

情報通信審議会 情報通信技術分科会
技術戦略委員会 次世代人工知能社会実装WG 第2回会合
議事概要(案)

1. 開催日時

平成29年2月14日(火) 10:00~12:00

2. 場所

総務省11階 第3特別会議室

3. 出席者(敬称略)

主任: 柳田 敏雄

構成員: 東 博暢、麻生 英樹、荒牧 英治、池田 尚司、上田 修功、宇佐見 正士、臼田 裕一郎、
栄藤 稔、大竹 清敬、加納 敏行、川鍋 一晃、松永 圭司(栗本構成員代理)、小林 哲則、
鳥澤 健太郎、原 裕貴、春野 雅彦、本田 英二、前田 英作、森川 幸治、山川 宏、

ゲスト: 盛合 志帆(NICT)、益子 信郎(NICT)、

オブザーバー: 内閣府 中川 実務研修員、文部科学省 栗原専門官、経済産業省 植木企画官、
事務局(総務省): (大臣官房)

武田官房総括審議官、
(情報通信国際戦略局)

吉田参事官、

技術政策課/野崎課長、山口企画官、寺岡補佐、

研究推進室/越後室長、出葉推進官、中川補佐、皆川補佐

4. 議事概要

(1) 第1回WGの議事概要の確認について

事務局より資料1に基づき第1回会合の議事概要(案)について説明があり、修正等がある場合には、2月17日(金)までに事務局まで連絡することとなった。

(2) 構成員等からのプレゼンテーション

鳥澤構成員から資料2-1、NICT盛合様から資料2-2、麻生構成員から資料2-3、荒牧構成員から資料2-4に基づき、それぞれ説明が行われた。

それぞれの説明における質疑等は以下のとおり。

■ 鳥澤構成員プレゼンテーション

栄藤構成員: NICTのデータをさわるにはNICTの肩書が必要だが、例えば私は兼業推進派だが、民間企業の肩書をもったままNICTで兼業できるシステムは考えられないのか。完全に出向だと、そちらに行ってしまうわけで、どちらかというサービスがわかっている人間やサービスに使いたい人間が、うまくNICTのポジションを兼業でやられて、どちらもウイン・ウインになれるという、そんなストーリーが結構あると思う。

鳥澤構成員: 現状そういった事例はあまりないかと思うが、検討させていただきたい。おそらく大丈夫だと思っている。

柳田主任：NICTのどなたか、事務的に答えることができるか。

益子様：それはすでにやっているチームもあるので、可能である。

原構成員：企業の立場から言わせていただくと、鳥澤先生がおっしゃっていることは、ほとんどそのとおりだと思うが、やはり学習済み、ラベル付きのデータがない。すごく単純な発想だが、お金でそれを解決することはできないか。つまり、鳥澤先生がおっしゃるようにデータ作成者と自然言語処理研究者のモチベーションがずれているならば、そもそもデータをつくるころにお金を投入して解決することはできないだろうか。ただ、その場合、税金を投入するのに特定のシステム向けだと批判が出るかもしれないが、特定のシステムに限らず汎用的に使える学習済みデータを増やして皆で広く使うようにすることはできないか。

鳥澤構成員：要するに今まで我々もお金は投入してきたわけだが、それはあくまで我々が最終的につくらなくてはいけないシステムのための学習。そういう形なので、その学習データをそのまま皆さんが持って行って、すぐに皆さんのサービスに直結すると言われると、ちょっと距離があるかもしれない。もちろん、我々のサービスをそのまま引き継いでいただけるというようなことであれば、それはそのまま使えるということになるわけだが、一方でお金を使う。だから、我々は実際お金を使ってデータをつくってきたわけで、今おっしゃりたいのは、要するに皆さんが共有できるような学習データをお金を使って作れないかということかと思うが、どうお答えするのがいいのか。そもそも、皆さんが使えるということで、仕様に合意ができるかどうか。要するに総花的なデータというのは、えてして使えない話で。申し上げにくいけれど、過去にもそういうプロジェクトがあったかと思う。なので、もちろんお金があれば、何でもできる、というのがお答えだが、ただしそういうお金の使い方でお金を出してくれる方がいるのかと言われると、それはいないのではないかな。

原構成員：特定のシステム向けであればデータも作りやすく活用しやすい一方で税金を投入すると批判も受ける。反面、様々なデータを作って汎用的に皆が使えるようにすることを目指すとかって使いにくいデータになったりする。このバランスをどう取るかだろう。かなり合意のとれた医療用の正解を入れたセットのようなものであれば、それに税金を投入するロジックも成り立つかもしれない。

鳥澤構成員：医療用だと荒牧構成員はどうか。

荒牧構成員：データをつくるというところ等が先ほどから話題になっているが、つくるのは非常に大変だと思うが、既にもうラベル付きの教師データがあるのではないかな。こんな症状の人はこんな病気を起こすという診断のようなものが日々たまっていくし、例えばそういうデータをつくるのは非常に鳥澤先生がおっしゃるとおり大変だが、少し発想を変えたら、既に実は持っているのではないかなというのがワーキンググループで見つけられればいいなというのが私なりの希望である。

鳥澤構成員：それは、私のプレゼン資料でいくと2種類のデータというお話をさせていただいたと思うが、このデータのタイプというところで、生データで、機械学習等でそれだけでいいような場合はまれということであるから、そのまれなケースがあるかもしれないということ。それはぜひこのワーキングで発掘していただければ、皆さんハッピーなのではないかと思う。

加納構成員：鳥澤構成員の資料の中で、一番最後のパラグラフのところだが、「NICTでは、サービス、アプリケーションの開発を行う研究者が学習データの構築法、設計まで考える」というところが、私ども企業にとっては非常に悲壮感の漂う世界でもある。つまりもしかすると

データサイエンティストを育成していくという努力、今までの方向が、間違っているのではないかと思ってしまうところがある。どういうことかということ、中立なデータサイエンティストというのはもう役に立たず、今実際にビジネスをやっていたりサービスを提供したりしている人がデータサイエンティストにならなければ、実は機械学習の学習データというのは作っていくことができないという風に思えたのだが、そのあたり、鳥澤さんのご意見をいただきたい。

鳥澤構成員：そこまで悲壮感という意見は持っていないが、やはり垂直に皆がコミュニケーションをちゃんととって、作業を行うというのは必須だと思う。では、純粹にデータサイエンティストというのが不要かと言われると、そこまでは考えていない。実際、我々は言語学者を雇用している。データをつくるということになるとどういうことになるかということ、要するにプログラムをつくったりアルゴリズムを考えたりする研究者と、その言語学者と、さらに実際にラベル付けをする作業者と、その3組の人間がよくコミュニケーションして作るということになる。やはりいかなるアルゴリズム、プログラムを考える人間でも、どういうふうマニュアルをつくったら、皆が欲しいデータをつくってくれるかといったところまではなかなか気が回らないのが現状。そういった意味で、一種のデータサイエンティストは必要だと思うが、一方で、それではそういう人だけ育成すれば何とかなるかということ、それはならないと思う。

■ NICT 盛合様 プレゼンテーション

鳥澤構成員：NICTの中で議論すればいいかもしれないが、ちょっとわからないので教えてほしい。6ページのこの例だが、これはデータの中でBMIと血糖値の具体的な数値がどこに書いてあるのかというのがわかっている状態で始めるのか、それともどこにBMIや血糖値が書いてあるのかという話も暗号化されていて普通にはわからないという状況で始めているのか。

盛合様：これはデータフォーマットがわかっている状態である。

鳥澤構成員：データフォーマット、自然言語処理への適用を考えると、データフォーマットというのはあるような、ないようなというような状況である。そういった状況でもこれは使える可能性があるのか。

盛合様：まだ今やっているのは、いわゆる構造化されたデータに対してなので、自然言語に対しては、どこか暗号化されたまま解析をするというところで使えるかというのを検討する必要があるかと思う。汎用的というよりは、問題に合わせてカスタマイズしていくような話になる。

鳥澤構成員：これがうまくいくとすると、さっき私が申し上げたようなデータの共有や、ちょっと著作権はクリアできないかもしれないが、そういったところでは、かなりプラスになるのかなという認識でいる。

もう一点質問だが、7ページの次のスライドについて、これはニューラルネットワークのアーキテクチャーは皆共有されている、皆知っているということか。

盛合様：その通りである。最初にニューラルネットワークの構造は共有する。

鳥澤構成員：なるほど。ちょっと馬鹿な質問かもしれないが、おそらく皆さん競争していく中で、独自のニューラルネットワークアーキテクチャーをつくっていて、データはみんな共有したいけれど、そこは皆知らせたくないというような話があるかと思うが、さすがにそういうの

は無理なのか。購買の情報を共有するとなると、そもそもどの購買を共有するのかわからないという話になると思うので。

盛合様：ここでやっているのは、もともと NIPS2012 で提案された方法の改良で、事前にニューラルネットワークを皆で共有して、皆で学習していく、しかし自分のデータは漏らしたくないというようなモデルであった。そこをプライバシーを保護する形に改良した方式になっている。

鳥澤構成員：分かりました。

栄藤構成員：やはり何が犠牲になるか、時間やメモリーがどういうふうに犠牲になるかというのをちゃんと見きわめた上で、そういう懸念はあるけれども、やはりなかなかデータが出てこないという日本独特の問題を克服するというか、こういった研究は非常に重要だと考えている。特にそういう医療系とか金融系、むしろ政府の情報、そういったものをうまく扱える仕組みが出てくると、すごくいいなという期待感がある。

盛合様：ありがとうございます。今回もCRESTのプロジェクトを進める上で、ほんとうにまれな例だと思うが、具体的にデータ提供を検討していただける銀行があり、そこからデータをいただく予定である。最初にご紹介したように個人情報保護の壁があり、今回は3ページにもあった「第三者への提供」の2番目にある委託という形であれば、銀行の業務の一部としてデータをお預かりして、我々のほうで解析させていただくという整理である。そういう枠組みでまず研究をスタートし、その後、そこでいい成果が出れば、例えば暗号化したまま、よりデータ漏洩した場合のリスクも下がった状態でデータが活用できるということで、その先の利用につながっていけばいいなと考えている。

上田構成員：質問だが、3ページで、第三者への提供というところで、匿名加工情報とある。これも定義もやや曖昧ではあるが、個人が特定できなければよいということで、そういう動きで今データが共有できるようになってきている中で、先ほどのプライバシー保護データ解析技術、6ページだが、それなどは個人が特定できていない状況になっていたら、別に普通に分析をすればよい話で、こういうのが本当に必要な場合は、糖尿病ということを分析して個人にちゃんとフィードバックできるような状況の応用だったら価値があるけれども、これはどういう使い方になるのか。また、プライバシープリザービングデータマイニングというのは大昔からあって、ほとんどそれはデータのシェアリングによるものなので、あまり使えないということで一旦ポシャったのだが、また暗号化技術ということで、多分暗号化であれば問題ないと。ただ暗号化というのは、本当に個人が特定できない、誰にもわからない状況であるから、その辺全体を見たとき、何か一貫していないなという気がする。その辺どうなのか。

盛合様：暗号化したまま分析の結果は、復号すれば個人がわかってしまう状態になる。匿名加工情報だと個人の情報ではなくなってしまうので、わからなくなってしまうため、その分析結果を個人に戻すということはできない。こちらの暗号化したままでデータ解析をするというほうは、分析結果がきちっともとに戻せるので、用途が違うのではないかと考えている。

上田構成員：だから暗号化の場合は、分析するときも完全なデータがわからない。

盛合様：他者には漏れないけれども。

上田構成員：データが流通するときになっていくと、だから暗号化による技術が重要で、いわゆるシェアリングみたいなものではないと。わかりました。

野崎課長：せっかくの機会なので、事務局から1つだけ教えていただきたい。今NICTはCinetでfMRIを使って脳画像を集積していこうとしていて、そういったfMRIの画像は

様々な分野で利活用が期待されている。そういう脳画像を集めて様々な産業応用を進めていく上で、私という人間の毎年の脳画像を撮って、経年でどう変化しているかということを追っていくことが重要である。私と画像のひも付けを切れれば、匿名加工情報として、私の同意を得なくても自由に産業利用、第三者の利用に使えるのか。脳画像というのは、匿名加工情報にするにはどうすればよいのか。

盛合様：とても難しい、まさに議論中のところかと思う。匿名加工情報も、例えばいわゆる銀行の取引データや購買履歴のような表型のデータ、位置情報、顔画像情報、脳画像のようにジャンルによって、匿名加工の方法というのは全く違うと思われるし、それもまさに個人情報保護委員会のほうで事務局レポートを書いて、どんどん例を入れ、それから J I P D E C のほうでも例を積み重ねていこうというところになっていると思う。匿名加工は一意の方法ではない。その基準のほうも委員会等で議論されているところ。またその方法がほんとうに安全なのか、個人のリスクはないのかということ、学会でも並行してやっているところは今日ご紹介したところである。そのため、これからということがかかなりあるかと思う。また、脳情報のほうだが、まず個人に同意をいただければ、ある程度活用できるかと思う。そうでない場合は、匿名加工情報に加工して、第三者への提供なども含めて利活用するという形になるかと思う。医療機関で利活用するで一たについては、結構同意をいただいていることが多いと聞いている。

野崎課長：最初から、第三者提供の同意をもらっていただければいいということか。

盛合様：その通り。

荒牧構成員：自然言語文の暗号化は匿名化されているのか。例えば、京都の再生医療の第一人者といったら、何となくあの先生かなと思ひ浮かぶ、そういった普通の人ならば推測可能なこともあると思うが、そういった自然文に対する暗号化とは何か、現状はどうか、教えていただければと思う。

盛合様：今の議論の中に、そういうのに関してはないような気がしている。ただ、データとして特異なデータは外しましょうというような処理は入っているが、今の発案に関してはわからない。

■ 麻生構成員 プレゼンテーション

鳥澤構成員：幾つか異なるというか、相矛盾するアプローチを同時並行でやっておられるというふうに理解したのだが。

麻生構成員：あれはうちではなく、世の中である。

鳥澤構成員：端的に言うと、テキストと画像と結びつけるというと、要するにテキストの内容は数百次元のベクトルであらわされるというような、昔構文解析の研究等をしていた人間からすると、ふざけるなという感じだが、それが非常にポピュラーで、実際やってみるとそのほうが性能が出るという、ある種悲しい現実があるわけだが、この Semantic Parsing というのは、学会ではやっているのはもちろん承知しているが、逆行する流れである。もちろん、だからいずれ 10 年後、20 年後にはやっぱり Semantic Parsing したほうがよかったですねと、ニューラルネットも同じような過程を経て復活しているので、今はどちらとも言えないと思うが、その辺はどうなのか。産総研としてはどんなふうに考えておられるのか。

麻生構成員：産総研としてと言われると困る。私の考えとしては、結局今の段階では、どちらがいいというふうな結論は出ていないということで、もちろん言語は非常に構造的な対象だから、最終的には構造が重要であるというふうな考えで研究されている。でも、そうはいつでも人間が使う範囲の言語であれば、その複雑さの上限みたいなものがあるので、ニューラルネットワークを使って、高次元の空間に意味を埋め込んでしまえば、その構造も何かよくわからないけれどもコーディングされてしまう。今現在、ニューラル機械翻訳がかなり成功しているというところは、そういう要因もあると思うけれども、結局どっちというのは振れ続ける振り子のようなもので、決定はつかないような気もしている。さらに鳥澤構成員もおっしゃるように、どっちがいいかというのは応用にもすごくよると思うので、今の段階でどっちだというふうには決めずに、相矛盾するというのはそのとおりだが、どっちが優れているという段階だというふうに。ただし、両方をハイブリッドしようというふうな話もちろんあり、構造的な表現の中の末端の表現をヒューリスティックに書くとか、そういうふうな話も提案はされている。

荒牧構成員：私なりの解釈だが、こういった Semantic Parsing 自身、社会環境に依存するものではなくて、医学の分野でいけば論文とかは1回しか書かれなくても非常に大事だけれども。1回きりのきちんとした情報といったものは、Semantic Parsing できっちり押さえて、もっと大量な事例などは、ニューラルネット的な流れでふわっと捉える。そういった使い分けみたいなものを社会実装ごとに理想的なつけ方を必要かと思っている。

■ 荒牧構成員 プレゼンテーション

栄藤構成員：大変興味のあるお話だった。一番気になるのは、よくAIで画像の読影医は要らないとか、内科の問診ももう要らないとなっているが、今のところ、まだ実用化になっていないのは、どのあたりに課題があるのか。

荒牧構成員：画像の読影医が要らないといったことは、多分次からいけば、近いうちにそういうふうな浸透をして、要らなくなるのかなと思っている。まだ現状の精度などまだ不十分かと思うが、そうなりつつある方向の途中という感じか。

栄藤構成員：それで、問診のところも医者が要らないというのは極端な意見で、それはかなりいろいろな法律の問題があるので難しいが、ある程度の、例えば看護師のサポートだとか救命士のサポートだとか、そういったシステムができると思えるのに、なぜまだできていないのか。

荒牧構成員：質問の意図を勘違いしていた。まず読影医などは要らなくなると思うが、そこに残される唯一の大事な仕事として、問診で直接患者さんに対応するといった部分が出ると思う。すなわち人工知能的に何か疑われるとして、それについて深掘りして見る。そういった能力は人間でないとできないかと私なりに思っている。

栄藤構成員：それでは内科医のある程度のサポートは難しいということか。

荒牧構成員：疾患がいろいろある。サポートはできるが、代替は無理だと思う。

栄藤構成員：代替のそういった本質の話をしているわけではなく、ある程度症状を見て、今、これはインフルエンザの確率が80%で、胃腸の感染症がこれぐらいでなど、そういったことを言えるシステムができると思ったのだが、それはできないのかと聞いている。

荒牧構成員：それはできると思う。

栄藤構成員：質問は、それができていない何かその大きな課題というのは何なのか。

荒牧構成員：それは時間の問題で、近いうちにできるものだと思う。

柳田主任：多分時間も長かったり遅かったりするのですが、ここは一応総務省でも良いけれども、加速するためには何が必要なのかという質問でもあったように思う。今、何が足りないから5年後になっているのか、これをすれば1年後になってしまうなど。

荒牧構成員：現状、医療とAIが非常に注目されて、協力体制もとりやすいが、それでもこのぱっとすぐできないといったものには、やはり今の医療の情報のデータが島状にあるといった状態だと思っている。それを、ある程度の数をそろえてやれば、戦国時代みたいにぱっとまとまるのかなと思っている。その過程かなと思っている。要領を得ない回答になってしまったが。

鳥澤構成員：今の質問とも関係しているが、細かいということで飛ばされた9枚目のスライドがあるが、症状を特定と症状を標準化というのは、ターゲットは単語なのか。それとも例えばくじくと頭が痛いとか、何かそういうフレーズまでカバーしているのか。

荒牧構成員：これはフレーズはカバーせずに、複合名詞レベルの、単語ではなくて、名詞句レベルになっている。

鳥澤構成員：名詞句レベルということか。わかりました。それで、先ほどの栄藤構成員の質問とも関係している、もっと過激な質問かもしれない。2枚目に、要するに製薬会社頑張れみたいなスライドがあったが、こういう現状から製薬会社の売上げがブーストする、あるいはほかのスライドでもあったけれども Watson に対抗するみたいなどころに至るまで、一番障害となっているのは何だとお考えか。あるいは、そこは目指していないと言うかもしれませんが。要するに今後医療AIを実用化して皆さんに使っていただいて、皆さんハッピーになってもらうために一番重要なこと、一番解かなきゃいけない問題は何なのかということを書いていただくと、議論が深まるんじゃないかと思う。

荒牧構成員：わかりました。先ほど言ったように、島状に今データが集まっているので、その島ごとにしたいことが違っているということ。医療全部をできるというふうに言うと、ちょっとぼやける感じがして、それぞれの持っているデータごとに自治医科大学のしたいこと、東大病院のしたいことといったような理由があり、それに合わせたゴールをともにつくっていくことによって、そこがまとまると。それを中程度のゴールといったものが、多分今IBMの Watson とかほかのAI とかに欠けているのかもしれないなと思っている。

鳥澤構成員：ということは、結局はやっぱりデータが必要だと。現状いろいろなところから集めておられるということだが、それも集まってハッピーなんだけれども、まだ問題がいろいろあると。そういう認識で良いか。

荒牧構成員：我々のところは集まってハッピーだが、もっといっぱいデータはあるはずなので、もっとここに参入する余地はあり、いろいろなところで同時多発的な中規模のAIみたいなものができるような形なのか。

鳥澤構成員：ただ、そこではさっきおっしゃられていたようにそれぞれデータの由来も目的も違うので、それでは全部一気通貫でそれを集めて学習したりしまえという話にはならないと。

荒牧構成員：というのは非常に荒っぽい、そこが今の問題だというふうに思う。

鳥澤構成員：さっきも私も同じような質問を受けたけれども、要するに日本全体で共有できる医療データみたいな、ある種の目的に向かってそれを大量に集めるみたいな話というのは、現実的なのか。

荒牧構成員：それはこちらとしてナショナルデータベースがあるが、しかし、言語処理というふうな観点から言うと、ナショナルデータベースで扱っているのは、そういった言語は含まれていないし、もう少し言語が含まれているならば、実際もう少し臨床寄りだと思っている。そこは国レベルで当たるところにはそういった言語が入っていないので、そこはちょっと言語処理と医療という組み合わせがちょっと違うかなと思っている。

鳥澤構成員：では、今現状、医療系の臨床に関する言語データというのを、何かある種校閲的な観点から集めるというような試みは、現在はされていないと。

荒牧構成員：していない。

鳥澤構成員：される見込みというのはあるのか。

荒牧構成員：私が質問したのにも近いけれども自然言語文といったものは、人間が読むと結構わかってしまう部分が多いので、自然文のままではうちの施設からは出せないといったところが結構多いと思っている。

鳥澤構成員：なるほど。今のは大変おもしろいと思った。どうもありがとうございました。

(3) 意見交換等

加納構成員：鳥澤構成員にした質問と同じ質問になるかもしれないが、今の荒牧構成員のお話の中で、やはり学習データとか医療診断にデータを使っていこうとすると、医療や診断を熟知し、かつ経験のある人間がデータサイエンティストとして働かなければいけないという事情は、やはりかなり強くあるというふうにお思いか。

荒牧構成員：私は、今非常にそれを強く思っている。サンプル1の例だが、よく、ある医療分野の大御所の先生が言われるには、病院の先生は情報もそこそこ勉強していると。しかし、情報の先生は、医療はあんまり勉強していないのではないかと。そういったデータには興味があるが医療には興味がない方もおられますので、そういったケースも多いのかとっております。

加納構成員：ありがとうございます。今で感じたことを少しコメントで入れさせていただくと、情報通信に携わる企業。おそらく教育も研究も今まで水平分業という形で進化してきて、それぞれ個別に水平分業の水平の層の中で研究を推進してきたわけだが、ここに来て、どうも先ほど一番冒頭に鳥澤構成員がおっしゃったように、垂直統合というところがだんだん見えてきて、どうもICTそのものが大きく変わらなければいけない何か変化点に来ているのではないかなというふうに少し感じた。

柳田主任：いろいろ3センター長と話をするけれど、おっしゃるように情報系の先生、大大学者の先生は、文字列には興味があるけれど、そこに含まれているメカニズムにあまり興味がない等。そういう事態もあるのは確かなので、おっしゃるように、もうメカニズムなり仕組みなりも理解しながらデータの処理をする、というセンスを皆に持っていかないといけないだろう。そうすると、人材育成か。頭のかたい年寄りにそれを求めるのも無理ということもあるかもしれない。

柴藤構成員：自然対話のアプリケーションとしては、やっぱり医療というのは一番、一丁目一番地ですごく重要だと思っている、そういった症例データがちゃんと集まるということと、それに対して推論して、診断するというのは、かなり遠隔医療とか通信の絡む話で、重要かなというふうに思う。医者の方と話す、我々情報屋が思っているみたいに、if then elseで診

断できるようなそんな教科書がないようで、情報系モデルがあって、あと観測しているシンボルはあるんだけど、その間の理詰めが全然、医者 of 全く暗黙知でやっているという話らしい。間違っていたら直してほしいが。そこら辺をいわゆる情報系、情報通信系が埋めることができる結構良いと思う。先ほどでもやっぱり文章から症例が出るというだけではだめで、おそらくリーズニングの過程が見えないと医者は信用しないと思う。どっちかという、判別機のほうがいいのではないかといいぐらいで、熱がこうだからこう。熱が低いから、じゃ、こうだとか、そういったシステムができる結構いいかなと。日本にとってはいいかなと思う。

荒牧構成員：こういったものをご紹介したが、私は、でも情報が一番自由に羽ばたけるといいうか、乗りおりできるのは、医療の枠に縛られないことかもしれないと思っている。健康な人を扱うヘルスケアのほうが、分野として伸びしろがある。栄藤構成員が言われたことは、まさに同意しまして、お医者さんはすごいリーズニングというか、そういったものを、根拠を求められている。説得できないというか、信用しないと思っている。

池田構成員：今の話を伺って思うのは、やはり社会実装というのを考えたときは、何を最初のターゲットアプリケーションとするかというのをよく考えないといけないと思う。例えばさっきの荒牧先生の医療の例でも、症例検索から診断支援をして、あるいは診断医に置きかえをするというふうに、研究としては一定のアプローチはあるけれども、最初はどこを目指すのかというものを一旦明確に決めて、そこに対して集中的に投資をするというビジョンが大事だと思う。私は医療のことをよくわかっていないので、それが症例検索なのか診断支援なのかとなると、私はよくわからないけれども、そこをやっぱり医療関係者とか、あるいはもう少しビジネスの関係者とか、そういう議論をして、最初のターゲットというのはいっしょに明確にすべきだというふうに思う。

宇佐見構成員：池田構成員のお話と通じるけれども、前回も申し上げたとおり、社会実装ではサービスを限定し、その効果を明確化することが重要と考えている。その中で、荒牧先生の最後の17ページはすごく興味あるデータと感じている。健常者の語彙能力に関する経年データと認知症患者の状態変化について、例えばスマートフォンの会話記録等からこういったことが検出可能となれば、スマホによる脳疾患の診断というサービス分野が開拓でき、大きなインパクトになるのではないかと。これは質問になるが、これらのデータは数多くの試験で取れたものと思われるが、研究として見たときには、まだ基礎段階の段階か。

荒牧構成員：こういった研究は、ここ数年注目を浴びていて、ほんとうに来年ぐらいに商品として発売されるようなレベルまでは行っていると思うが、しかしまだそういったものはないので、ちょっと火がつく直前の技術かなと思っている。

宇佐見構成員：質問の意図として、測定方法は医学的にしっかりした理論に基づいたものであり、それに対して自然言語処理でどこまでできるかが知りたい。横軸はおそらく、患者の会話、縦軸はそこから何かしらの症状が出ているかを特定するものだと理解している。現在、研究分野として縦軸、症状特定の高精度化の方へ向かっているかについて知りたい。

荒牧構成員：そういった方向感の研究が、この数年間。

原構成員：医療分野においては、医事法の問題が障壁になることもある。例えば、診断支援 AI を医者が用いたときの説明責任や、診断支援 AI が間違っただけの責任など、分かる範囲でご教示をお願いしたい。

荒牧構成員：今現在つくろうとしている診断支援システムは、医師の見落としを防ぐといったものなので、補助的な使用として、こういった可能性もあるのではないかと、念のために聞いてくださいといったようなシステムなので、そういった役に立つことはあっても訴訟リスクなど負うようなことはないかと思っている。

原構成員：それは、あくまで支援システムとして導入する分には問題はないということか。

荒牧構成員：はい。

東構成員：いろいろとご発表を聞いていて、荒牧構成員にちょっとご質問したいのだけれども、今回、社会実装ということで、データをいただくときに、ビジネス化したらいただけるという話もあったけれども、おそらくかなり事業化を目的として、こういうことをやりたいんだということをししないと、産業界が後押ししないようになってくると。最近少し気になっているのが、医療AIが結構バズワード的に入ってくるのだが、やはり創薬分野で、例えば創薬の発生に使うのか、自然言語の要素があるのかとか、例えばさっきの認知症のところであれば、大学の研究機関であれば、脳の中でのアミロイドベータの蓄積を生み出したりとか、大分初期のころから暗号しろっていう研究されている方々と、言語に出てきて、自然言語で対象を考えるとやっている方々とか、結構それぞれやっている研究対象が紐づいていない状況にあるかと思っていて、そのあたりのつなぎ込みって、いろいろな研究機関とお話しされていると思うが、どういうところまで進んできているのか、連携ができていないのか、今の状況をお伺いしたい。また、いろいろとデータサイエンティストのお話が、医療業界の方々にも必要かという話もあったが、最近私のほうでは結構医療系のスタートアップベンチャーなど支援しているが、30代のドクターの起業がここ最近2年で大分増えている。このあたりが結構、政府としても応援するところかなと思っている、30代ドクターぐらいの中では横の連携が強くて、コンピューターサイエンティストとか仲間が周りにいるということで、若い方々が集まってチームアップをして起業するというケースが多い。大抵、社長とかやられる方々に医師、ドクターの方とかクリニック開業医がいらっしゃるから、みずから臨床できると。先ほど少しそんな話もあったけれども、最近出てきているのが、どちらかというと認知行動療法。薬を与えるのではなくて、アプリを処方するということに着目して、それを医療機器認定していこうとか、保険点数をとっていこうという取り組みが出てきているということで、かなり医療分野と情報機器領域の融合が進んできたのがここ一、二年かなと思っている。このあたり、より加速するために、何らかこのWGでサポートする体制をつくれたらと思っているけれども、このあたり、現場感からご意見とか課題点とかあるのか。

荒牧構成員：最初の、連携ができていないのかという質問からいくと、現状のほかの研究と連携ができるかということ、できつつあるという段階で、認知症に関して、うちは東大病院ともやらせていただいているが、実は言語で認知症というのは、今日本国内で3つぐらいほかにもやっているところがあり、それぞれ各施設と連携してやっている。そういった形で、後者の話とも関連するが、医療系のほうにもAIに対する理解が広まっているので、組みやすい感じだなと思っている。相乗りのイメージのところと我々のところも一緒に乗ろうとしている。後者の医療者のほうのAIの理解というのは、ほんとうにそうだと思っている、私の前の前の職の東大病院ですと、東大病院の情報部というのは、ほとんどMDがいなかったけれど、今は非常にAIブームで、AIをしたいというお医者さんがいっぱい出てきて、企業だけでなく大学のデータサイエンティストにお医者さんのほうからなろうとしている動きもあるの

かなと思っている。そういったものに追い風を受けて、医療機器としてのAI、それが保険点数がおりとかがそういうところまで行くかという方向にも、医師の方が先導して入られると多分大分スムーズに行きやすいかと思うので、近い将来そうなる方向性に行くのかと思っている。

野崎課長：荒牧構成員に、せっかくの機会なので、1つ簡単なことを教えていただきたい。IBMのWatsonは、今、例えば東大の病院とか、大量の論文を入れて診断支援にいろいろ使われているというのをよく聞く。先生の資料の8ページにも、この症例検索から診断支援のデータへということで、症状をもとに診断名を推定と出ている。IBMは医療分野の用語データも充実させて、論文の中で因果関係等をかなり緻密に分析して、場合によっては創薬シーズまで見つけるぐらいの深い解析をしているのではないかと思う。日本でも8ページのような研究が進んでいるけれど、どういう差があるのか。それとも、もう解析レベルは追いついているのか、教えていただきたい。

荒牧構成員：症例報告といったものは論文の一種ではあるが、感覚としてかなりカルテに近い状態で、一症例について書かれた報告なので、学会で報告されますが、基本的には1人の患者さんについて書かれたカルテっぽいもので、そういったNは1のものを扱うような今の日本のシステムとIBMは論文、こういったものはこうだという知識を扱ったものと、そこに大きな差が出るということだと。解析のレベルということだが、英語と日本語という違いがあるので一概に比べるのは困難だが、IBMのほうが進んでいるだろうかかなと思っている。というのは、鳥澤構成員が言われました質問応答のデータを大量につくっているし、日本のほうは生データはいっぱいあるが、それに対してアノテーションといったものは今行われていない状態なので、こういったレベルの解析の深さは、IBMのほうがすぐれていると思っている。

野崎課長：こういう症例の解析だとできるけれど、例えば創薬シーズを見つける等はやっぱり論文の解析まで行かないとできないとか、そういうことはあるのか。

荒牧構成員：利用シーンの違いかと思うが、IBMが目指しているのは、患者さんが来て、治療法がなくてどうしようもない、どうしたらいいのかといったときに、あ、こんな治療法があったのかといったものを見つけるシステムかと思っている。だから、医科研とかそういった高度な医療の施設で使われているのかと思っているが、しかし実際の現場のお医者さんがやっていることは、ほとんどがそういった感じではなくて、今回来て、多分こうかなと思ってとりあえず治療法を出して、どんどんやりとりしているうちに、3カ月かけて診断を確定させていくと。その際に、そういったクイズのように、これだというようなものではなくて、一連のプロセスなわけである。そうして、その最中に見落としがあったら困るといったことで、常時診断支援してサポートしていくと。そういったようなイメージですので、全然目的も使う材料も異なってくるのかなと思っている。

野崎課長：要するにいわゆる高度医療の病院が対応するような難病で、治療法がわからないみたいな病気は、Watsonのような論文解析システムが要るけれど、いわゆる一般の町のお医者さんが、見落としがないようにするためには、この症例を中心とした解析で十分役に立つということか。

荒牧構成員：はい。

鳥澤構成員：いろいろなアイデアがあって、研究者としてうらやましい限りだが、インパクトというかマーケットというか、一押しはどれなのか。いろいろな話があったかと思うけれど、例え

ば、よくわかっていないので、これは失礼な言い方になるかもしれないが、確かに認知症が進んでいるかどうかという話がスマホでできるとはおもしろいけれど、でも、それって何か申しわけないけれど、何かアンドロイド何とかになるのか、アプリ一個でき上がったらもうそれでおしまいみたいな話に、下手するとなってしまう。一方で症例報告をうまく使って、見落としを防ぐというシステムは、ひょっとしたらお医者さんが1人1台持ってないと話が始まらないという時代が、ひょっとしたら来るのかもしれないし、あるいは今アドレスしていないようだが、じゃあ、やっぱり Watson がこの後マーケットを支配してしまうんだみたいなふうに思っているのか。ちょっとその辺をばかっと、一押しはこれですと言ってくれれば、皆さん納得するんじゃないか。

荒牧構成員：一押しはSNSを使ったものが一番いいと思っているが、しかしこのワーキンググループでは、脳があると思っているので、脳と言語処理が融合している大変少ない例として、最後認知症などというのも非常に検討する価値はあるかなと思っている。

鳥澤構成員：SNSが一押しというのは、要するに、もうお医者さんは、すべからく10年後にはSNSの情報を使っているということか。

荒牧構成員：いや、お医者さんが使うというよりは、普通の人。健康の範囲で、医療とは違う範囲でも使うといったものが、総務省が目指すような人工知能としては。それを置いておくと、お金が一番もうかるのは認知症だと思う。これは、認知症がもし治れるようになるとすると、自己メンタリングが欠ける。今、治らないけれど、もし治るとなった途端には、すごいマーケットであると。

柳田主任：多分、僕も一応医学部の教授をしていた立場からすると、認知症とか鬱病とかいうけれど、ものすごい多要因で、個人個人でもものすごく違うし、細胞レベルまで落とすと、ものすごいバリエーションを持っているから、結局はすごいビッグデータをとって、この特徴はこうだということまで特定していかないと、お医者さんも動きようがないということになって、そこまで逆にビッグデータをきちっとそろえると、大きなマーケットにもつながっていくんじゃないかなというような気がする。情報通信だとセンサーだとか、もちろん計測器を開発するという立場だけれど、お医者さん側からすると、ある表面的な状況がわかったと言われてもねという感じになるので、多分おっしゃるように、もう少し次はSNSでもすごいビッグデータをとると、こういう特徴の人はこういう予想が立つとかいうところまで持っていくところに行くんじゃないかなという気がしている。というわけで、そういうことをするためには、総務省はかなりお金をかけて、システムをつくっていただかないといけないというのはあれですけど。

(4) その他

第3回次世代人工知能社会実装WGは3月8日(水)に開催予定。