まりつつある。それがスタートすると、Blog ポータル事業者の監視チームからスパム判定プログラムの誤判定の結果が日々報告されてくることになる。その段階で、「研究者からの技術提供」→「企業によるサービス化」→「実サービスから研究へのフィードバック」という一連のサイクルが継続的に行われる段階に至ると期待している。

2.5 研究機関へのデータ提供

企業が研究機関から技術提供を受けたり、共同で研究する形とは別に、企業側から研究機関にデータ提供を行い、研究者が自主的に研究を行うという形の取組みもある。

東京大学石塚研究室の修士に編入した古川忠延氏には、NTT データと共同開発した Doblog ポータルにおけるユーザの書込みデータやユーザの閲覧行動データを提供した。Blog 上の情報の伝播のしくみの研究につながり、国際会議の論文としても発表されている。また東京工業大学の高安美佐子研究室や筑波大学の香田正人研究室にも Blog などのデータを提供しており、クチコミと売上との関係など、クチコミを活用した新しい研究を行っている。

とはいえ、必ずしもうまくいってばかりいるわけではなく、ご迷惑をおかけした例もある。2008 年、国立情報学研究所の神門典子教授より、NTCIR のテキストマイニングタスクのための Blog データの提供を打診されたときのことだ。神門教授より最初に話があった際には、私は二つ返事で問題がないとお話した。しかし、いざ提供する直前になって、そもそもユーザに著作権のある Blog データをホットリンクが第三者に提供して問題がないのか？ というチェックが法務部から入ってしまった。そもそも現行の著作権法からすると Yahoo! や Google のような検索エンジンでさえ違法（もしくはグレー）になってしまうので、何をいまさら、という部分はあった。しかし、個別に大学の研究室にデータを提供するのは異なり、NTCIR を通じて大々的に Blog データを提供する、というのはかなりセンシティブな問題だった。結果的に、提供することができず、神門教授にはご迷惑をおかけした。

NTCIR のような難しい例もあるが、実は、我々企業としては、サービスから得られたデータを研究機関に提供することは大きな負担ではなく、逆に、企業内ではできない研究をやっただけという面で非常にうれしい。

データを提供してもらえるのだろうかと一歩を踏み出せていない研究者の方も多いと思うが、積極的に企業に

アプローチしてみられることをお勧めしたい。研究成果を積極的に開示さえすれば、共同研究や実サービスへの組み込みの道もおのずと開かれると思う。

2.6 東京大学松尾 豊氏、産業総合研究所濱崎雅弘氏とのレコメンド ASP サービスの事業化

2007 年の夏、松尾氏と米国の有名なエンタープライズ Web 2.0 企業を訪問中、ホットリンクがレコメンド技術をもっているというのなら動くものを見せてみる、といわれたのがきっかけだった。「だったら見せてやる」ということで、帰国してすぐに世界最大のソーシャルブックマークサービス「del.ici.us」のデータのクロールをホットリンクで行い、そのデータを利用した Web ページのレコメンドエンジンを、これまで社内にあった技術とは別に、新規で東京大学松尾氏、産業総合研究所濱崎氏、東京大学の古川氏で研究開発した。結果的に、その米国企業にレコメンド技術を提供することにはならなかったが、日本国内で ASP サービスとして事業化することにした。

ほかのケースとは異なり、もともと研究機関に技術があったわけではなく、我々からの依頼に基づき、さまざまな学術研究をリサーチし、新規に研究開発したというケースである。

実際に事業を進めていくと、競合と差別化して戦っていくには、アルゴリズムの精度のみならず、大量のアクセスをどう分散処理するか、というシステム構築面での洗練が重要になるということがわかってきた。当初は、サービス化のスピードを重視し、研究時点でのソースコードをそのまま活用してシステム化した。その後の事業展開に伴い、順次リファクタリングしていった。

実サービスとして運用していく中で、さまざまな改良ポイントが見えてきて、実際に導入した企業の商品データ、ユーザの閲覧・購買履歴、レコメンド有無の結果など、さまざまなデータを研究チームにフィードバックし、アルゴリズムの改良研究を行ってもらい、ホットリンク側ではシステム面での改良を進めている。

この分業体制は非常に効率的であるが、さらに一歩進めて、データの提供のみならず、実際のソースコードも研究チームと企業側で共有した。これにより、研究チームからの改良されたアウトプットが即時実サービスに反映させられるようになった。ただし、今後研究チームに所属する学生メンバーが、競合企業に就職する場合の機密漏えいなどのリスクは課題として残っている。

3. 今後の人工知能技術のビジネス活用に必要なこと

研究機関の技術が実サービスとして利用されるまでには、多くのケースで研究機関と企業とのコラボレーションを必要とする実践研究という段階が必要となり、ま

た企業側では周辺システムや大規模化などのシステム構築、サービス企画、インタフェースデザインなどが必要になる。

3・1 研究機関側での課題と努力

ところが、通常研究機関では、研究成果を企業に利用してもらいたいが、「そこから先は大学がやることではなく、企業の役割です」という考えで、純粋研究を超えて泥臭い実践研究の部分まで足を踏み込むことを嫌いがちである。その場合、企業がその研究成果を自分達で積極的にサービス化することはかなり困難だと思う。

レコメンド事業や、スパム Blog 排除技術のところでも記載したが、「研究」→「実サービス化」→「研究へのフィードバック」というサイクルを『一度』構築してしまえば、その後は純粋な研究に戻れ、かつ実サービスから生のデータやニーズが得られ、それをもとに改良した研究成果をすばやく実サービスで再検証できる。研究がより効率化されるサイクルが『永遠に』回るのである。ぜひ、その一歩を踏み出していただきたい。

3・2 企業側での課題

では、研究機関側が一步踏み出せば、すべてうまくいくのだろうか？ 実はこの場合であっても、まだ問題は解決しない。なぜならば、企業側に研究機関の研究内容を理解し、研究者とコラボレーションを行えるだけのできる人材が極めて少ないからである。ネット業界はまだ生まれてから十数年しか経っておらず優秀な研究者が就職したいと思うような成熟した企業が、まだ多く存在していない。したがって、これまでに新興企業に就職した研究者の数は非常に少ないのである。一部少数の研究者が Yahoo! や楽天やサイバーエージェントに行き、また最近の1~2年でようやく Google やミクシィなどに就職する人が出てきたくらいである。さらに、それらの企業が、積極的に先端研究をサービスに組み込む文化をもっているかというところに道が狭くなる。

3・3 研究機関側および企業側の努力

この問題を解決するには、企業側に、もっと研究を理解できる人材が必要である。そのためには、研究者がネット新興企業に就職するという流れが必要だろう。それによって、実践研究を共同で行う下地が整うはずだ。

最近では、企業と組まずとも、研究機関自身で、実サービスまでを構築するケースも出てきている。この動きを加速するには、研究者が一度ネット企業や IT 企業に就職し、プログラミングやサービス企画の能力を身につけて、再度大学に戻る、というステップを踏むことが良いのでは？ と考えている。たかだか『数年』の修行で、『一生』の研究を効率化させられる武器を身につけられるのであれば良い選択肢かもしれない。

また、別の解決策としては、研究機関のアルゴリズム

を理解しつつ、実サービス構築のためのシステム構築やサービス企画を発注できる、専門の請負企業が存在してよいのではないかと思う。そうすれば、研究機関は予算さえ獲得できれば、自分達が主体でありつつ、実サービス化までの本質でない部分をアウトソーシングできるようになるだろう。

インターネットサービスの場合、研究室の研究成果をそのまま企業が利用できるというケースよりは、その知見をもとに、実サービスや実データに照らし合わせて実際に適用する際の実践研究が必要になるケースが多いように思う。そういう意味では、人工知能技術が問題ではなく、技術が必要とされるマーケット環境ができていくのかというところは、技術活用の大きなポイントになると思われる。

4. 最 後 に

Web 2.0 サービスが一般ユーザに普及し、人工知能を学習させるための元データが大量に手に入る環境になった。また、同時に、ネットビジネスも成熟期に入ってきており、単にアイデアや営業力のみでビジネスとして勝てるフェーズは終わりを告げつつある。そういう面で、今がまさに、人工知能技術がビジネス分野で必要とされ、活躍できる環境が整ったタイミングである。

この原稿を書く機会をいただいたことで、ここ10年近くの自分自身のサービス開発の歩みを「人工知能技術のビジネス活用」という視点で初めて振り返ってみた。それによって、日本は、研究機関側も企業側もそのチャンスをものにするにはまだまだ未成熟かつ発展途上であるということをはっきりと認識した。とはいえ、確実に改善の芽は生まれてきている。

今後のネット業界の競争は必ず世界レベルで戦わなくては行けない。日本の研究者の方々の研究レベルはけっして世界と引けをとらないと思っている。その技術が、実サービスとして世に出され、世界と戦えるサービスになるように、本稿が少しでも一助となれば、幸甚である。

◇ 参 考 文 献 ◇

- [畦地 01] 畦地真太郎, 福原知宏, 藤原伸彦, 角 薫, 松村真宏, 平田高志, 矢野博之, 西田豊明: パブリック・オピニオン・チャンネル: 知識創造コミュニティの形成に向けて, 人工知能学会誌, Vol. 16, No.1, pp. 130-138 (2001)
- [筒井 00] 筒井貴之, 石塚 満: キャラクタエージェント制御機能を有するマルチモーダル・プレゼンテーション記述言語 MPML, 情処学論, Vol. 41, No. 4, pp. 1124-1133 (2000)
- [内山 06] 内山幸樹: 1時間でわかる図解 Web 2.0, 秀和システム (2006)

2009年5月22日 受理

著者紹介



内山 幸樹

1995年東京大学大学院在学中に株式会社マジックマウスの設立に加わり、日本最初期のサーチエンジンの開発に携わる。1997年株式会社マジックマウス(現デジット株式会社)入社。1998年デジット株式会社常務取締役就任。2000年6月、株式会社ホットリンクを設立し、代表取締役役に就任。検索エンジン、ソーシャルブックマークサービス、mixi ツールバー、

Blog ポータル「Doblog」、Blog 分析サービス「電通バズリサーチ」、「クチコミ@係長」、レコメンデーションエンジン「レコナイズ」など、Web 2.0 的先端サービスの開発に多数関わる。著書に、「仮想世界で暮らす法 (ブルーボックス)」講談社 (2007)、「1時間でわかる図解 Web 2.0」秀和システム (2006)。