

第3回AI専門部会

日時 平成29年4月20日(木)

10:00~

場所 PMDA会議室1~4

<開会>

○光石部会長 おはようございます、定刻になりましたので只今より「第3回 AI 専門部会」を開催いたします。最初に事務局より出席状況と資料の確認をお願いいたします。

<委員出席状況確認及び資料確認>

○事務局(江原) 事務局です、委員の出席状況についてまず御報告申し上げます。14名の委員のうち、10名の先生に御出席いただいていることを御報告いたします。

配布資料の確認をお願いできればと思います。まず、お手元に席次表、議事次第等、資料目録がございますので、資料目録を片手に御確認をお願いできればと思います。

資料につきましては資料1から資料5までございます。資料1は荒牧先生の御講演資料、資料2が岸本先生の御講演資料、資料3は報告書作成の整理表(案)、資料4は技術分類と応用分類(案)、それから資料5といたしまして第2回の論点整理(案)となっております。本日の資料、すべてお持ち帰りいただいて結構ですが、取扱注意となっておりますので保管の御留意をお願いできればと思います。不足資料などがありましたら事務局までお願いできればと思います、以上です。

○光石部会長 ありがとうございます、よろしいでしょうか。

第2回の専門部会は宮野先生と清水委員に御講演をいただきました。先ほども説明がありましたが、第2回の議論を受けて資料5にまとめています。これは場合によっては、また後で説明ということでもあります。

本日は2つの講演を用意しており、その後に報告書作成に向けた全体討論を行っていきたいと思います。

最初の講演ですが、「WEB ビッグデータからの医療情報抽出」というタイトルで奈良先端科学技術大学院大学の荒牧英治先生に御講演をいただきたいと思います。荒牧先生は自然言語処理を中心とした情報処理を用いてインフルエンザサーベイランス、花粉症サーベイランスの研究に取り組んでいらっしゃるということです。20分程度の御講演よろしくお願いたします。

<議題1: AI(人工知能)等に関するご講演と意見交換>

<①Web ビッグデータからの医療情報抽出(奈良先端科学技術大学院大学 荒牧英治特任准教授)>

○荒牧氏 御紹介ありがとうございました。奈良先端科学技術大学院大学の荒牧が

御説明いたします。

先ほど自己紹介させていただきましたが、自身のバックグラウンドは情報系、その後、東大病院の大江先生のところで医療情報に関することを扱って以降、医療と情報の両分野に渡って研究を続けております。今日はその中でビッグデータに関する内容を2つ御紹介したいと思っております。

まず、近年、医療と情報との間で一番注目を浴びていると思っているのはIBMのWatsonかと思えます。このIBM・Watson、医療に関して次のように言っています。「私は4,000以上もの白血病の論文を1秒もかからずに読めます」。これはある意味正確に言っている面もありまして、4,000以上もの論文というのはかなり速く論文を読む能力があるのですが、読んでいる対象といったものは論文であるということです。

しかし、実際の医療現場におきましては論文以外の様々な業務にも用いる文章があると思っています。一番多いのは現場で使う診療録ですし、それ以外にサマリーや症例報告などがあります。現在扱っているのは一番綺麗に書かれている論文だけですが、今後10年ぐらいの間に徐々に、普段使っている文章のほうも対象になって真のビッグデータが使えるような感じになるかと思っています。

こういった文章を扱う技術には、こういった技術が必要なのかを御紹介します。ほとんどの文章には自然文が含まれています。構造化されている項目もありますが自然文が入っており、その部分の書き方、表記が様々な存在します。

これは気分が悪いことを示すような概念をどのように表現するかといったバリエーションです。様々なバリエーションがあって、これらから全部、詰まるところICDコード、病院のコードで言うところの気分が悪いことを示す。こういうように紐付けるような能力が必要となってきます。これが自然原語処理でできる非常に大きな武器の一つかと思っています。

実際にはほかの技術も色々使っております。文が与えられた際に、この文はどうやら何か副作用が起きているような文なのですが、ここからどの部分が薬であって、どの部分が副作用であって、その関係を捉えてあって、しかもその副作用は正式な言い方、MedDRA/Jで言いますとどういった表現になるのか、薬も商品名ではなく成分ではどういったものなのか。こういった紐付けも全部成功して、初めてこの1文を解析できるといった技術です。

現在の技術なのですが、文が与えられて、薬品名や症状を抽出する精度は大体80%から90%ぐらい、そこそこできることになっています。それ

を正式な表現に標準化するという、自動標準化の技術も 80% ぐらいやっています。この 2 つの関係を捉えるといったことはちょっと難しく、50% ぐらいやっています。一個一個は結構研究が日進月歩で進んでいます、全部成功して初めてこの部分を解析できるのですが、それはかなり難しく、全部成功するのは 30% ぐらいとなります。一個一個の文章の精度は低いのですが、しかし、大量に集めることによって、量で質を担保しようといった考え方がビッグデータの基本的なアイデアかと思っています。

こういった病院内文章を解析するようなツールなのですが、これまでは企業との共同研究などで各企業であったり、研究者間がインハウスで造ってきたりしたのですが、良いものをちゃんと交換しようといった動きもあり、これは厚労科研でやっているものなのですが、病名や症状を抽出するような構文解析器、こういったパーサーをオープンソフトウェアとして公開しようといったことも今行っています。今年度、これは班内で試験的に公開しており、来年度から実際に Web 上でダウンロードして利用できることになるかと思っています。

本日お話しするビッグデータなのですが、病院内の文章を扱うような技術を発展させていくような形、延長線上にあるかと思っています。これは私の私見なのですが、一施設であったり複合施設が扱うような病院システムを主に医療情報分野では扱ったりしていました。しかし、それを超えて、今では SNS、具体的にはツイッターやフェイスブックなどでそういった文章を扱ったり、センサー情報を扱ったり、様々な形で発展するかと思っています。このような病院内の文章が使うので培った言語処理技術を Web の文章に適合したのが後で御紹介する 2 つ目の研究事例になります。

その 2 つ目の研究事例なのですが、最初に材料の規模のすごさを御紹介したいと思っています。これは主な SNS 上での活動の統計量なのですが、フェイスブックに「いいね！」をする数というのは 1 分当たり 400 万もあります。つぶやくといった声は若干重たいのですが、それでも 34 万あります。これだけの数がありますと、この中の 0.1% でも何か、医療に関することをつぶやいていると毎分何万と溜まっていくことになります。これを何とか利活用できないかといったものが基本にあります。

そういった 1 アプリケーションの例をこちらのムービーで御紹介させていただきます。これは風邪の例なのですが、風邪を引いている人の症状をつぶやきから取ってくるというシステムになります。風邪の主な症状は 6 個ぐらい製薬会社の方と相談、喉の痛みや寒気といったものをピックアップして表示しています。

こちらはそのインフルエンザバージョン、インフルくんというシステム

になっています。このインフルくんは既に公開していますので、皆さん興味を持たれましたら、この後御自身の携帯などで検索していただければこの場面を見ていただければと思います。

このように現在のつぶやきの数が出まして、それがどのぐらい合ってくるのかということなのですが、これがこちらの精度になっています。過去数年間の精度なのですが、この青色がつぶやきの量になっています。この赤色の部分について感染症情報センターが出している患者数になっています。こうやって見ますと、過去5年間の間、今年度は若干外したのですが非常に傾向をとらえていると思います。しかも、これは全国統計なのですが、実際のつぶやきには10%ほど位置情報が判明しているものがあります。その10%を可視化してやるとこういう感じになり、北のほうで流行っているとかいったことも表示できます。今、こういった技術が徐々にできつつあります。

これを実際に使うとなると非常に大きい問題もあります。まず、一番大きい問題は誰でも素朴に思い付くのですが、田舎のお祖父ちゃんやお祖母ちゃんがツイッターでつぶやくのはとても考えられないといったことです。これは本当にそうです。これは東京と北海道の札幌なのですが、札幌は相当都会だと思えますが、それでも東京と比べたら本当にまばらな感じになっています。例えば、ネット上ではインフルエンザが流行っているのはほぼ都会側だけになってしまっています。これはどうか。例年例年動かして行って、北海道でのつぶやきがこのぐらいの量は患者数で言うところのぐらいという感じで、1年分試行して、そこから補正してくるといったことを行います。

今は量的な補正でしたが質的な補正もあります。つぶやきといったものがインフルエンザに関する調査用に発しているわけではない、様々な表現があります。それが本当にインフルエンザと関係するのかどうかといったことを判定するといった技術が必要で、全部、単に単語をピックアップするだけでは相当ノイズを拾ってしまうことになります。

そのノイズの例がこちらです。「死ぬほど腹が痛い」と言っているのは多分、おなかが痛いのだろうと分かるのですが、ここから下が難しい例、下から2つ目など特に難しいのですが、「来年大台に乗るかと思うと寒気がする」と言いますが、これは精神的な寒気であって肉体的な、風邪の寒気ではないことを判断しなければなりません。これは極めて難しいかと思いますが、実は今まででは相当の精度で出てきます。大体、80%ほどの精度で出てきます。

これをどうやるか。皆さんが使っている技術でも非常に近いものがあり

まして、迷惑メールをフィルタリングするような機能があるかと思えます。あれは相当精度がいいかと思えます。色々なメールが来るにしても、要は迷惑な宣伝メール若しくは自分個人宛てのものかというのは二値判断であります。文章が与えられて、2 つに分離することは非常に高い精度でできますので、つぶやきのような 140 文字ぐらいであったとしても、要は症状がそこにあるのかないのかといった位置判断ですので結構な感じでできることになります。具体的にはキーワードをピックアップして、このキーワードの組み合わせだったら風邪、そうでなかったら違うという感じで分類することになります。

そのような技術を使い、これは過去 3 年間分の全国データなのですが、各都道府県、大抵の都市で非常に高い精度で一致するといった結果が出ています。このピンク色が感染症情報センターの値、赤い色の線がつぶやきの値になっています。

こういう技術はすぐにでも使えそうなのですが、どうやって使っていくかということについても検討を行っています。インフルエンザに関しては感染症情報センターが正確な値を出しています。どのように使うかといいますと、ここが使い方なのですが、最近海外からの渡航者が増えて、特に東京オリンピックに向けて増えている。そういった際に未知の、これまで日本では想定していなかったような、サーベイランスしていなかったような感染症を持って日本に来る方がいて、想定外の感染症が広がる可能性があるかもしれません。そういった想定外の情報に対して、ネットの情報をもうちょっと使ってサーベイランスしようといった形で、この AMED の研究ではつぶやきの研究も使って今実用化を目指しています。

具体的なシステムの画面もありますので御紹介します。つぶやきくんは「つぶやきくん IG」といったシステムウェアになっています。この IG というのは、Infectious Gastroenteritis、感染性胃腸炎に感染サーベイランスになっています。感染性胃腸炎のような、インフルエンザほど大規模ではなくて、そこそこのボリュームで起きるのですが、なかなか予測が困難なものに対してつぶやきを適用しようとしている例です。

この場合、難しいのは、「感染性胃腸炎」とつぶやくのは滅多に考えられません。病名でつぶやく頻度は非常に少ないです。ですから、この上のほうを見ていただくと、手元の資料のほうがよく見えるかもしれませんが、具体的には下痢であったり寒気であったり症状ベースで見つけてやって、感染症の察知をするといった方法でやっています。

こういったシステムを動かしていると、実際に食中毒が発生した際、このシステム上で見ると若干増えたりします。これがどの程度の規模の

感染症などの発生をどのぐらいの精度で検知できるのかといった、パフォーマンスのチェックを今年度やっています。極めて予備的な段階ですが、どうやら300人規模で起こった場合にはネット上でも結構ピークが出そうだと、という結果が今のところ得られている状態です。

更につぶやきだけではなくて、ネット上の情報には様々なものが存在します。これは「Yahoo!クラウドソーシング」といった、Yahoo上に登録してあるようなワーカーに対して同意的にアンケートを出すようなネット上の仕組みを作ったものです。これにより、医薬品の副作用に関するような体験はないですかと聞くような研究も実施しております。これは慶應大の薬学部の先生と一緒にやっている研究なのですが、こういったネット上での公募でかつてないような、大規模な副作用を収集するという研究です。

ただ、これも非常に難しい例がありまして、どのような薬を飲んだかといったこともちゃんと分かっていなくて、何か胃の薬であったりなど極めて抽象的な感じで書かれてあったりします。症状のほうも「キリキリ痛む」「顔がパンパンに腫れた」など、普通の平易な表現で書かれています。こういった、大規模に集まるのは集まるのですが、若干標準化されていないような表現を言語処理で標準化してやって収集する試みが今行われつつある状態です。以上、駆け足で研究を紹介してきました。

最後、よく聞かれる質問についてここに挙げました。こういった技術が使えるのですが、こういった点はどうなのかといった点です。

まずはインフラの不安定性の問題、つぶやきでしたらツイッター社のツイッターを使うわけなのですが、もしツイッター社がそのつぶやきを集めるのをやめるとか研究利用を停止するとどうするのか。これは本当に大問題だと思っています。重要なインフラとして使う場合にはその辺を担保したくなってくるのですが、この場合、我々の答えとしては複数のSNSを使うことを考えています。Yahooの検索クエリも使えばつぶやきも使えば、フェイスブックも使う。研究上は一つで検証するのですが、実際には複数使うことで解決するのではないかと思っています。

あとはユーザ・バイアスの問題、先ほどお伝えしたように都会の若者がよく使っていますので、田舎の、お年寄りの情報をどうやって見るのか。これも非常に難しいのですが、しかし一つ救いなのは介護者の方は若いので、お祖父ちゃんやお祖母ちゃんの病気のことをつぶやく可能性もありますし、お子さんの病気のことをつぶやく可能性もあります。ですから、使い方を工夫すれば何とかなるのかもしれないと思っています。

最後に正確性の問題なのですが、もし仮にバイオテロみたいなものが起

こった場合には、こういった SNS というのは誰でも簡単に書き込みますので、大量のノイズを入れることもできます。まだ考え始められたばかりなのですが、今後検討していく必要があるかと思っています。

最後、詰まるところは、こういった不安材料もありながら、情報技術を使うかどうかといった場合、そして使った場合、誰がどのように責任を取るのかという問題です。これについては、一研究者の範疇を超えていると思いますので、皆さんの考え方に委ねたいと思っています。

以上、簡単に御紹介しました。こういった技術が非常に注目されている面もあり、この 3 か月間でも 3、4 回ほどテレビ取材や新聞取材がありました。こういったものがどんどん周知されるに連れ、早く世に出したい、使えるものにしたいと思っていますので、皆さんのお知恵など拝借できればと思っています。以上、私からの報告を終わりたいと思います。

○光石部会長 荒牧先生、どうもありがとうございました。それでは、質問やコメントをいただければと思います。いかがでしょうか。

○原田委員 原田と申します、貴重な御講演をありがとうございました。
ビジネスモデルについてお伺いしたいのですが、例えば行政や製薬会社など、そういうところへの情報提供というビジネスがあると思います。例えば医療機器になって、医師の方に診断のヒントを差し上げるとか、そのような医療機器としてのビジネス形態がもしあり得るようでしたら教えていただきたいのですが。

○荒牧氏 実際の臨床の現場で使う場合としては、最近、すごくインフルエンザが流行っているから気を付けてくださいといったような、若しくは今流行っているインフルエンザの A 型なのか B 型なのか、そういった情報とかは事前にお医者さんに出すシステムとして使えるかもしれません。その際、そういったビジネスと言うか、機器としてどうするかというと、形態についてはちょっと私自身も考えていないのですが、今のところフリーの Web サービスとして一般の方が誰でも使う。ビジネスとしてはそこに広告を載せて、広告商材として使うみたいな形は今まで経験があるのですが、これ自身を販売するといったことは今のところしていない、検討もしていない段階です。

○原田委員 研究段階だと思いますが、ツイッターや Yahoo などのデータは実際に例えば製品、サービスになった場合、有料でそういうところから御提供いただくという形になるのでしょうか。

○荒牧氏 商品の形態にもよるのですが、つぶやきに投稿されたデータを Web サービスとして誰でも見られる形として公開することに関しては別に法律には抵触しないので、それは一応可能なのですが、その情報を販売するよ

うになりますと多分問題になるかと思っております。

- 原田委員 今、ツイッターとかのデータは無料で提供いただいていたから、特に費用は発生せず収集されているということですか。
- 荒牧氏 はい。
- 原田委員 ありがとうございます。
- 光石部会長 ありがとうございます、ほかにいかがでしょうか。
- 許委員 大変面白いお話をありがとうございます。先ほどデータに地域性がある、あるいは年齢で例えば子供さん、成人、それからお年寄りというようなバイアスがかかる可能性がある。逆に、1人のツイッターを縦に評価することで、そのツイッターを出しているキャラクタースティックみたいなものをある程度推測することが可能だと思います。
- そういう形で、例えば選挙の時などにどの年齢層がどの人を支持しているとか、そういう形でビッグデータを一つ一つ、発信している人のグループ分けなどをすると大分精度が上がるのではないかと思います、そういう点はいかがですか。
- 荒牧氏 ありがとうございます、非常に興味深い御提案です。そういったことを既に何度か試みている研究者もいるのですが、どうやらそれほど長期間つぶやく人があまりいない。そう言う問題がある。つまり、位置時点でのボリュームが大きく、更にその時間軸を追えばもっと正確になるというのは本当にそうだと思いますし、非常に貴重なデータかと思えます。
- どうやら、ツイッターのようなものを1、2年ぐらいやってやめていく人も多い。あと、一般的にWebサービスとして、いまだかつて10年以上続いたWebサービスがあまりない状態ですのでツイッター自身も10年続くかどうか分からない。ちょうど、今、10年目ぐらいになるのですが今後多分衰退していくと思われまますし、数年間ヘビーユーザーだった人が離れていく。数週間ぐらいの短い時間軸でしたら多分追跡は可能なのですが、もう少し大きい時間というのは難しくなってくるかと思います。
- 光石部会長 ほかにいかがですか。
- 鎮西委員 例えばガセネタを噛ませられた時、ガセネタに特有のパターンなどを検出するという話がありますか。
- 荒牧氏 ありがとうございます。ガセネタというのはネット上に非常にたくさんありますので一定数あります。そのガセネタが一定量増えてきますと、何とかと言ったものはガセだと言う人も存在します。たまにピックアップして、訂正する人の言い方といったものはパターンが決まっていますので、そういったものを収集してガセネタを捉えるといった研究も結構行っています。「何とかというのは、でもな」、「何とかというのはガ

せだ」と言って引用する。そういったフォームごとに疑わしいツイートを選別することも可能になりつつあると思っています。

○鎮西委員　　もう一つ、ツイッターやこういうものは例えばどこから発信されているという、場所の情報は必ず付いているのでしょうか。

○荒牧氏　　場所の情報が付いているのは1%ぐらいしかない状態です。先ほど、10%ほど場所が分かると言いましたが、残りの9%ほどは位置情報が分かりませんが住んでいる住所を一応書いている、若しくはその人自身は分からなくても、その人の友達が住所を書いているという形で、周りの関係から攻めていって強く推察されるという形で、県レベルでは10%ほど分かるということになります。

○鎮西委員　　分かってしまうのですか、怖いですね。あと、我々はかなりの距離を移動することもあるわけです。例えば、移動とこういう話を組み合わせるとい解析はされていますか。

○荒牧氏　　そうです、住所というのは2プランあります。住んでいる地域と今現在どこでつぶやいたかという2つの位置情報があります。こういったサーベイランスで使っているのは住んでいる位置でやっていますので結構分かります。どこに移動したかといったものはつぶやいている内容をちゃんと注意深く見ると、スカイツリーと言っていると東京都という感じで判定するような技術もあるのですが、それは今現在研究中であり精度もよくない状態です。ですから、旅行先での発言はノイズになってしまつて、今、住んでいる地域を頑張って当てているといった状態になります。

○橋爪委員　　非常に面白い。しかも、Webからデータを吸い取ってこれだけのことができるというのは非常に興味を持ちました。最初に先生がおっしゃられた診療録からデータを吸い上げてというところ、もしこれが本当にできるのであれば、個人情報さえクリアできれば毎日全国から外来であれ、入院であれ、診療録から自動的にそれを吸い上げて何が全国で起きているかというのは可能なのでしょうか。

○荒牧氏　　近い将来、それが可能になれば良いと思っています。実際、これは多分、大江先生のほうがお詳しいかと思いますが、診療録自身は結構書き方が相当バラエティに富んでいて、やはり相当難しいかと思っています。しかし単位サマリーとか、ちゃんと報告用には書いているようなサマリーでしたら多分、現状の技術でも結構解析できると思っています。その部分が近い将来、実用化するのかなと思っています。

○橋爪委員　　ある意味では、ドクターが診断した情報を記載していますので、通常のツイッターよりは発信元というか、主催している人達はある程度は信頼が置ける。ドクターが診断した情報を記載していますので、そのあたり

は結構統一されているのではないかと思います。ですから、もし診療録から自在に情報を得るためにハードルがあるとしたら、どのようなことがあるためにそれができないといけないのかなと思いました。

○荒牧氏

私の理解している限りハードルは多分2つあります。現場で書く文章はその場のために書きますので、若干汚く書いているので、それを自動処理するのが困難であるといった解析精度の問題がある。あとは自分用に書いているものですので、それを医事利用することに対し、現場のお医者さんも抵抗があるかもしれないといった、あまり診療録を情報として出したいと思われない先生も多分多いかと思います。その2つが問題かと私は理解しています。

○橋爪委員

今後どうなるか分からないのですが、カルテも患者さんが自由に見られるということでオープンにしているところもありますよね。そういうところから段階を経て情報を吸い取る、ということが研究的にやられると課題などがしっかり見つかってくる。

それができると、どの地域からどういう病気、あるいは症状が起きているというのがリアルタイムに出てくるのではないかと先生のお言葉を聞いて思いました。

○荒牧氏

ありがとうございました。

○光石部会長

カルテみたいに、ある意味で信頼性の高いデータとここで扱っているような、問題のところに書いてありますが、しょせんと言って良いのかどうか分からないのですが不安定なデータと大分扱いが違うのかなという気がしています。多分、マクロな動向というのはこういうことでかなり捉えられるのかなという気がするのですが、例えば個別、最近でも投薬や医療行為などで結構問題になっている例もあるように思います。そういうものをあまり確実でないというか、信頼性の置けないデータから抽出してどうのこうのというのは結構危険なのかなという気がします。どの辺にしきい値、しきい値と言って良いのかどうか分からないのですが、そういう点はもう結構確立されているのですか。

○荒牧氏

ビッグデータの一番大きい考え方として精度を量で担保するという点があります。これだけの精度であったとしてもそれなら10倍用意しよう、若しくは1000倍用意しようという、量でカバーするといったものが大きい。そういった点では非常に相性が良いと思います。現場で書いているお医者さんの文章自身を利用するというのは、ビッグデータの観点から見るとちょっと量的には落ちるので正確に処理したいというものがあります。そういった点ではビッグデータというより、違ったデータの扱い方が要るのかもしれない。

一点、私自身も今検討していますのは、お医者さんが自然に書きながら同時に収集できればと思っています。システムの遅れ、仮の案でちょっと見づらいかもしれません。Wordの文書のように、書いていると標準化されていない表現の下側に線が入ったりする感じ、文法チェックのようなものが入ります。ここを書き換えると、この部分はこういうように言い換えても良いですかとなって、それを標準化した形で提示して可視化する。こういった現場の書きやすいツールみたいなものを提供してやれば、自然と標準化した表現が入って行って、データを吸い上げるといったこともできるかと思っています。こういった形で医療現場自身に介入すれば非常に良い形でデータが集まるかもしれないと期待しています。こういったものも一方では試みられていることを御紹介いたします。

○大江副部長 今の橋爪先生の御質問に関連してですが、以前、荒牧先生が私のところにおられた時に先生がおっしゃったように、電子カルテの記載から色々知見を抽出することを試みたことがあります。現在も複数の大学病院の電子カルテの自然言語処理を継続してやっています。

何が難しいかと言うと、退院時サマリーや看護記録は比較的ちゃんとした文章なのですが、それ以外、日常医師の書く診療記録はツイッター以上に文ではない、ほとんど文ではないのです。事実と所見の羅列で、かなり医学知識のある人が関係を汲み取らないと、なぜここにこれが書いてあって、書いてあることは何を意味しているのかが結構分からない記述が大量にある。多分、そういう領域にある程度特化した言語処理を開発していただくと橋爪先生がおっしゃったようなことができるようになるだろう、おそらくそれはできるだろうと思います。そのような状況かと思しますので補足させていただきました。

○光石部長 ありがとうございます。まだ質問があるかもしれないのですが、後の総合討論でもしあれば質問していただくこととし、次の講演に移らせていただければと思います。荒牧先生、どうもありがとうございました。

次は資料2にあります、「機械学習を活用した精神疾患の重症度評価システムの開発」、慶應義塾大学医学部精神神経科の岸本泰士郎先生に御講演いただきたいと思います。岸本先生はバイオマーカーの設定が困難な疾患領域での患者の表情、音声、日常生活の画像などを定量化する先進的な研究をされているということです。岸本先生、よろしく願いいたします。

<議題1：②機械学習を活用した精神疾患の重症度評価システムの開発（慶應義塾大学医学部精神・神経科学教室 岸本泰士郎専任講師）>

○岸本氏

慶應義塾大学精神神経科の岸本です。本日は、このような場で発表の機会をいただき、誠にありがとうございます。私は、精神科の臨床医で、普段は患者さんの診療に当たっておりますけれども、1年半ぐらい前に、日本医療研究開発機構 (AMED) からチャンスをいただき、機械学習を活用した精神疾患の重症度評価システム、あるいはそれを可能にするような医療機器の開発のプロジェクトに取り組んでまいりました。本日は、この研究開発のバックグラウンドであるとか、あるいは研究開発の経過についてお話をさせていただきます。

代表的な精神科の疾患を挙げてあります。うつ病、認知症、統合失調症といった精神疾患の診断は、患者さんと我々精神科医との会話を通じて行われます。患者さんがどういう気分でいらっしゃるのか、気力はどうか、考え方はどうか、日常生活はどのように送っていらっしゃるのか、患者さんの表情、声、動作といった雰囲気、我々がある程度観察しながら、典型的な患者さんとどう類似しているのか、あるいは我々が正常と考える範囲をどの程度逸脱しているのか、というところをいわば自分自身の物差しと照らし合わせながら、患者さんの診断、重症度評価、治療を行っております。

例えば、うつ病の患者さんであれば、表情は非常に暗くて、声も暗くて、抑揚のない声になります。動作は緩慢になって、話の応答そのものも遅くなります。こういうのを精神運動抑制と呼んでおりますけれども、そういう特徴が出てきます。発言内容は悲観的、そして自責的になります。

認知症は、記憶力低下は代表的な症状です。それ以外にも語想起障害と言いますが、言葉がパッと出てきにくくなることで、指示語が増えたり、抽象語が増えたりする。記憶の低下を繕うような形の発言が増え、迂言という回りくどい発言が増える。その他に意欲や関心の低下というものが認められるようになります。

我々がうつ病や認知症の患者さんの診療に当たる際に、どのように患者さんの重症度を評価しているかという、色々なレーティングスケール、評価尺度があります。代表的なものをこちらに挙げています。うつ病、躁うつ病においてはHAM-Dと呼ばれているスケール、あるいはMADRSというものが有名です。具体的には患者さんの自覚症状、気分、エネルギー、自殺念慮といったものを点数化する。食欲、睡眠といった生理活動に関する質問を通じて点数化する。我々の観察を通じて点数化する。それを合計点で表すような仕組みになっています。

ただ、色々な問題があって、判断基準が非常に曖昧になりやすい。アンカーポイントは当然用意してあるのですが、どうしても評価者

によってバイアスを受けてしまい、なかなか安定しないという問題があります。こういったレーティングスケールをしっかりと行おうとすると、30分程度かかります。忙しいクリニックでは5分、10分で患者さんを回していますので、現実的にこういうレーティングを行う事は不可能というような状況もあります。

認知症に関しても、色々なスケールがありますので、その代表的なものを挙げています。評点の仕組みはものによって少しずつ違うのですが、記憶力を中心に、認知機能を広く浅く取るものもあれば、特定の認知機能にある程度焦点を絞って、非常に深く聞くようなものもあります。問題点は、簡便なものは非常に大雑把で天井効果、すなわち非常に優秀な方がある程度認知機能が低下してきても、なかなか点数に表われないという問題があります。床効果は、ある程度悪くなってしまうと、底点を突いてしまっていて、なかなかその後の変化が分かりにくいという現象。練習効果、これは非常に大きな問題だと思いますけれども、何度も同じ検査をやっているうちに、患者さんがそれを覚えてしまってテストにならなくなります。患者さんによっては、私が伺う前から「桜と猫と電車でしたっけ」などとおっしゃったりする。一方で詳しい検査は長時間かかるものですから、我々にとっても患者さんにとっても非常に負担が多くて、しょっちゅうやることはなかなかできないです。

こういった、どうしても曖昧になってしまう精神科の重症度の評価の良い方法がないということが、色々な問題につながっていると考えています。治療開始のタイミングが不明確になったり、治療反応が分かりにくくなったり、さらにこれは新薬の開発にブレーキをかけてしまっているような状況にもつながっていると思います。レーティングがうまくいかないことだけとはもちろん申しませんが、色々な背景はありますけれども、重症度の評価がうまく行えない事も1つの要因です。これは2013年の一般向けの『THE NEW YORKER』という雑誌の記事ですけれども、THE PSYCHIATRIC DRUG CRISIS、精神科の薬が危機的な状況にあるという記事です。pharmaceutical giant と呼ばれるような製薬企業が、精神科領域から撤退するという状態にまで発展しています。

そのような中で、世界中の科学者が色々な方法でバイオマーカーを探し、定量化できないかということで色々な研究をなさっていると思います。私はICTとか、機械学習に注目して、こういう新しいデバイス、あるいは解析方法等を用いて、何らかの方法で精神疾患の病状を定量化できないかと考えました。海外からも、似たような研究、これはプレリミナリーなレベルが主ですけれども報告されるようになっていきます。

これは、携帯電話の会話のパターンから、躁うつ病の患者さんの気分を同定するのに成功したというレポートです。これは、同じく携帯電話で、活動量であるとか、位置情報を用いて、その気分の同定を行うといったアプローチ。これは表情の解析ですけれども、人間によってうつっぽい、そうでないと分類するものと、機械はかなり類似のレベルの結果を出してきているレポートです。これは荒牧先生のご専門分野ですが、自然言語処理というアプローチを用いて、統合失調症の、非常にハイリスクと考えられるような患者さんが、将来発病するかどうかというのを、自然言語処理を通じてある程度同定できたというレポートです。こういうものが出てきています。

我々も、類似のアプローチを用いて、何とかこういう新しい方法で、患者さんの状態を定量化したり、病気を早い段会で気付いたり、再発の予防を早めに気付けないか。そういうことで、国内の多くの会社が集まっていたいただいて、産学連携プロジェクトとして、1年半前からAMEDの委託研究として、我々はPROMPTと呼んでいるプロジェクトを進めております。具体的には診療場面をカメラ、赤外線カメラ、そしてマイクで記録します。患者さんの表情を捉えて、これは主にオムロンの技術を使っていますけれども、画像をもとに、表情の程度などを定量化していく。それから体の動きについて赤外線カメラを通じて定量化する。声は、患者さんの声のトーン、お話になっているスピード(話速)を計量し、お話になっている内容そのものを文字化して、そのものを解析できないかということを行っています。

診療場面だけではなくて、ウェアラブルデバイスを用いて、日常生活をモニタリングすることで、日常の生活の活動レベルであるとか睡眠の状態、こういった様々のデータを集めてきて、我々がゴールドスタンダードと呼ばれているようなレーティングスケールとある程度照らし合わせるような、すり合わせるような形で機械学習のアルゴリズムを作り込んで、診療場面で患者さんの重症度、特にうつ病の場合には精神運動抑制のようなものを定量化するような数値を提供して、それを診療に役立てられないかと考えています。

実用化のイメージですけれども、患者さんが我々、あるいは診察前に心理士さんとの会話でも良いかもしれませんが、患者さんと医療者との会話の様子を機械が観察して、そのデータをクラウドに送って、そこで解析を行って、リアルタイムで診療場面に戻ってくる。こういう形で実用化できないかと考えております。

研究を開始して、1年半経ちました。初年度は半年程度でしたので、患

者さんのデータを集めるためのシステム、基盤構築を行いました。昨年度1年間かけてプロトタイプ、試作段階のアルゴリズムを作り込んできたところです。現在までに、大体450データセットぐらい集めて、患者さんの状態をある程度定量化して、病気であるのか、その重症度はどうなのかということ推定するようなアルゴリズムを作り込んできているところです。

以後のお話を少し整理するために、

[REDACTED]

これは、表情分析の技術なのですが、これそのものを使っているわけではなくてサンプルです。このように機械が、この人が笑っているのか、怒っているのか、悲しい顔をしているのかある程度推定をして、一枚一枚の写真でその推定の結果を返してきます。

これは、機械がどのぐらいの確度で喜んでいると考えているかという数値が上に表示されています。実際には色々な表情が常に混ざっていますので、その推移を非常に短い短時間で推定しております。御覧いただいて、人間が笑っているときには確かに機械も笑っていると判定しているとお感じいただけたかもしれません。

[REDACTED]

[Redacted text block]

これ以降は特許出願準備中のため、お手元の配布資料にはスライドは入
れてありませんが、ほんの一部だけ紹介させていただきます。

[Redacted text block]

[Redacted text block]

[Redacted text block]

客観的な情報を診療場面にお返しするというのは非常に

重要だと思っておりますけれども、こういう技術の開発を通じて、何とか患者さんの予後全体を良くしていきたいということを考えた中で、客観的指標だけではなくて、患者さんそれぞれの予防、再発する前の早期サイン、病気に特徴付けられるような早い段階で出てくるような特徴量を見つけていくことで、全体の患者さんの QOL を高めていくことにつながればという思いで取り組んでおります。

これは AMED からいただいて、委託研究として行っているのですが、最終的には医療機器を作るのが一番の目標です。AMED からは、PMDA ともしっかり話し合いをしながら開発を進めるように、というアドバイスをいただいております。既に何回かは全般相談でお邪魔させていただいて、相談には乗っていただいております。我々として機械学習、こういうアプローチで医療機器を作ろうとしているわけですが、いくつか懸念点を挙げて議論してまいりました。

例えば

学習データを基に解析・推測する。あくまでもこれは推定値であるという部分がどうなのだろうか。曖昧で、それぞれ曖昧なものを、更にまとめてマルチモーダルデータにして、総合的な解析を行うという部分がどうなのでしょう。それから、導入後に診療データが蓄積されて、それを学習することで更に予測精度が高まっていく可能性があるわけですが、そういうものは今までの審査過程とは予想を異にしていると思いますので、どのように考えていったらいいのか。

おおむね左の部分に関してはかなりポジティブなお返事をいただいているように個人的には感じています。もちろんこれからの議論は必要だと思いますので、是非良いディスカッションが今後できればと思っております。以上です。御清聴ありがとうございました。

○光石部会長 岸本先生、どうもありがとうございました。それでは、また 10 分程度質問、コメント等をいただければと思います。

○加藤委員

同じ精神科ですので、岸本先生も意識されたかもしれません。同じような機械学習でというのを、ちょっと違う手法というか、違う方法でATRの研究者がやっています。こういうやり方も非常に面白いと思います。ただ、先ほどの荒牧先生の感染症を対象にしたようなものと、この精神疾患を対象にするというのは、釈迦に説法でとっくに分かっておられると思いますが、相当意味合いが違います。抗うつ薬がもう藪に入ってしまったって、大企業はほとんど手を引いてしまったというのは、おそらくそのためだと思います。要するに、数を増やせば増やすほど藪に入っていくわけです。個々人の背景とか素質という類の人間の脳の色々な問題を全部背景にした上で出てくるような病気を、表面的なものだけでやると、色々な人が、色々なことを感じるというので、できなくなるように思います。

それと、私は発達障害という対象を今やっています。この場合は、そもそも表情が認知できませんので、ある表情をみて、あれは笑っている顔だろうという前提がそういう人たちは成立しません。しかもそういった人たちはかなり多いというようなことがあります。そういう意味合いである程度限定的にやるほうが一般的かと思います。もう1つは評価者が面接場面からといっても、面接自体が問題です。面接のときに、患者さんが一番の問題点を言うとは全く限りませんので、逆に言わないことに一番問題があったりするほうが精神科ではむしろ重要です。そういう意味では、面接者が誰でも良いとは全然言えない。同じ専門家でも、専門性によってその面接で得られる情報量と情報の内容が全く違うということがありますので、その辺を含みに入れてやっていったほうが良いのではないかと思います。

発達障害のほうで言うと、自閉症指数というのがあります。通称 AQ と言います。これはコンピューターでやると、コンピューターが勝手に計算してくれるものだから世界中で普及しています。あれで信頼できて、自閉症を判断できるかというのと、私どもが見ている限りでは全くできません。全然それに合わない。だけど、個々人の、自閉症の人の中で言えばある程度、つまりこっちが囲い込んでしまうと、その中の重症度は比較的私は信頼性があると思います。そういう意味合いでちょっと限定的にするのが重要かと思いました。

○岸本氏

ありがとうございます。私からのコメントですが、1つには、精神運動抑制というお話をしました。やはり、メランコリー親和型の典型的な、我々が考えるいわゆる内因性の典型的な患者さんを基本的にはターゲットにしています。それから診断に使えるか。診断補助、支援という言い

方を本日はしなかったのですけれども、あくまでそう考えています。精神科医が既にうつ病であろうと診断している場合に、精神運動抑制を定量化することで、薬への反応であるとか、それをより見える形にしていきたいというのがコンセプトになっています。先生がおっしゃる、ある程度限定してというところは、我々がもともと考えていた部分ではあります。

[REDACTED]

○光石部会長

他にはいかがでしょうか。

○武田委員

私は医療のことは全く分からないので、純粹に質問させていただきます。今もお話を聞いていると、精神科疾患というのは、本当に病気であるかどうかということの判定も難しい。実際のお医者さんも難しいというお話のときに、こういう機械学習で、例えば正に正例、負例で判定するというのは、その閉じた系ではそこそこの精度が出るというのは、既に実験されて分かっていると。ただ、今もお話があったように、そもそもそれは正例として正しいのかというところが曖昧だと、途端に問題が不安定になると、その前にあるセットだとこれは入るけれども、あのセットは入らないような問題は、その正例を変えるとどんどん答えが変わってきてしまうというので、このアプローチをそのまま使うことはちょっと不安だと思いました。

もう1つは、今のお話を聞いていると、本当に病気であるかどうかよりも、安定した症状の同定が、こういう機械学習でできるのかと。最後の1個のゴールを出すのではなくて、いくつかの症状を機械学習である程度定量化する。その定量化の尺度は、もう少しスケールとして出せるものなので、それを見てお医者さんが総合的に判断する。もう一歩手前側で機械学習について出すというのは、そういうこともこの中で検討されているのでしょうか。

○岸本氏

実は色々議論を今までしてきており、倫理的な部分も含めてですけれども、総合的な指標が果たして、本当に役に立つのかという議論もしています。ユーザーインターフェースを少し手前の段階に落としたほうが、臨床家にとってもむしろ使いやすいのではないかという議論もしています。

[REDACTED]

正例で世界動向、確かにそのラベル付けそのものというのは当然なので、すけれども、それを言うと元も子もなくて、ある程度マスで見えていくと、

共通の部分が見えてきたり、あとは我々がレーティングする上ではかなりトレーニングを積んだ人たちだけにやらせていたりする部分があって、ある程度それを軸にやっていくしかないのかと思っております。

それから地域差等のファクターもノイズとして入ってくる可能性は考えていて、東京だけにかかわらず、あちこちの病院から集めてくることで、共通する特徴例をうまく引き出してくると良いのかとは考えています。

○加藤委員

すみません、1つ補足です。このことそのものを言っているわけではないのですが、NIRS という、光トポグラフィというのがあります。それで変化を見るとうつ病が分かるというようなのが先進医療で今は既に認められています。これは、非常に補助的な診断機器というのか、ひょっとすると補助以前かもしれませんが、そのぐらいのレベルなのです。こういうものが世間に既に出ています。これは現に起こっていることですがけれども、そうすると、クリニックの機械化と言いますか、クリニックの自販機化が一部で起こっています。

どういうことかと言うと、それをパッとやらせると、あなたはうつ病ですよ。この治療は高い治療です。この治療をやりなさい、絶対に効きますという情報がいっぱいもらえる。10回やらないと効かない、10回で数十万円というお金を一遍に取ってしまうのです。それでやっても効かない、それは回数足りないからでしょうというパターンでやって、問題になりつつあるように私は思っています。そっちのほうに行かないようにしたほうが良いです。ある程度きちんと、これは本当にそうなのかということ限定しないと、これは PMDA そのもので問題になってくることあり得るかという気がします。

○光石部会長

先ほどの意見に関連するかもしれないのですが、私はお話を聞いていて2つ思いました。1つは、先生方はお忙しいところで、問診と言っているのかどうか分からないのですがけれども、診断をされる際にこういう技術を使うと、先ほどの機械化というのはポジティブに取るか、ネガティブに取るかというのはあります。結構支援はしてくれるのではないかという気がしています。お忙しいのに2、3分しか取れないのにというのは、これは1時間機械が結構データを取ってくれると、ある意味では良いのかなという気もしないでもないです。例えば質問も、その答えをあらかじめ患者さんが覚えているというのだけでも、機械だったら色々変えることもできるのではないかという気がするのです。

もう1つは、生活の中で結構長時間データを集めてこられるというのも1つメリットとして、こういう技術があるとあるのかという気がしています。ポジティブに取るとその2点を思いました。

○岸本氏

ポジティブなコメントをありがとうございます。1つ目の点なのですが、これも私の紹介が十分ではなかったのだと思います。これは良い面と悪い面があって非常に難しいと思っています。我々がやろうとしている診断支援機器は、あくまで医師ないし心理士との会話の状態を観察して結果を出します。そこは逆に言うと、標準化がされていないもので、定量するのが本当に正しいのかという批判はお受けしています。同時に、むしろ我々はコミュニケーションをしっかりと、患者さんから症状を引き出すような臨床家の技量がないと成立しない医療機器でもあって、それはそれで医療機器として良いのかという議論は当然あると思います。そういう部分がありますので、そういうのを使いながら、かつ今まで我々が暗黙知として感覚で感じていた患者さんの状態を数値で表現する。良い感覚の先生はもちろんそれで良いと思うのですが、ある意味患者さんとそのデータを批評しながら使うということでも良いと思います。つつい精神科医の独りよがりになりがちな治療も、ある程度患者さんと状態について共有して、新しいディスカッションをしながら診療を進めていけると、むしろ良い方向に行くのではないかと。

それから、治療の経過の中で、患者さんが良くなっていないのを何となく気付いているのだから気付いていないのだから、延々と治療を続けるケースはあり得ると思います。そこをちゃんと評価して、本当に良くなっているのか、治療がうまくいっているのかどうかということ判断するという、新しい情報が入ること自体は、むしろ非常にポジティブな方向に行くのではないかと期待しています。

2点目は、本当に睡眠が取れているかどうか。そんな単純な情報ですら、非常に患者さんのそれぞれのバイアスが入っていて、結局我々が治療対象としているものが、患者さんから報告を受けることと、実際の出来事とは大分乖離がありますので、そういう客観的なデータを診療場に持ち込むことが、他の領域では当たり前のように行われていること、例えば白血球の数が上がった下がった、熱が下がったと当たり前のように行われていることが、なかなか精神科領域ではできていません。そういう新しい指標を持ち込むことで良い面があるのではないかと考えています。

○荒牧氏

治療、診断支援といった面でも非常に有効だと思います。もしかしたら私がこういう可能性があるかと思ったのは、患者さん自身に渡してあげて、患者さん自身が、自分で回復する過程を目で見えるようになると、私は今ちゃんと今治療が効いていると分かると、患者さん自身も非常に自分自身で自分を回復していくことができるかもしれないと思います。そういう患者さんが使う医療機器という可能性はないのでしょうか。

○岸本氏

はい、これもなかなか難しい面があるとは思ってまして、完全に世の中に勝手に出てしまっていて、望んでもない人が勝手に診断されるというか、そういう状況も非常に危険かなと思いますので、まずは完全に精神科医ないし、医者がコントロールできるような形において診療室で使うということを想定して開発しています。ただ、先ほどの回答と同じですけれども、結局は良くなっている、良くなっていないということがどうしても分かりづらい領域ではあるので、診療場面で計測したデータが、患者さんにお返しできるということ自体は、先生がおっしゃるように、患者さんに常にその状態をフィードバックしながら行う医療につながるのではないかと考えています。

○光石部会長 . はい、どうぞ。

○大江副部会長 ある種この精神医学療域の診断に関わる部分で、従来ならば専門職の技能の違いだとか、主観などにどうしても結果が左右される部分がある領域で、ある種別の違ったシステムで客観化をする仕組みを併用するという医療というのは、精神領域だけではなく他の領域でもこれから重要な役割を果たしていくのではないかと考えてお聞きしていました。

ちょっと技術的なこととお聞きしたいと思ったのは、医療の領域でマルチモーダルな機械学習ディープラーニングというのは非常に重要な領域だと思います。それに挑戦されているのだらうと思うのですが、特徴量の選択を事前に統計的な手法などで、選択をしてから掛けるのか。ディープラーニングのように特徴量の選択もネットワーク側で学習させて任せるのかで、すごく手法が変わってくるように思います。前者の手法だと、従来の医学の領域で多分このあたりが特徴量なのだらうと思うものを、やはり取捨選択しやすくなるのですけれども、思いがけない特徴量を利用するという意味ではディープラーニングのように、特徴量選択自体も学習させたほうが面白いと思うのですが、その辺を今回はどのようにされているのでしょうか。

○岸本氏

はい、ありがとうございます。ディープラーニングを行うには正直データセットが足りないのではないかと。もう少しトラディショナルな機械学習を用いて行っています。その特徴量を選ぶときにもモダリティとしては我々がやはり精神科医が、ここに一番精神運動抑制であれば、精神運動抑制が現われるだらうなというところを、より抽出しているつもりではありますので、そういう意味ではそこでかなり集約されているなと思います。



○大江副部長 では、将来的に N が非常に大きくなれば、さらにそういう特徴量の自動抽出なども考えられるということですね。ありがとうございました。

○加藤委員 よろしいですか。いかにも私がネガティブなことばかりというような論調ですので、一言だけ。決してそういうつもりはありません。今のお話でいうと、ディープラーニングにも関係しますが、あくまでインタビューや何かをベースにしているということは重要だと思います。そこは非常にポジティブに評価できると思います。というか、逆に言うと、ディープラーニングはなかなかできないだろうとも思います。そんなツイッターみたいにわっというような数はあり得ませんので、人間が 1 時間か 2 時間かけて患者さんから聞いているというのがデータベースですから、たかだか万にいくのも苦しいかもしれませんね。同じような、実は脳プロでやっている研究で、私どもは発達障害をやりましたが、うつ病のグループもいて、うつ病を同じように機械学習で、これは安静時の脳機能画像で判断できるかをやっています。まだこれは出ていないので、詳しいことは申し上げられませんが、うつ病を広く取ってしまうと、難しいようですね。ですからそういう意味で、できるというのは結構限定したほうが良いし、その辺をきちんと今の段階では人間の側が判定していくということは必要だろうと思っています。

○岸本氏 はい、ありがとうございます。それぞれの先生方、加藤先生らが行っていらっしゃる FMRI を日常的に取るのは非常に難しいと思いますので、日常的にある程度フォローに使えるような、もう少し浅いレベルでの評価ということにはどうしてもなってくるかと思っています。対象を実際に今のところはかなりメランコリックな患者さんだけ集めてきていますので、典型的な患者さんを対象にしていますけれども、ある程度広げて本当にどのぐらいうまくいくのか。発言内容と精神運動抑制との乖離が出てくることは、むしろ期待していますけれども、今のところは典型例に集約して、そして広げていく中でどのぐらい使えるかということは見極めていきたいと思っています。

○許委員 今のところ典型的な病気の例を対象にやっておられるというのですけれども、我々の世界の中でやはり本当に正常な精神の持ち主というのを、

最終的には定義することができるようになるのでしょうか。世の中には変わった人がたくさんいるのですよね。ボーダーラインというのは難しいと思います。けれども絶対的な病気と絶対的な正常の、絶対的な正常のほうはこの方法で定義できるようになるのでしょうか。

○岸本氏

そうですね、コメントが思いつかないのですけれども。周りや本人が、困っていらっしやらない人は基本的には正常で良いと思っています。正常と鬱病をどのぐらい正解したかというデータも出したのですが、そこが正直我々のゴールではなくて、やはり重症度、典型的なその症状、精神運動抑制に現われるような重症度を定量する。もう正常になってくると、それはほとんどゼロに近いので、もうそこはどうでもよくて、むしろ疾患のそれこそ個人の経過の中で、どのように指標が変わっていったかということをお返しするようなシステムにしたいなと思っています。正常の範囲をこの機械が規定するというものでは正直ないのです。

○加藤委員

今のような御質問の趣旨というか、そういう期待というか、非常に広いと思います。非常に多くの方がそういうことを色々思っておられると思いますし、だからこそ逆に危険があると、そういう先ほど言ったようなそういうビジネスが起こってしまう。決して性善説では世界はやっていけないので、そういう人たちがやってしまうようなところをちゃんと押さえておかないとまずいと思っています。

○光石部会長

それでは時間の制約もありますので、一旦ここで岸本先生の質疑応答というのは打ち切らせていただいて、次の議論に移らせていただきたいと思います。岸本先生、どうもありがとうございました。

次は資料3、資料4といったところですが、今までもどこまでAIと呼ぶのか、既存の機器との関係はどうかということ、多少議論はされてきたのですけれども、もう少しシステムティックに議論したほうが良いのではないかとということで、資料3、4、があります。これは佐久間先生を中心にまとめていただきましたので、まず佐久間先生から少し説明をいただいて、その上でこういうまとめ方で良いのかどうかも含めて議論したいと思います。佐久間先生よろしく願います。

○佐久間副審査センター長

既存のもの、それから今後出てくるものについて検討いたしました。資料3を御覧いただきますと、まず応用というところで分けていくと、診断支援とか治療計画とか治療支援、それからリハビリテーションその他の応用が出てくるかと思います。これは何人かの先生からいただいたものをまとめてみたということでありませう。

2番目の分類軸として、技術分類とかアルゴリズムの特徴という観点、それからAIシステムが出力した結果を人間が妥当かどうかを見ることが

できるのかどうかという観点があります。それからルールが AI システムより生成された場合に、結果に問題がある場合に修正できるのかどうかといった観点もあります。また学習も含めたそのプロセスの細部を設計者が理解できるのかどうかという観点からの分類もあります。

いわゆるデータを専門家がしっかり診断を付けた形で持つていくのか、今日お話があったように、全くそういうことがなくて、広く一般に存在するデータから学習データ取ってくるのか、といったデータ収集方法の観点からの分け方があるだろうと考えられます。

これまでビッグデータということを議論しているのですが、一方で物理シミュレーションや、医療知識がちゃんと構造化されてデータとなっており、それに基づくシミュレーションが比較的重要な位置を持っています。シミュレーションを基礎に持ちながら、支援をするシステムというものもあると考えました。例えば血管内圧の解析、冠状動脈内狭窄前後での圧力差を画像診断データから推測するような製品が存在しますが、そこでは流体力学のある非常に簡単なモデルが使われています。骨強度の解析には計算力学を使っています。診断支援の分野では、例えば心電図の自動診断などは、これは波形の分類ルールをしっかりと構築してシステムが構成されています。本日話題提供された精神疾患の診断支援で用いられている手法は、新しい形だと思います。

治療計画では、歯科の矯正のシミュレーターでは歯列をどのように移動させることが良いかを推定するものでありますが、その過程で形状モデルを用いるのか、一部力学モデルを用いるのか色々な問題解法があるのではないかと思います。角膜のレーザー手術計画は、多分形状があってそれに対してどうレーザーを照射すればよいかを、おそらく物理的に考えているでしょう。このようなシステムはある意味で、AI というよりもコンピューターを使った計画です。

放射線の治療計画システムでは、ある部位に線量を集積させる。一方で正常部位に放射線を当てないにはどうするかということで、照射方向とかを考えていきます。最近研究で出てきているのは、そういうものの計画と結果を見て行って、過去の優れた計画を参照しながら、人間がやったものを修正するような、一部人工知能的なアプローチが研究されています。研究的には股関節のインプラントの最適な形状を自動的に出力するものも検討されています。

治療支援では治療計画支援がありますが、薬物投与量の最適量を算出するシミュレーターとか、透析治療の計画を出すというようなものは、単体の測定で出始めてきているということだと思います。リハビリテーシ

ョンも、患者さんの回復度合いに従って、負荷を自動的に調整するとか、これが単なるバイオメカニクスから演繹的に出てくるのであれば従来型のシステムですが、臨床の色々な経験を取り込みながらコンサルテーションするといったシステムも出てくるかもしれません。

その他、自動切開とか処置をする。これは一部これが AI なのかと言われるかもしれませんが、バイポーラ電気メスで自動的にパワーをコントロールして、組織の最適な焼灼を行う製品が現実にあるわけです。これは AI ではなくて自動制御ですけれども、一般からは AI 的にみられるかもしれません。

技術分類については、この部会では主としてディープラーニングというのがかなり出てきてはいるのですけれども、外は異なる技術があるであろうということで、これらも列挙しておく必要があると思っています。アルゴリズムの特徴なども、アルゴリズムそのものが例えばルールそのものが少し増えてくるとか、これは御専門の先生に見ていただかなければいけないと思います。

この 2 から 7 については、資料 4 も見ていただいて、議論していただければと思っています。これは実は応用分野と技術分野と 1 次元に出るものではなくて、2 次元あるいは多次元配置して分けるべきかと思ったところがあります。多次元にすると非常に分かりにくくなるので、まず 2 次元で分けてみました。

まず応用分野というのは 1 つの形でできてくると思うのですけれども、技術分類としては例えば古典的な AI で言う推論エンジンみたいなもの、先ほどの ECG の自動解析などはそうだと思いますけれども、今日話があった自然言語処理というものとそういうものは似ているところもあるし、違うところもあるだろうと。これは多分独立にいうよりかなり相互に関係した分類になると思います。

ディープラーニングは他の学習アルゴリズム、例えば古いと思われるかもしれませんが、非通知モデルを使いながら、パラメーターを選ぶときに遺伝的アルゴリズムを使うとか、そういうことをやっている研究があると思いますので、そういうところもあるだろう。このあたりをどういうものがあるかということ、少し技術的に俯瞰しておくのも必要だと思います。

シミュレーションを学習に応用するという考え方もあります。実際に先ほど症例数が少ないという話があって、学習がしにくいという話がありましたが、シミュレーターだと色々なパラメーターを入れることができ、これを活用して学習を行うことが可能です。確か 1 回目の専門部会での松

尾先生のところでロボットを並べて学習するところを工場で学習させる。これは現場ではないのですが、モデルを使って学習させることをやる。そのときにモデルにパラメーターを振りながらやるということもあるので、そういう意味ではシミュレーションというのは結構重要だという議論があると思っています。

そのあたりの分類は、アルゴリズムとか実装の仕方のところがあると思っています。それらがある応用分野と組み合わせられて分類されるのではないかというイメージを持ちました。学習データの収集法は、本日の議論を伺うと、ここで分類したような簡単なものではないだろうと思っています。専門家がデータを選択して入力する。今日お話があったように、自動的にツイッターからデータを持ってくる。あるいはそこからもう一度それを自動的に分類するのだけれども、それをさらに専門家がうまくフィルタリングしてやると、色々なレベルのことがあると思います。

この点はデータの信頼性に関わることなので、学習データがどれぐらい信用できるのかということ、データの収集方法で少し分類していく必要があると思っています。それから結果の活用法による分類があります。これも今日議論があったと思うのですが、推論結果を直接患者さんに使われてしまうということ、あるいはそのまま、先ほど自販機化という話がありましたけれども、オートマチックに使われてしまう。

多分これが機器に組み込まれて例えば機器の制御に使われるということになると、これは完全にそれを使うということになる。専門家の診断の補助にするといったレベルで、どういうレベルがあるのかというあたりがあると思います。直接適用されないと書きましたが、結果の活用法においても懸念される点がかかなり変わってくるので、これによってリスクあるいは信頼性ということが変わってくるだろう。このあたりについて少し自由度を上げて記述しなければいけないのかなと思いました。

それから今日事前の議論の中で、結果の活用のところで市販後のメンテナンスと言いますか、そのやり方がオンラインでずっと学習されるのか。あるいはある程度データを現場で取ってきたものが、どこかのクラウドセンターに集中して出てきて、そこである程度データの厳選をされて元々のアルゴリズムをもう一回トレーニングして、ある時期それが評価された上でダウンロードさせて使うのかとか、さまざまなケースを考えなければいけないと思います。したがって資料では太線を用いて、やや異なる視点を記載しています。本当は多次元になるのだと思うのですが、これをやってしまうと難しいのでまずはお示しした形といたしました。

今まで出てきた議論だと、こういう整理の仕方では座標軸が決まるであろう

うという観点から、議論のたたき台として出しました。決してこれが完璧だとは思っていませんので、議論いただければと思います。以上です。

○光石部会長

ありがとうございます。委員の方々から御意見いかがでしょうか。

○武田委員

すごく良いまとめだと思う一方、色々言いたいこともあるのです。そういう意味で五月雨式にちょっと申し上げます。今日も学習型のシステムのプレゼンがあったのですが、特に荒牧先生の方でもクラウドに置いた色々なサービスはどうするのだという議論が出てきました。一方で機器としてどうなのかという話もあったときに、こういうシステムにおいて今後当然データが更新されて、むしろ向上することが期待されるような動的なシステムであるべきなのですね。そう考えると最初において初期状態である種の性能を出して、それを審査するというのは当然なわけですが、多分これは継続的にいわゆる審査を受ける仕組みをこういうシステムには持ち込まないといけないと思います。

つまりどのぐらいのタイムスパンが分かりませんが、そういうものが定期的にチェックされるような機構も付かないと、こういうデータが更新されてそれによって性能が向上する。場合によっては性能が向上しない場合もあるのでしょうか、そこをある種のクライテリアを作って、こういうタイプの機器は初期の審査もあるけれども、今後あるタイムスパンでもう一度再チェックがかかる、継続的にチェックがかかるといった仕組みを取り入れるべきだと、こういう動的データを扱う仕組みとしては重要だと、1点思いました。

もう1つ佐久間先生の中で、応用分野等で分類されて、これはこれで分かるのですが、一方で多分期待される応用分野を考えたときにもう1つ別の軸があると思いました。それは今やっている診断であり治療であるものの置き換えにあたるようなものの機器、サービスなのか、それとも全く置き換えでないようなサービスとなるのかということも、結構大きな違いかと思いました。例えば荒牧先生のものだと今までないようなものを作るという意味では、置き換えではないようなサービス、機器になるし、診断の補助という意味では今まで医者がやっていたのを部分的に置き換えるという意味では、そういうカテゴリーのものです。

これから AI 応用で出てくるものというのは、正にこれまでなかったようなものが出てくる。それもある意味当然受入れなければいけないでしょうから、そうするとこれも1つの応用分野を切る視点として必要だと思いました。一旦これぐらいで。

○光石部会長

ありがとうございます。とりあえずは良いですか。ではよろしく申し上げます。

○山根委員

武田先生のお話をちゃんと理解したのか、自分なりに歪めてしまったのか分かりませんが、非常に大事な話だと思います。ソフトウェアというのは機能をどんどん追加していくことができるということで、例えば佐久間先生が作ってくださった一番左の行です。第1バージョンは診断支援だけでした、第2バージョンは治療計画まで立てられるようになりました、第3バージョンは治療支援ができるようになりました、というのは有り得ることなのです。そうするとこの表についてコメントするならば、左端のところは分類しないでまとめて1つのコラムにしても良いのではないかという印象を持ちました。

そう考えるに至った例が、私の知っている数値流体力学ソフトが第1バージョンの診断だけで出てきたのが、どうも第2バージョンはディープラーニングの機能が入っているらしい。ということは先生が言われたように再審査しないと機能が猛烈にアップする、そこで評価方法も変わることが十分起きてくる。本当は、私は数値流体解析がとても大変なので、ディープラーニングをやっても無理だろうと思っていたら、もうカタログに書いてあるのです。びっくりしました。ということで、どんどんグレードアップしてくるソフトウェアに対応する審査体制というもの、先生は非常に鋭い点を指摘されたと思います。

○武田委員

一つだけ補足しますと、測定のアップデートはまだ分かりやすいのです。一応バージョンアップするという、明示的な人間が手を入れて高度を変えろという、それは明示的に言えるのですが、例えば学習データの更新といった問題になると、それは非常に見えづらいところにあるのですね。あるときスッと学生たちに入れ換えられるということは十分にあり得るわけですね。ですからバージョンアップでより問題を複雑にしています。ソフトウェアは全然変わらないのだけれど、投入する学習データセットが変わるだけで結果ががらっと変わってしまう。それをデータで固定するなんて言ってしまったら、逆にその性能が発揮できなくなってしまいます。そのためにはデータを変えても良いのだけれど、そのときにはどっかのタイミングでもう一回その性能評価が入ると、そういったものをプログラム化する。要するに審査側もプログラム化するという工夫が必要かなと思います。そういった意味で、データはもっとややこしいと思います。

○光石部会長

そのデータはどのぐらいでアップデートするのかとか、どのぐらいでダウンロードするのか、どこのタイミングで審査するのかというのは、審査も時間が掛かるようですと、またこれは結構大変な話になると思うのですが、そのあたりは議論かなという気がします。

- 鎮西委員 成長する AI の 1 つの派生形として、何々先生バージョンであるとか、あるいは地域差が入っている AI システムが出てくる可能性があるということ。製品としてはライフサイクルを考えると色々な形態、もう 1 つは、今度は先生方が作ったデータセットもバリューを生むものですから、これも経済的な商品としての対象になり得るわけで、現在、既にそういうことも少しずつ始まっているかと思います。現在の薬機法だと、そこまで縛りを掛けるかどうかという話はまだ結論は出ていないと思うのですが、将来的にはそういうことも考えなければいけないかもしれないといったところもあります。ライフサイクルの中にアルゴリズム、要するに我々が考える古典的なアルゴリズムとかシステムといったもののほかにデータそのものも入ってくる可能性があるかと。
- 光石部会長 タッチーですが、慶応病院で学習したものと東大病院で学習したものとなるのかとか、ならないとも限らない。
- 荒牧氏 1 点、補足なのですが、私自身が今、総務省の次世代人工知能ワーキンググループで構成員をやっているのですが、今、正に武田先生がおっしゃった問題で、作った成果物よりも学習モデル自身のほうが著作物なのではないかといった議論を行っていて、学習モデル自身がむしろ本題であるような感じのような認識になりつつあるという状態かと思っています。ちょっと補足しました。
- 橋爪委員 医療機器という観点から見たときに、今、既に行われているのは効能・効果ですね。もう 1 つは、使用目的ということです。ですから、あくまでそれを基本に考えて、自動更新することでそのあたりが変わることを明確にうたうのであれば、それは評価方法をきちっと考えないといけないと思います。自動更新できるシステムはあるけれども、それがある時点から、このように変わりますということを、PMDA なりにちゃんと評価を受けて、公表する上でもそれを出すということ。そこに縛りをきちっと置くべきではないかと思います。ですから、先ほどからおっしゃっていただいているように、AI の自動的に学習できて、どんどん性能が上がっていくという本来の目的、AI の良いところがそれで消えるようだと困りますが、あくまで医療側から見ると、それが分からないうちに効能あるいは効果が変わってしまうというのは非常に問題がありますから、そこが変わることにに対しては責任持ってそれを出す企業側が、その時点できちっと評価を受けないと認可できないようなシステムはやはり必要なのかなと、ちょっと感じました。
- 岸本氏 私どものシステムに限って言えばですが、オンライン学習のような常に結果が変わるものではなくて、教師つき機能学習をしていますのでち

らで新規のデータ入力を制御していて、あるときにアップデートしますというコントロール自体は可能だとは思いますが。そのときに、また一から臨床試験をやらなければいけないがというと、それは正直無理な話で、そのときに例えば過去のデータセットを用いて、もう1回新しいアルゴリズムで検証したところ、精度は落ちていないと。そういう形での承認が得られるとよいと思います。当然、新しくなるときにはチェックは必要だと思うのですが、そのような形でいくのが現実的なのかなと、勝手ながら考えていたのですけれども。

○橋爪委員 正にそのとおりで、それは一部変更で十分可能だと思います。そのあたりの仕組みづくりが、ワーキングなり、そこで出てくれば良いのかなと思います。

○光石部会長 資料3を佐久間先生に作っていただいている、私が言うのも恐縮なのですが、治療支援のあたりはすんなり入るのですが、番号を振ってある3から7のあたりが何となくごちゃごちゃとなっていて、このあたりはもうちょっと整理の仕方があるのだろうなという気がしています。これを資料3のこっちのほうにまとめるのが良いのか、資料4のほうにこういう項目を入れ込んでいくのが良いのか。むしろ資料4のほうに入れ込んでいくのかなという気もするのですが、どうですか。

○佐久間副審査センター長 光石先生、そのとおりで、私が悩んだのはそこです。結局、これはある意味で関連するものを、とりあえずまとめたものが資料3であります。まとめる観点として、技術分類というものもありますが、今言った学習ということではデータの取り方とも関連します。分類のためにはさらに細かな項目をもう1つぐらい作らなければいけないかもしれません。あと、結果の活用という形で分類していくと、先ほどの1次分析、1.5次分析といった部分での考え方を、ここに組み込めるのではないかと思います。そのあたりを御議論いただいて、4の所の大きなサブのグルーピングとしての視点として足りないところをご指摘いただけると、まとめる側としては有り難いという感じはいたします。

○光石部会長 資料4のほうは応用分野と書いてありますが、こういうものを見る人が必ずしもAIの専門家ばかりではないと思いますので、例えばアルゴリズムであれば、その特徴だとか、そういうものも説明があったほうが良いかなという気がします。あと、これは科学委員会なのであれなのですが、審査のポイントなどがこういうものに更にあると、例えば医療機器の開発者とか、そういう人もとても参考になるのかなという気もするのですが、これは書いて良いのか、ちょっと書かないほうが良いのかというあたりはあるのかなという気もするのです。そのあたりは、またこれ

からの議論かと思えます。

○佐久間副審査センター長 今回の科学委員会のまとめの議論で PMDA 側からの要望として考えられるものとして、先ほどおっしゃったアルゴリズムの特徴や限界の現段階の科学的知見のまとめだと思えます。これは審査側から見ればどれだけ信頼性があるのか、どこに弱いところがあるかといったところが明らかになることであり、有用な情報となります。この観点からは学習データの収集の仕方でもこれが信頼性のあるデータを取るためには、あるデータ収集方法にはこういう利点があり、こういう弱い点があるといった整理ができると良いのかなという議論だったかなと思えます。

○光石部会長 大丈夫でしょうか。そうしましたら、時間の関係もありますので、次に資料 5 が佐久間先生を中心に事務局側で作った、ある意味では報告書の目次にしてはどうかということなのです。これも配られていると思うのですが、横長で参考資料と書いてあります。これが前回も使わせていただいた議論なのですが、技術の俯瞰と課題抽出、新要素、規則、ベンチマーク、同承認の在り方、倫理・責任というのになっていて、資料 5 が目次案として出ているのです。これも佐久間先生から、少し説明をお願いします。

○佐久間副審査センター長 参考資料を御覧ください。非常におおまかな話ではあるのですが、報告書として、まず骨子として挙げることで技術の俯瞰と課題抽出があります。そこで資料 4 の技術などについて、先ほど光石先生が御指摘になられたような技術の特徴を記述することも含めておく必要があると思えます。

2 番目の新要素という所では、特に繰り返し出てきた、常に機能を更新する、向上する、あるいは場合によると劣化するものもあるといったところに対して、どういう新しいこれまでに考えられていなかったような要素があるのかといったこと。例えば学習の形態などの議論もここに含まれると思われま。前回からお話がありましたが、出てきた結果が本当に正しいのかどうかということについて、物理モデルとか、化学の知識の下に生成されたものであれば、科学的根拠は明確なのですが、様々な要素を総合的に判断したといったものの場合に、その客観的な評価をどのように行うのかという議論が今日もありましたが、ここではこれららに関連する議論を記述すべきであると思っております。

規制、ベンチマークと承認の在り方ですが、先ほどお話があったようなことを書くことを想定しています。先ほどの山根先生の、あるときにこういう機能を標榜してきたものが次回変わるということは、医療機器の中での機能を明確にどう定義するかというところで出てくることでしょ

うから、そのあたりとも関連する議論かと思っています。変更の管理と
いったところが、現行においてはどういう形があり得て、どういうもの
が妥当なのか、あるいは将来こういうものを考えなければいけないとい
ったことがあるだろう。

倫理・責任の所は、次々回のところで御講演いただく先生方に少しお話
いただくことになるかもしれませんが、これに関連することとすれば、
先ほど武田先生のほうからお話がありました、今のものを置き換えるの
か、支援するのか、全く違うものを考えるのかといった議論は、この部分
に記載されるものだと考えます。

それを資料5はこの議論のまとめ方をもとに、別添のところに書かれて
いるこれまで委員の先生方の御発言の中に出てくる論点やコメントを構
造化していくことができないかと考えたたたき台です。

<議題2：検討の方針と今後のスケジュールについて>

○光石部会長 ありがとうございました。資料の説明を先にさせていただくと、委員の
先生方だけだと思うのですが、タイムラインの案を配らせていただいた
のです。今まで何をやってきたのかということ、これから先どうする
のかということです。今日は3回目ということで、今までこういった先生
方に講演をいただいたということ、次回は鎮西先生に「医療用ロボッ
トの自律性、situation awarenessに関する国際標準での議論」というも
のです。第5回は7月になりますが、
どちらか
という先ほどの言う第4章というか、そちら側に関係するものには
なっているということです。今日の所の右側を見ると、第4回までにどの
柱を執筆希望かアンケートを取り、最終的な担当を決めると書いてはあ
るのですが、目次が決まらないことには、どこを執筆すると言ってもな
かなか難しいかなという気もしているところです。これで見ただくと
分かるように、夏休み頃に色々執筆の宿題が出そうだとということが見
て取れるかと思えます。このスケジュールと先ほどの資料5等の目次あた
りについて、あと10分ぐらいで御意見をいただければと思います。いか
がでしょうか。

○鎮西委員 厚生労働省のAI懇談会がありましたよね。あれとの重複などは、これ
でもう取れているのでしょうか。

○大江副部会長 あちらのほうは、3月末に報告書まとめを出すという当初日程だったの
ですが、3月末の最終回では少し先送りをして、3月末の時点では今後の
議論の素材という形で提示された段階ですので、そちらのほうは今後の

資料状況を見て、そちらと調整していきたいと思っています。

- 光石部会長 繰り返しになるかもしれないのですが、1章は俯瞰と課題抽出で、2章が割と新しく出てきている技術的な話を書いて、3章あたりがどういったところに使うのかという、臨床的な位置付けとか、評価の仕方などを書いて、4章で責任・倫理という感じということですよ。
- 大江副部会長 今日の岸本先生の最後のほうのスライドを見ていて、ちょっと追加したほうが良いかなと思ったのは、おそらく今後の処理はスタンド・アローンではなくて、クラウド上の処理系に頼らないといけなくなるのではないかと。そうすると、処理系そのものがクラウド上にあるシステム、医療機器、ソフトウェアシステムの課題というか、担保しておかないといけない条件は何かということ、どこかで記載する必要があるかなと思いました。場所はどこが良いのか、ちょっと分からない。新要素なのでしょうか。この辺は佐久間先生と調整いただければと思います。
- 光石部会長 クラウドもリアルタイムでつながっている必要があるのか、オフラインになっていても良いのかと。オフラインになっていてもよくしておかないと、ちょっと危険という感じはするのですが、データ量から何から全部、手元に持ってこられるかどうかというのは分からないですよ。
- 大江副部会長 もはやオフラインで動かすというのは無理なシステムも増えてくるのではないかと思いますね。
- 武田委員 これは■■■■のあたりで出るかもしれないのですが、一応、最近の人工知能の倫理の話が議論できているのは、暴走するかどうかみたいな議論が必ず出てきて、私どもはそういうのは議論としてあまり適切ではないなと思う一方、その懸念は社会的にはかなり指摘されるので、ある種のシステムの誤動作になってしまうかもしれないのですが、そういった外れたときの対応をどう考えるかみたいなことも少しどこかに入れるとよいかもしれないですね。システムが壊れたときという言い方はおかしいのですが、いわゆる意図しない方向になるとき。もちろん、本来正しく動くはずだという前提は良いのですが、動かないときにどうするのだという問題はより深刻にはなると思うので、それもどこかでやれば良いかなと思います。
- 光石部会長 重要な点だと思いますが、それは技術的なところで書くとしたら、多分2章になるでしょうし、もう少し医療との関わりということで書くことになれば、今ここに書いてある佐久間さんの3.4がそれになるのですかね。これはもうちょっと違うことを書くのですか。
- 佐久間副審査センター長 そこまで十分やっていませんが、多分リスクということとすれば、先ほどクラウドということがありましたが、この手のシステムが

今までの医療機器とちょっと違うのが、設計をやって、それをしっかりした製造ラインで作って評価をやって、その後それをちゃんとメンテしてということなのですが、そのメンテした人は運用する人のところの、運用の適正性みたいなものが実は入ってくる話なので、この議論はソフトウェアシステムの国際安全規格にその拡張の議論が入ってきているので、そのあたりはもしかすると議論になってくるところかなとは思いますが。

- 光石部会長 技術的には多分未完成であっても、多少エラーというか、外れるものが出るにしても、それをうまく使い方をすれば医師がカバーできるという使い方が当然あるわけです。
- 佐久間審査センター長 医療機器における安全の定義というのは先生が指摘された考え方が含まれていると考えられます。結局受容できるリスクと安全性の兼ね合いで議論を展開します。残留リスクをどこまで許容するのか、またどの程度まで抑え込むのかというのは、重要な議論であり、応用分野によって考え方が変わってくるところがあります。
- 武田委員 そうですね。広い意味でリスクですが、今までみたいに割と一元的にここで切るというよりは、リスクとベネフィットをどこで切るかというのを、逆にそれを定義するのがある種の規制なり、そういう仕組みだというので、確かにそこを3章あたりで少し、逆に言うとリスクを高く取る分ベネフィットもある、そういう装置もあり得るとか、そういう議論はここでできたら良いかなとは思いますが。
- 光石部会長 ありがとうございます。例えば第1回目で自動運転のところの話もいただいたのですが、そういう既存のものは1章に書くとすれば書くのですかね。厚生労働省のほうにも、そういうのはもう当然書いてあるのですか。
- 大江副部会長 この1章の技術の現状という所に短くまとめるしかないかなと思います。
- 光石部会長 車でもそうかもしれないのですが、医療の場合には自動というものは少ない可能性もあるというか、そうしないと安全性が担保できないかもしれないというのがあって、そのあたりがいわゆる機械が全部自動で動くものとは大分違うかなという気がしますね。使い方だとは思いますが。
- 武田委員 その辺も機器の種類とか、当然あり得るわけですね。例えば適切な点滴をやるための最適化するプログラムが、若しくはそういう措置が組み込まれるというものは、AI的に学習して動くということもあり得て、それはかなりリスクを押さえたときは自動運転もできます。そういう意味では、先ほどの資料4の結果の活用法というのは、確かに自動運転と同じようなレベル分けはできそうですね。それは十分、完全自動運転も、機器

の種類によってはあり得ると思います。

○原田委員 自動という言葉なのですが、ソフトウェアではなくて、特にハードウェアのロボットとか、鎮西先生が講演される予定の内容に含まれると思うのですが、自動は必ずしも学習を使わないというか、学習を使うもののロボットは自動が多いと思うのですが、自動という言葉の議論と今回の議論の学習の話はしっかり明確にして、議論を進める必要があるかなと思いました。

○光石部会長 だから、自動というものの定義できるのかどうかというのもあるのですけれども。

○原田委員 例えば手術ロボットの自動化の話は、別のガイドラインとかでも既に議論もされているところだと思いますし、自動で動くときに一切、今、学習していないものなども結構あって、計算して自動で動くわけなのですが、そういうものは多分今回のスコープからは外れるのではないかと思います。ただ、一般の方がロボットで自動というと、もう AI というように思われてしまうと思うので、そのあたりの区別をしっかりと、今回議論するのはこのスコープだけですよということをクリアに。

○光石部会長 だから、そこはむしろそういうことを明確に書いておかないといけないということかと思います。

○原田委員 そうですね。

○武田委員 ただ、それをやろうと思うと、その区別は意外に難しいのですね。例えば一番初歩的なのはフィードバック制御だって自動と言えば自動制御であって、それをちょっと拡張して、過去の室温をベースに室温をコントロールするのも、ある種の自動と言えば自動。そのときに過去の1日分のデータを取ったら、それはもう学習データなのかと言われたときに、境界線は危ういのです。もちろん、どこかで議論は分けなければいけないのです。ただ、そこは必ず明確に分けられるというよりは、意識して分けるというようなことになるかと思います。

○光石部会長 ありがとうございます。そうしましたら、資料5がある意味では目次のたたき台になりますので、これとか、もちろん資料3、資料4についても意見をいただければと思いますので、これは事務局のほうから改めて委員の先生方にメールをいただいて、それで意見を出していただくことになるのと、もしここは是非、自分が書きたいということがありましたら、それもお知らせいただければと思います。過去の例では、ページ数が全部で10~15ページとは聞いていて、目次だけでも2ページになってしまうのではないかという気もしないでもないのですが、多い分には構わないということです。ただ、一部分だけ多いというのもバランスを欠くかも

しれないということです。

次回なのですが、先ほどのタイムラインで見ていただくと分かるように、5月24日、鎮西先生ということですが、何かアナウンスしておくことはありますか。

○鎮西委員 : 私の話は、おそらくもう少し先の話。具体的に言うと、治療計画であるとか、そういう患者さんに直接影響を及ぼす医療機器に、こういう人工知能、あるいは統計的な判断アルゴリズムが入ってくる場合に、それをどう分類して、どう考えるかという話が最近少しずつ出てきています。IECのテクニカルレポートという形でまとめた文書が、もうすぐ発刊される予定になっております。そういったものを紹介しつつということです。今、最後に出てきたような自動、オートマチックとオートノミーの微妙な違いなども定義されておりますので、そういったところを紹介していこうと思います。

○光石部会長 : ありがとうございます。そうしましたら、本日の議論は以上ということで、事務局のほうから連絡事項ということでよろしく申し上げます。

<議題3：その他>

○事務局(江原) : 事務局です。1点だけ御連絡です。次回の日程ですが、5月24日(水)16時からを予定しておりますので、どうぞよろしくお願いいたします。

<閉会>

○光石部会長 : 以上となります。本日はお忙しいところお集まりいただきまして、どうもありがとうございます。これにて閉会とさせていただきます。どうもありがとうございました。