

交通政策審議会第26回気象分科会

平成30年4月24日

【総務課長】 それでは、定刻となりましたので、ただいまから交通政策審議会第26回気象分科会を開催させていただきます。

事務局を務めさせていただきます総務課長の黒沢です。よろしくお願いいたします。議事に入るまでの進行を務めさせていただきます。

審議会委員、臨時委員の皆様方には、お忙しいところを気象分科会にご出席いただきまして、まことにありがとうございます。

まず、委員の出席を確認させていただきます。

本日、矢ヶ崎委員からは欠席との連絡をいただいております。家田委員からは、おくれる旨、連絡が入っております。30分程度という連絡でございました。

交通政策審議会気象分科会の定足数につきましては、交通政策審議会令第8条によりまして、委員または議事に関係のある臨時委員の過半数の出席をもって会議の定足数とされております。本日は分科会の委員総数9名中8名の出席をいただくこととなりますので、分科会が成立しておりますことをご報告申し上げます。

今回は新年度最初の分科会でございますので、気象庁職員の新任者の紹介を最初にさせていただきたいと思っております。

地震火山部長の土井でございます。

【地震火山部長】 土井恵治と申します。よろしくお願いいたします。

【総務課長】 参事官の葦澤でございます。

【参事官】 葦澤でございます。よろしくお願いいたします。

【総務課長】 参事官の長谷川でございます。

【参事官】 長谷川でございます。よろしくお願いいたします。

【総務課長】 以上3名でございます。よろしくお願いいたします。

次に、お手元の資料の確認をさせていただきます。

今回からペーパーレス会議のシステムを設置しております。使用方法につきましては、事務局のほうの担当がいますので、不明な点がございましたら、お申し付けいただければと思っております。今回は初めてでございますので、紙の資料についてもご用意してございます。適宜、申し付けていただければと思っております。よろしくお願いいたします。

それでは、紙の資料の確認だけさせていただきます。まず、第26回気象分科会の次第がございます。縦判でございます。それから、委員名簿がございます。それから、議事資料としまして、横判の右上に、資料1、資料2、資料3、資料4でございます。

以上でございますが、お手元がないものがございましたら、事務局までお申し付けいただければと思っております。

本日の会議につきましては公開されております。傍聴が認められております。また、庁内のテレビ会議システムで地方官署にも共有させていただくことになっておりますので、よろしくお願いいたします。会議後に、速やかに資料及び議事録の公開が行われますので、あらかじめご承知おき願えればと思っております。

連絡事項でございますが、前回もお願いしてございますが、マイクの使い方です。ご発言いただく際には、ご面倒でも、マイクの台の部分のボタンを押して、ご発言をお願いいたします。また、ご発言が終わりましたら、再度ボタンを押しまして、スイッチを切っていただければありがたいと思っております。

以上、連絡事項等でございます。

報道の方々のカメラ撮りはここまでとなりますので、以後の撮影はご遠慮いただきますよう、よろしくお願いいたします。

それでは、議事の進行につきましては、新野分科会長にお願いいたします。よろしくお願いいたします。

【委員】 おはようございます。それでは、議事に入らせていただきます。

前回の会合では、検討の背景となります社会的な課題や科学技術の動向を踏

まえた上で、中長期を展望した気象業務の方向性についてご議論いただきました。今回、この方向性の実現のために、重点的に取り組むべき方策のうち、「気象業務に関わる技術開発」について議論したいと思います。

議事の流れですが、前回の会合でいただいたご意見を踏まえ、事務局において気象業務の方向性を再整理しておりますので、まずはこちらを確認したいと思います。続きまして、気象業務に密接に関連する科学技術の現状と展望についてですが、前回、委員より、IoT等の最新動向に関して、大変興味深いお話をいただいたところです。今回、気象業務に関わる技術開発への活用も期待される人工知能につきまして、委員よりご説明いただくことになっております。最後に、「気象業務に関わる技術開発」につきまして、事務局が作成した資料をもとに、議論していただこうと思っております。

それでは、前回いただいた主なご意見について、気象庁より資料の説明をお願いいたします。

【企画課長】 企画課長の森でございます。私のほうからご説明をさせていただきます。

最先端の科学技術というほどではないんですけど、タブレットがあります。タブレットか紙なので、最後のページのところ、紙の26ページをごらんいただくと、今、新野分科会長から、これまでの流れをご案内いただいたところですが、タイトルは、「今後の審議予定」となっているスライドが、全体の流れとなっています。

今、分科会長からご案内いただいたとおり、2回目では、気象業務の方向性についてご議論いただいた。3回目、今日ですが、気象業務に関わる技術開発を、4回目のところでは、重点的に取り組むべき方策として、さらに、気象情報・データの利活用推進をご審議いただいて、8月の提言に向けて進めさせていただきたいと考えさせていただいているところでございます。

それでは、前回の主なご意見というところに参りたいと思います。資料1でございませけれども、タブレット、また、資料の3ページをごらんください。

これは第2回目の主なご意見についてでございます。後ろのページのほうで、

今日のテーマであります技術開発については、抽出した形で記載させていただいているところです。

前回の主なご意見でございますけれども、まず、気象業務の現状等についてであります。

この冬のところで、大雪もございましたけれども、そういう社会的に大きな影響をもたらした事例があった場合には、もう少し何かしておけば影響を軽減できたかもしれないという点があるんだったら、気象当局側と交通当局側の共同のチームを組んで積極的な方策を講じるというのもあるのではないかとお話がありました。

実際のところ、この冬の大雪を踏まえて、国交省では鉄道局さん、道路局さんのところで、それぞれ有識者による検討会、あるいは検討委員会を立ち上げているところでございまして、ここには気象庁もオブザーバーといった形で参画させていただいています。

ちなみに、鉄道局さんのほうは、まだいらしていないかもしれないけど、家田先生が座長を務めておられると伺っているところでございます。

次も、大雪に絡めたところですけども、例えば、ここの道路は通れませんといったようなことであつたら、迂回路として可能性のある方面をあわせて示すなど、パッシブより一歩進んだ戦略的な業務の推進をしてはどうかとあります。

今度は火山ですが、火山というのはなかなか予測は学問的に難しい段階だということであろう。そういった中で、長期的な火山の観測や噴火予知の体制、今後10年間といったものについて、ありようや意欲を積極的に示すべきだと。有珠山のような成功事例もあるので、火山の専門家の育成が大事である。

生産性向上の観点では、伸び代がまだまだあるところで、気象データを利用する企業を増やしていったらどうか。核となる企業や業種群をつくって、伸ばしていくのがいいのではないかと。

また大雪の話に戻りますけど、そういったときに、今度は、言ってみれば、情報の利用者側ですけども、車の運転をするお立場の方でもリテラシーを上

げていくことも重要ではないかという話がありました。

予測情報の精度は、どのようなものであるのかといったところにご理解いただくことが大事であろう。1カ月以上先でなく、もっと手前の予測精度を向上させるところまで、現状認識として妥当なのではないかとありました。

次のページへ参りますけれども、気象業務にも密接に関連する科学技術分野の現状・今後の展望です。

スパコンはハード、AIはソフトと分けずに、こういったものを融合させて考えるべきだ。それによって画期的に性能が向上したり、計算時間が短縮できる可能性もある。

基本的に、数値予報というのは、物理法則によるシミュレーションというところがあるわけですが、AIで予測を行える可能性がある。

その次にもありますけど、数値予報モデルで表現できないような部分へうまくAIを利用している部分があるのではないかと。

AIというと、よく、「識別」、「予測」、「実効」という3つに分けたりするんですが、その辺のところは慎重に切り分けして議論すべきである。

前回、委員から貴重なプレゼンをいただいたところですが、その中で、緊急地震速報というのは世界最高峰のIoTであるというお言葉も頂戴しました。

IoTというのは、現実世界からデータを取得する過程、データから知識を創出する過程、知識を現実世界に戻す過程がありますが、最後の過程が一番重要ですというお話があったかと思います。

データ・情報の提供については、品質というのは非常に大事なんですが、何でもかんでも品質を高めるじゃなくて、ベストエフォート型に移行することによって、品質保証型に比して平均的な品質を格段に向上させることができる。一方で、これは品質保証しないといけないよねというのは、例えば生命にかかわる分野、防災といったものがそうかと考えますが、そういったところの観点もあるでしょうというもの。

IoTを通じた流通については、標準化というのは難しいので、技術をオー

ブンにすることでソフトの開発効率を向上させ相互接続を確保していくしかないんじゃないか。

ビジネスの視点からは、予測技術というのは今後どの程度伸びるのか、その見通しを公表していただきたいというふうにあります。

3 ページ目ですけれども、科学技術を見据えた気象業務の方向性についてです。

100%の予測というのはどうしてもできないので、やっぱり減災に力を入れるべきで、AIが活用できる部分もあるだろう。

精度については、漠然とただ精度というのではなくて、空間的なもの、時間的なもの、インテンシティー（強度）的なものは使い分けて見るべきである。そういったところで、よりよい国民生活を誘導する側のプロアクティブな対応を行う方向への転換を打ち出してもよいのではないか。

その他、国民の皆様がふだんから気象情報に関心を持っていただくことも大事なので、2020年のオリンピック・パラリンピックはいい機会なので、そういったところで積極的な発信を行うことも一案である。

観測・予測精度向上の取組、それ自体は非常に重要な方向性だけれども、世界最高精度を何のために求めるんだろうかといった目的を明らかにすべきである。

方向性といったところだろうけれども、気象情報が楽しく役に立っていると思いつつ使っていていただくことが大事だ。観光とかそういう観点もあるんじゃないのか。

観測データ、民間事業者のデータについては、ビジネスになるということであれば料金が発生することもあるでしょう。

一方で、気象に関しては、空気中の情報は国民共有の財産だといった理念を、審議会でそういう観点で見ること大事だと話がありました。

その他、正確な情報に加えて、市町村等の信頼関係の構築、リテラシー向上等の取組など、地域防災のヘッドクォーターとしての役割を地台に担ってほしい。地域防災体制の中核となっていくという今後10年の意思表示を明快にし

ていただきたいと話をいただきました。

本日は、技術開発を中心にご議論いただくこととなりますけど、いただきましたご意見につきましては、提言に向けまして、可能な限り反映していきたいと考えているところでございます。

次のページから、業務の方向性の再整理というところで、基本的には前回の資料と、資料自体はほとんど変わっていないところです。

あり方としては、気象データ・情報が、防災・日常生活・社会経済に必要なソフトインフラとして社会のさまざまな方面で活用されることで、一人一人の生命・財産が守られ、誰もが生き生きと活力のある暮らしを享受できるような社会に寄与したいといったところで、下にありますとおり、観点ということでユーザー目線、それから、社会的ニーズの把握、それを踏まえた水準としてはどんなものか。さらに、不断の改善を進めていく。

その中には、観測や予測精度の向上に関する技術開発を進めるものもありますし、データ・情報を利活用いただくための取組、リテラシーの向上もありますけど、そういったものを進める。必要なときにデータ・情報を利用できる取組や環境整備を進めるといったところではないかということです。

次のページも、前回の絵とほとんど同じで、今言った観点などが真ん中の楕円の中に入っていて、上のほうが防災、下の左のほうが日常生活、日々の生活、右下のほうが産業活動、こういったところで役に立つようになるのではないかということです。

ちなみに、前回、上のほうの外国人旅行者等のところに、仲居さんの絵しかないというところがあったので、ガイドさんの絵も入れてくださいというので、それは入れてございます。

最後でございますけれども、今度は、キーワードとなった社会的ニーズ、それから、目標を定める、不断の改善、連携といったものについて、真ん中にありますけど、それで、観測・予測精度向上の取組、リテラシー向上の取組、利用環境向上の取組といった観点では、それぞれこういったものが考えられるというところで、下にありますとおり、ゴールとしては、先ほども申しましたけ

れども、気象情報データが社会のソフトインフラとしてさまざまな場面で生かされることで、国民の安全・安心の確保、生産性の向上、豊かな暮らしの実現へ寄与を目指していくのだらうというお話だったと考えているところでございます。

前回いただいた主なご意見と整理としては、以上のところでございます。

【委員】 どうもありがとうございました。

ただいまのご説明に関して、補足あるいはご質問等もあるかとは思いますが、それらに関しては、本日の3番目の課題の「気象業務に関わる技術開発」に関する議論のところでまとめてお願いしたいと思っております。

そこで、次の資料2をご覧いただきたいのですが、資料2の2枚目の絵に示されている「気象業務にも密接に関連する主な科学技術分野」のうち、前回、この図の下のほうにあるIoT等の技術を、委員からご紹介いただきました。本日は委員より、人工知能技術の最新動向についてご紹介いただくことになっております。委員、どうぞよろしくお願いいたします。

【委員】 よろしく願いいたします。

人工知能技術というと、世の中、華やかなイメージが強いかと思いますが、その華やかなレイヤーと、あと、基礎の地味なレイヤーがありまして、今日は地味なレイヤーのほうのお話をさせていただきます。

ちょっと自己紹介を、最初に簡単にさせていただきたいと思うんですが、私は今、理化学研究所のAIの研究センターのセンター長をしていますが、そこでは、いわゆるポストクの研究員を雇っていますので、職業研究者とともに研究をしています。

一方、大学でも教員をやっています、そこでは、修士とか博士の大学院の学生を中心に研究をしています。

まだ、それだけでも人が足りない状況ですので、いろいろな会社、技術顧問のようなことをさせていただいて、企業の若手エンジニアとも協業するということをやっています。AIの分野は世界的に人が足りない状況なんですが、我々の身の回りでも、圧倒的なニーズがあるにもかかわらず、それを研究できる人

はほんのちょっとしかいないという状況ですので、あらゆるチャンネルを使って人を集めているという状況です。

私自身、人工知能、正確には機械学習と呼ばれる分野の数学的な基礎研究をやってきておりまして、いわゆるコンピューター科学とか統計学にかかわるような研究分野で活動してきています。いろいろ英語で専門書も書いたりしています。

私、もともと工学部の出身ですので、研究そのものは基礎研究なんですが、それを実世界の応用に生かしたいという気持ちは常に持っているんですが、残念ながら、自分自身ではとてもそんなことはできませんので、プロのパートナーを見つけて一緒にやろうということで、いろいろな企業の方や国の研究所の方と連携しまして、画像の処理とか言語の理解とか脳波の解析、ロボットの制御、医療とか生命のデータ解析、こんな分野で応用研究もしてまいりました。最近、AIのブームがあったおかげで、いろいろ教科書を書く機会にも恵まれました。

今、世の中としては、ディープラーニングとビッグデータというキーワードが非常に華やかに盛り上がっていますが、研究としましては、それだけではないかなというのが今日のお話の趣旨にもなります。

ディープラーニングというのが何かというのは、非常に解釈するのが難しいところなんですが、この絵の中では、モデルの軸と解釈しています。昔々は、 $y = a \times x$ みたいな線形モデルで、身長と体重は正比例しますよというところからスタートしていて、でも、線形だと複雑な予測はできないので、非線形にしましょう。でも、ただ単に非線形にすると計算が難しくなってしまうので、いろいろ工夫しましょうということで、加法モデルであったり、カーネルモデルであったりといったものがこれまでずっと研究されていまして、最近、より複雑な深層モデルというのが使われるようになったと解釈しています。

一方、機械学習の手法というのは、モデルだけで決まるわけではありませんで、そのモデルにデータを実際に入れるわけですが、データを入れてどうやって学習するかという、学習法のアルゴリズムそのものが重要になるわけなんで

すが、それが縦軸にございまして、それがいわゆる教師付き学習とか、教師なし学習とか、強化学習と呼ばれるものがあるんですが、今日のお話は縦軸のほうのお話をしたいと思います。ですので、表向き、ディープラーニングは全く出てこないんですが、基本的には、全てディープラーニングと組み合わせることができるという話になっています。

話題を2つ紹介したいと思っているんですが、いろいろなところで必要になるであろう異常を検知するという問題と、ただ単に分類するという問題ですね。すごく基礎的なお話をしたいと思います。

異常検知の問題は何かといいますと、ざっくりデータがたくさん、観測値がいろいろ与えられたときに、1つだけ変なものがあったとして、これを見つきたいという、それだけの問題です。これはいろいろな研究が過去にありまして、ちょっと細かいことを書いてきてしまったんですが、有名な方法は幾つかあるんですね。

例えば、局所異常因子法というのが、これは30年ぐらい前に提案された方法なんですが、今でも結構よく使われているんですが、何か点が与えられたら、点のたくさんあるところは密度が高くて、点がまばらにしかないところは密度が低いとしたときに、密度が低いところにある点は異常度が高いだろう。あと、すぐ近くにたくさん点があるときに、ちょっと離れると、これも異常っぽいだろうというようなイメージで、異常を見つけましょうという方法なんですが、こういう赤いバツが実際に与えられたデータ点だとしまして、青い丸がその点の異常度で、これはすごく異常度が高い、これはそれほど異常度が低いという感じで見るんですが、こういう感じの結果になるんです。

結局、この方法は、今でもよく使われているんですが、何も学習はしてなくて、与えられたデータをもとに、何か事前に決めた計算式に従って計算をして、スコアを出すと、こういうスコアが出てきますということだけですので、やったらすごくうまくいくこともありますし、全然だめなこともあるわけなんですね。結局、データの特性というのは全く生かしていないということになります。

もうちょっと学習しようというようなアプローチも、これまでにいろいろ、当然あるわけなんですけど、最もナイーブなものは、ただ単にデータの密度をちゃんと推定しようということで、いろいろな方法がありまして、例えばデータ点が与えられたときに、そこになめらかな関数を置いて、足し合わせると、こんな感じで密度の関数が得られますので、こういうものをつくってやって、密度の低いところから出てきた点は異常とみなそう、密度の高いところから出てきた点は正常だとみなそうという方法も当然使われているわけなんですけど、これは非常に実は難しい問題でして、この関数を、精度よく推定するというのがものすごく困難だということが知られていますので、実は、実際にやると全然うまくいかないという状況です。

そこで、それを何とか避けようと提案された方法が、データ点がたくさん与えられたときに、そのほとんどを含むような円をうまく求めてやって、その外にはみ出たものを異常にしようといった方法も提案されていまして、これですと、関数を推定する必要がないわけなんです。密度を推定する必要がありませんで、実際にはレベルセットという、切ったところを求めていることになるんですが、それを求めるので、より解きやすいということで、いつか注目されたんですが、結局これも、どの点が異常かというのを明示的に学習しているわけではないわけですので、ある意味、あらかじめ決められたルールのもとで、何かを推定しているんですが、うまくいくか、うまくいかないかは、やってみないとわからないということになります。

ですので、結論としましては、実はさっきの問題は解けませんということなんです。データ点が与えられて、そこから異常を見つけたいというのは、みんなそれを当然、解きたいと思っているんですが、数学的にはそれはちゃんと定義された問題ではないので、結局、ちゃんと解くことはできないということですので、我々のアプローチとしては、もうちょっと情報をちゃんと足してやりましょう、数学的にちゃんと解ける問題をつくりましょうというのがスタート地点でして、それを弱教師付きの異常検知と呼んでいます。

さっきの問題では、全く何もラベルがついていない。ただ単に点だけが与え

られて、どれが異常かというのを見つけないという問題だったんですが、それだけではまともにできませんので、ここではちょっと問題設定を変えまして、もう一つ、データ集合を持っているとします。そのデータ集合を今、「 x 」と書きましたが、これは全部正常だとわかっていると仮定します。イメージとしては、例えば工場で製品をつくっていて、異常かどうかを判定したいという問題がよくありますが、そういうものを解くときに、例えば、きのうは一日、全然異常がなかったとわかっていれば、きのうのデータをとっておけばいいんですね。きのうのデータがこれになりまして、今日のデータがこれで、きのうの正常のデータを使って、今日のデータから異常を見つけないという問題を解こうということなんですね。

結論としては、こういう設定を考えると、数学的にちゃんと解けるようになりますということなんですね。この問題を解くときも、ナイーブには、正常なデータが既に与えられているので、これの確率密度を推定してやれば、ちゃんと異常検知できるんですが、それはやっぱり精度が悪いので、テクニカルにはそこは何とか回避しようということで、我々がつくった方法は、密度の比を推定するという方法なんですね。

正常なデータ、きのうのデータがこんな分布をしていたとして、今日のデータがこんな分布をしていたとして、これらの分布はわかっていないんですが、データ点を持っているだけなんですが、今日のデータに、実はここにちょっと異常がありますという状況を考えます。これを見つけないわけなんですが、もともとデータがほとんどない密度の低いところで、値をちゃんと推定するというのは非常に難しいわけなんですが、この2つの関数の比をとってやるんですね。

そうしますと、ゼロのあたりは、例えば正常ですので、ほとんど値は同じですので、比をとるとほぼ1になるんですが、異常があるところ、ほんのちょっとだけ突起があるんですが、もともとこっちはすごく小さな値ですので、比をとってやると、1から大きくずれるんですね。そうすると、すごい高次元の空間で、この小さな突起を見つけるのはものすごく難しいんですが、こっち側に

持ってきてやって、大きなへこみを見つけるというのはすごく簡単にできるわけなんですね。すごく簡単にできるということは、少ないデータ点からでも精度よく学習できるということなんですね。

ほかにもいろいろテクニカルに細かいことはあるんですが、こういう方法を使うと、非常に精度よく異常検知ができるということがわかりまして、実際これは、もともと考えたのは10年近く前なんですけど、いろいろな会社に使ってもらっています。製鉄プロセスで、何かつくっているところに、マイクを並べて、異常が起こらないかどうかというのを検知したり、あるいは、さっきの工場の例とか、プリンターのローラーをつくっている、今では、まだ人をいっぱい雇って人力で検査しているんですけど、それを自動化しようということであったり、あるいは、ローンを借りに来たお客さんの、その人は大丈夫かどうかというのを審査するのに使ったり、そういうところで活用されました。

実は全く同じ技術が、変化検知というものにも使えまして、時系列データが与えられたときに、それが突然傾向が変わって、変化があったかどうかというのを検知するのにも、実は使うことができまして、そういうものを活用すると、例えば画面の中から何か注目すべき物体を見つけましょうという問題であったり、あるいは、動画の中から何かイベントを見つけましょうとか、あるいは、ツイッターのデータから、何かイベントが起こったかどうかというのが判定できたり、そういう応用も全く同じ技術で使うことができます。

ここまでのまとめとしまして、比をとってやるというのがキーだったんですが、結局、1つのデータセットだけ与えられて、異常を見つけましょうというのは、みんな当然、それを解きたいと思っているんですけど、それは解けないということなんですね。ですので、そのまま頑張って、何となく解けたようにして我慢するか、ちょっと問題設定を変えて、追加で正常なデータを集めてやるというところまで我慢してやれば、ちゃんと解くことができるということなんですね。

そのときに、比を使うとうまく解けますというのが今回の方法の趣旨だったんですけど、実用的な面から見て大事なものは、ただ単に1セットのデータだけか

ら異常を見つけようとする、結局、主観的に答えを決めているのとほとんど変わりませんので、チューニングのパラメータをちょっと変えてやると、どの点でも異常にできてしまうんですね。ですので、先に答えを聞いていれば、それに合うようにチューニングしてやれば、合っているように見えてしまうんですね。そういうことが実際、論文の中でも行われているのが事実なんですね。

でも、一方、この方法ですと、客観的に交差確認という方法で、チューニングのパラメータを決められますので、誰がやっても同じ結果が出せるというのが、メリットがあると思っています。

そういうこともあって、この方法、そんなに新しい方法ではないんですが、いろいろな企業の方に使っていただいています。

今のが1つ目の、異常検知の話でした。

今度は、普通のいわゆるパターン認識の問題です。

今、画像認識とか音声認識、あるいは言語の機械翻訳といった分野で、人間と同等か、場合によってはそれ以上の性能が出ることが報告されていますが、これらの応用分野というのは全て、ある意味、ネット系のものでして、ネットからどんどんデータがとれるので、ビッグデータが使えるという分野なんですね。例えば、我々理研の中でも、医療データの解析とか、橋とかトンネルなどのインフラの管理であったり、あるいは防災の話、自然災害の話であったり、あるいは材料開発といった分野の応用もしているんですが、こういうところはデータをとるところに人が入ってしまいますので、大量にデータをとれないわけなんですね。ですので、何とか限られた情報からうまく学習するというのが永遠の課題で、非常に重要だと思っています。

この部分をほんとうに基礎から、ずっと取り組んでいるんですが、今、考えている問題は、一番簡単な2クラスの教師付き分類と呼ばれる問題でして、青の点と赤の点が与えられたときに、それを分ける緑の線を求めましょうという、これだけの問題なんですね。雨が降るかどうかとか、地震が起こるかどうかとか、そういうのを分類しているイメージです。

これも昔から研究されている問題で、赤とか青とか色がついている点のこと

を、ラベル付きのデータと呼びますが、大量のラベル付きデータを用いれば、分離境界が精度よく学習できるということが既に知られていまして、もうちょっと正確には、ラベル付きのデータの数を n としたときには、緑の線の推定誤差がルート n 分の 1 の速さで減っていく、こういう感じで減っていくということが知られていますので、これでビッグデータを取りましようという理論的な根拠になっているわけなんですね。

ビッグデータがとれる課題では、実際にこれでうまくいくようになっているわけなんですけど、残念ながら、応用例によってはなかなかビッグデータはとれないこともございますので、そういう場面で、逆の極限として、全くラベルをとらない、教師なし分類というのを仮想的に考えてみたいと思います。

今、データ点はいっぱいあるんですが、まだ、青か赤かわからない、黒の状態になっています。この状態で青と赤を分けたいという問題を解こうとしているんですが、これは原理的にはできないわけなんですね、全く色がありませんので。ですので、実際には、いわゆるクラスタリングというのをしまして、データを固まりごとに分けてやって、そうすると、結果的にここで分けることになるわけなんですけど、たまたま左上の 1 つのクラスターが全部赤とか、右下のクラスターが全部青という状況になれば、これですごくうまくいくわけなんですけど、実際の問題はなかなかそんなきれいにはなっていませんので、これでクラスタリングをしても、赤と青に分けることにはうまくつながらないわけなんですね。

そこで折衷案として、半教師付き分類という研究が、過去 20 年ぐらい、非常に盛んに行われてきました。これはほんのちょっとの赤と青のラベル付きデータと、あと、大量のラベルなしデータ。ラベルなしデータはほとんどコストなく集めてこられるという前提なんですね。それを使って、うまく青と赤を分けようという話なんですけど、さっきのページと同じ絵を使っているんですが、結局、教師なし分類と同じやり方をすることになるんですね。

基本的には、やっぱりクラスターごとに分けて、今回はこのクラスターのここが青だとわかっていますので、この青を、何となくこの辺に伝播してやると、

このクラスターは全部青だと。今、この点が赤だとわかっていますので、この情報をこのクラスターに伝播してやると、このクラスターが全部赤だというふうに色づけるだろうということなのですが、これも結局、背後の仮定として、1つのクラスターが、1つのクラス、1つの色に対応しているという仮定が成り立っていない限り、うまくいかないわけなんですね。

ですので、随分はやったんですが、私もいろいろな企業の技術顧問みたいなものをさせていただいて、いろいろな会社が、やっぱりラベル付きデータを取りたくないの、こういう方法を使いたいということで、頑張っているいろいろやってきたんですが、残念ながら、ほとんどうまくいかなかったというのが正直なところなんですね。

ここまでの状況をまとめますと、いろいろな分類の方法がありまして、学習の精度の高い・低いを左右にとりまして、データにラベルをつけるコストの高い・低いを上下にとりまして、一番最初の青と赤の点を使って分ける教師付き学習は、非常に精度はいいんですが、ラベル付けのコストが高いということで、右上に来ます。一方、2番目の全部黒の点を使う教師なし学習は、ラベル付けのコストは事実上ゼロに近い状況なんですが、精度は出ないということで、左下に来ることになります。最後の半教師付き学習が、ほんとはもうちょっと精度がよくあってほしかったんですが、残念ながら、あまり精度が出ないということで、やっぱり左側に来てしまいます。

ということで、結果的に右下のあいている部分、精度がよくてコストが低いという方法が、今のところないという状況ですので、この部分を埋めるような新しい方法をつくらなければいけないんじゃないかというのが、最近の我々のグループの研究の意識でして、そういう方法をいろいろつくってきました。

そのイメージをざっと紹介させていただければと思うんですが、例えば方法1番は、教師なし分類を考えていますので、ラベルを全くとらない方法なんですが、ちょっと特殊な設定を考えていまして、ちゃんとそれでも学習できる場合があるということなんですね。

これはちょっと不思議な方法なんですが、今、ラベルなしの黒いデータセッ

トが2つあると仮定します。その2つは全く同じではなくて、背後に、この中に青と赤がまざっているんですね。ここも青と赤がまざっているんですが、その比率が違うということを仮定します。これは、気象の文脈でそれができるかわからないんですが、例えば病院のデータ解析をするときは、病院Aから集めてきたデータと病院Bから集めてきたデータを2セット用意して、それぞれの病院で、例えば、がん患者の比率が違いますよということが保証できればいいわけなんですね。何とかうまく、ラベルなしで比率が違う2セットのデータを持ってくることができれば、実はそれだけから、ちゃんと青と赤を分ける線を学習できますということを理論的に証明しました。

これはちょっと不思議な方法なんですけど、一つもラベルをとらなくても、ちゃんと青と赤を分けられるんですね、この場合は。しかも、もうちょっと理論的にちゃんと調べますと、さっきのルートn分の1の速さで誤差が減るのが、この文脈でも実はちゃんと言えますので、ラベルを一つもとらないんですが、ラベルを全部とったときと同じぐらいの精度が理論上は出るという、ちょっと特殊な方法なんですね。こういう方法を1つつくりました。教師なしでもちゃんと解ける場合があるという問題です。

2番目、これが最近いろいろ研究業界でもブームになってきているんですが、青と赤を分けたいんですが、赤のデータが全くとれないという状況を考えます。

例えばインターネットの会社が、あるユーザーがこのリンクをクリックするかどうかを予測するという、クリック予測という問題を解いているんですが、これは広告業界で非常に重要な問題なんですけど、そういう分類をするときに、ユーザーがどんなリンクをクリックしたかというクリックログのデータを使って学習するんですが、クリックしたリンクというのはすぐとれますので、それは正のデータなんですね。そのユーザーが好きなリンクということで、正のデータは簡単にたくさんとれます。青のデータですね。

一方、負のデータ、そのユーザーが嫌いなリンクのデータというのは、実は簡単に集められません。クリックしなかったリンクというのは簡単に後からとってこられるんですが、クリックしなかったから、嫌いかどうかというのはわ

からなくて、ほんとうは好きだけれども、そのときたまたま忙しくてクリックしなかったかもしれないという可能性がありますので、クリックしなかったリンクというのは、赤ではなくて、まだ青か赤かわからない、黒のデータなんですね。

そういう状況ですと、青と赤を分けたいんですが、青と黒しかとれないということになりまして、青と黒だけから青と赤を分ける線がこの辺にあるんですが、これを求めたいという問題になります。

これはほかに、SNSで友達かどうかを予測するとか、そういう問題にも適用できるんですが、これに対して、結論としては、赤のデータが全くなくても、ちゃんと学習できますということを証明しました。ですので、青と黒だけあれば、もう完全に学習できます。赤は一つもなくとも大丈夫ですということを示しました。

これを使うと、先ほどの半教師付き分類もうまく改良することができまして、半教師付き分類というのは結局、黒のデータがたくさんあって、青と赤がちょっとだけあるという設定なんですけど、さっきのページで、青と黒だけあればちゃんと学習できますということを既に証明しているわけですので、それに赤が加わっても、特に損はしないわけなんです。

ですので、もうこの時点でできましたと言えるんですが、せっかく赤もありますので、赤もさらに活用しようということで、いろいろ考えた結果、青と黒を分けるという先ほどのページの学習法と、一番最初のところで出てきた、青と赤を分ける普通の教師付き学習の方法、2つを組み合わせると非常に精度が出るということがわかりまして、それを理論的にちゃんと調べると、ちょっと余計なスライドを1枚持ってきちゃったんですが、機械学習の人は、将来出てくるデータに対する予測がどれくらい当たるかというのを理論的に評価しようとしているんですが、この不等式の左辺が将来の予測の誤差なんですけど、誤差が最悪どれくらいかというのを上から押さえるんですね。

いろいろごちゃごちゃあるんですが、ここにルート n_0 分の1という項がありまして、これがすごく重要な項なんですけど、 n_0 というのが、先ほどのラベ

ルなしの点の数ですね。まだ色がついていない点の数なんですけど、誤差がルート n_0 分の 1 でちゃんと減っていきまうということが保証されますので、半教師付き学習において、ラベルのついていないデータをたくさんとれば、ちゃんと誤差が減りますということが保証できる初めての方法なんです。

今までは、クラスターに沿ってラベルを伝播しようみたいな方法ばかりずっと研究されてきたんですけど、今回、全く違ったアプローチで、ラベルなしデータを活用することによって、ルート n_0 分の 1 という項が初めて出てきた方法なんです。ですので、こういうものを使いますと、ラベルなしデータをほんとうに有効に活用できるかもしれないということで、人手をかけずにデータを集めても、うまく学習できる可能性があるということなんです。

ほかにも、先ほど、ラベルなしデータはとれるという話だったんですけど、問題によっては、ラベルなしすらとれないこともあります。

例えばこれは、企業のデータ解析みたいなイメージなんですけど、自社のデータに対しては当然アクセスできますが、他社のデータベースに関しては、ラベルなしですら、プライバシーの問題があつて、アクセスできないだろう。自社と他社を分類するような分類器をつくるときに、結局、自分の会社のデータしかとれないという状況で、何とか学習できないかというのが問題意識ですね。

ほかには、サイエンスの研究とかで、例えば、『nature』とかああいうジャーナルに論文が出てくるわけですが、そういうところに載る論文というのは、実験がうまくいったものしか、当然アクセプトされなくて、実験を失敗したものというのは表に出てこないわけですね。こういうのをパブリケーションバイアスと呼んだりしますが、そういう中で、成功した結果しか集まらない中で、成功と失敗を分けるような分類器をつくれなかつたかということでも、原理的には使えるわけなんです。

結論としては、そのままではできませんということなんですけど、正のデータで、どれくらい正かという信頼度がわかれば学習できますということを示しました。この点は 95% 正ですよ、この点は 5% くらいしか正じゃないですよというような信頼度がわかれば、ちゃんと、正のデータしか見ないんですけど、青

と赤を分けることができるようになるということを示しました。

最後、もう一個おまけなんですけど、今まで青と赤の2種類だったんですが、今度はもっと多クラスのもので、3つ以上の場合を考えます。実際には1,000とか1万という数を考えているイメージなんですけど、例えば画像に何が映っているかというのを予測するとき、空であったり、人であったり、犬だったり、猫だったりとリストアップすると、何千何万となるわけですね。何千何万のクラスに分類したいという問題を考えているんですけど、データをつくるのが非常に大変なんです。1万個ラベルがあったとすると、ラベル付けするときに、1万個のリストを順番に見て行って、正しいものを選ばないといけないわけなんです。そうすると、何千個を見て、2,3,5,7番目のこれが正しかったというのを入力するというのを何回も繰り返すわけですので、100万データとか、とてもとれないわけなんです。

そこで我々が考えたのは、正しいデータは諦めるので、間違っただラベルを1つ下さい、そうすると間違っただラベルからでも学習できますよということを示しました。間違っただラベルを「補ラベル」と呼んでいるんですけど、集めるのは簡単なんです。1万個ラベルがあったら、正解は1個しかないですので、基本的に、適当に選べば、ほぼ確実に間違っているわけなんです。適当に1個選んで、これは間違っていますかと聞いて、間違っていますと言ってもらえれば、それで使えるわけなんです。たまたま運よく、合っていますと言ってもらえれば、それももちろん使えるわけなんですけど、間違っているものもちゃんと使えるということで、正しいデータだけじゃなくて、正しくないデータからでもさらに情報がとれるという新しい方法でして、こういうものを使うと、データのラベル付けのコストがさらに下げられるわけなんです。

ということで、いろいろ細かい話をしましたが、右下の部分ですね。精度が良いままラベル付けのコストを下げるという部分にこだわって研究をされていて、これに関しては、一般論ができるわけではなくて、問題に合わせてつくっていく必要がありますので、いろいろな問題で起こり得るような状況を考えて、それに対して、ちゃんと理論的に保証があるような学習法をつくっていこ

うということ、いろいろな方法ができていまして、こういう話を今、実はいろいろな企業さんに裏で実験していただいています、そろそろ成果が具体的に出てきそうな状況になってきました。

最後は、AIの人材に関する話をちょっとだけしたいと思います、今、AIの研究者・技術者が世界的に不足しているという状況で、日本では、この数年で特にAIのブームが来たということで、AI技術者の雇用が非常に活発化されていますが、我々の機械学習の学会を見ても、思えば2000年ごろから、当時できたばかりのグーグルとか、もうかなり大きかったマイクロソフトの研究所であったり、あるいは、当時元気がよかったヤフーの研究所とかそういうところが、こっちの業界から人をたくさん雇っていたんですね。

それが2000年ごろですので、十七、八年前とかそんなころなんですが、そういう会社はそういう人材をどんどん集め続けていますので、十何年かけて、ものすごく人が育っているわけなんですね。一方、日本はここ数年ぐらいでやっと、急にAI人材が必要だという状況になっていますので、人が全然いないというのは、明らかに日本とアメリカで差があるわけなんですね。

例えばアメリカの大学で今、コンピューターサイエンスの分野を見ますと、客観的なデータがなかなかそろわなかったんですが、これはMITの博士課程の学生の数というので、過去10年ぐらいを見ると、倍近く増えていますということなんですが、実態としてはこんなレベルではなくて、我々の周りのアメリカの主要大学のコンピューターサイエンスの先生とかに聞くと、大学院に2,000人とか3,000人という規模で応募が来て、合格するのは数十人と言っていますので、100倍ぐらいの倍率なんですね。

一方、我々がいる大学院とかで見ると、3人応募が来て、2人採るか、3人採るか、どっちにしましょうかという議論を会議でやっている、そんなレベルなんですよ。ですので、人材の数も100倍違うという状況ですので、なかなかそのままでは厳しいなというのを感じている状況です。

そうしてしまうと、人工知能の技術が理解できる人と理解できない人の差がどんどん開いていってしまうわけなんですね。そこをやっぱり何とかしたいと

思っています、1回目か2回目の会議のときに、ちょっと話をさせていただいたんですが、今の日本の大学の学部学生の割合というのをどこから調べてみますと、いわゆる文系と呼ばれている分野が半分ぐらいありまして、理学、工学が合わせて18%ぐらい、農学が2.9%、保健とか医歯薬、教育とかこういうものが残り10%、10%ぐらいある感じなんです。

この中で、どこをどう強化していけばいいかというのは、非常に難しいところなんです、まずは、コアの部分であるAIの技術を高めていくような最先端の研究をする人というのも当然、一定数必要ですので、それは理学、工学のあたりの一部に対応するかなという気はします、この辺の人たちに対しては、AI技術の発展のために、プログラミングも当然できないといけません、それだけじゃなくて、AIの技術、数学が必要になりますので、数学をやるという、この両輪でしっかり教育していく必要があるだろうと思っています。

あと、残りの2割の部分が、「その他理系」という書き方がいいかわからないんですが、今、理系のほとんどの分野で、AI技術を使わないとそもそも研究が進まないという状況にもなってきましたので、彼らも当然、AIを活用する必要性が出てきているわけなんです。そういう分野の人にも、AIを使ってデータ解析に対する素養を高めていただこうと考えています。

あとは、最後、いわゆる文系で、政治・経済・ビジネス・社会系の人たちが、やっぱり技術とは無縁で生きていけない時代になってきていると思いますので、AIと聞いただけで、魔法の箱があるように考えてしまったり、あるいは、数式を見ただけで、数式は全然だめだからとアレルギー反応を起こしてしまうというのは非常にまずいなと思っています、実際、数式といっても、大したことを書いていない数式も多いわけなんです。

例えば、数式でも全然、中学生でも理解できるようなことを書いてある場合もあるんですが、数学と言っただけで、嫌ですと言われてしまうことも結構ありますので、そういう部分は、ちょっとアレルギー反応みたいなものをうまく抑えていかないといけないなと思っています、技術の中身をそんなに細かく知る必要はないわけなんです、最低限の技術を活用するための技術リテラシ

ーというのは当然、持っていただかないといけないなと思っていました。

日本はやっぱり文系、理系というのが分かれ過ぎなんですよ。中学生のころに完全に分かれてしまって、一般論として、私立の文系の大学に入る人というのは、15歳で数学を終わっているんですね。高校生ではもう一切、数式はさわらないという生活をしていますので、それはちょっとまずいかなという気はしているんですね。ですので、大学入試で数学をやれば、問題が解決するのかわからないですが、せめて18歳ぐらいまで、ほんとうに基礎の数学で結構ですので、みんながやることになれば、それだけでも随分違うんじゃないかなという気がするんですね。

理系と文系というふうに見たときに、理系の人が大人になって、政治の話をちょっと勉強したり、経済の話を勉強したりということは、ごく普通にあるかと思うんですが、あるいは、企業にずっといる方とかは、ビジネスとかマネジメントのことも勉強される方は多いと思いますが、逆のケース、文系の人が大人になって数学を勉強するというのは、ほぼ聞いたことがない状況ですので、そう考えますと、やっぱり数学みたいな地味な分野というのは、若いうちに無理してでもやっておく必要があるのかなという気がしますので、それは大学ではなくて、もしかして中学、高校あたりの話になるかもしれないんですが、そういうところで、ほんのちょっとでもいいので、みんなが最低限の技術にかかわるような数学の基礎の部分を学ぶことができれば、その先がずっと変わってきますので、随分よくなるんじゃないかなという気がするんですね。

最後にちょっと我田引水的になるんですが、こういうものをやるためには、受け皿となる情報科学系の教員と学生が必要になるわけなんですよ。今、我々の研究室も、AIブームのおかげもあって、すごくたくさん応募が来ているんですが、何十人も来ても、採れる学生はほんの数人しかいないという状況なんですよ。世の中のニーズとしては、AI人材をもっと輩出しなさいと言われてるのに、大学では受けられないというすごく矛盾した状況で、我々も結構悩んでいるところなんですよ。そのためには、長期的に見て、今ブームだから増やせばいいというのは、あまりいい解決策ではないのかもしれませんが、でも、

ニーズがあるのは事実ですし、大学とかも定員をもっとフレキシブルに変えていくような制度が必要なのかなと感じています。

最後は大分脱線しましたが、以上で終わりたいと思います。ありがとうございました。

【委員】 どうもありがとうございました。こういうお話を聞かせていただく機会はありませんので、非常に貴重な勉強をさせていただいたと思います。

それでは、委員からいただいたお話に関して、ご質問等ございましたら、委員の方からお願いできればと思います。どなたかございますでしょうか。

では、委員、お願いします。

【委員】 今日はいろいろありがとうございました。

今日の先生のお話、比較的技術の、また、AIのテクノロジーの話が多かったんですけども、この手の問題、この技術にいかに関係の問題をマップするかということだと思います。今日のお話、要は、空間分割といった内容だと思いますが、先生からごらんになられて、気象のこういった問題の中で、こういうAIのテクノロジーに帰結できて、非常に有効に働くことが予想される分野がありますでしょうか？気象はどっちかということ、過去と現在のデータから未来を推測するという、時系列っぽいところもある中で、一番帰結しやすい、適用するとすごく効果が出る領域が、先生の目から見て何かあれば、ご教示いただけますでしょうか？

【委員】 ありがとうございます。そこをまさに今、ほんとうは皆さんとご相談させていただければいいなというところであるんですが、我々は我々で基礎のほうから技術をつくっていて、現実世界ではどういう問題があってというので、そこにギャップは当然ありますよね。

ですので、そこでうまく解ける問題を見つけられるかどうかというのが非常に肝かと思っていて、理研のセンターでも、いろいろな研究所とか企業の方と共同研究を始めようとしているんですが、全部そこが問題なんですね。こういう技術があります、こういう問題があります、その中で幾つマッチングがとれるかというのがほとんど全ての状況でして、私も気象のことは、一般ユー

ザーのレベルでしかわかっていませんので、具体的にこの問題というのは、なかなか申し上げるのは難しいところですが、実は今の機械学習の問題、手法は、お話にあった時系列の解析にはちょっと向いていないところはあるんですね。データ点が全部、統計的に独立だという仮定を置いて理論ができていますので、時系列のように依存性があるものというのは、原理的には直接は使えないんですね。

ただ、それを近似的に使うような方法はいっぱいありますので、そういうものを使って、今までいろいろな問題を解いてきているんですが、そういう中で、多分、気象のデータで大きく問題になりそうなのは、欠損値が多いとか、人手がかかるところは、やっぱりデータの数がとれないとか、あと、異常値が多いのもありますよね。

そういうものに対するアルゴリズムとか理論というのは、ここ10年ぐらいでものすごく進んでいますので、そういうものは、今まで既に統計解析の手法が使われている問題に対して、より最先端のほうを適用していくというだけで、精度はかなり改善できる可能性があるのではないかなと思っているんですね。

ですので、既存のものを改善していくという研究と、まだできていないものに対して、新しい方法をつくって適用していくという、当然両方やっていかないといけないと思うんですが、後者はかなりハードルが高いと思いますので、まずは最初、その分野同士で仲よくなるという意味も込めて、前者のほうから、既にある程度問題が定式化されているところに新しい方法をどんどん適用していくというところからスタートできると、将来性はあるかなと感じています。

【委員】　もう1点、よろしいですか。これはちょっと気象と話がずれちゃうようですが、先生とお話しできる機会もあまりないので。今日お伺いしたいのが、一番最後の人材ですけど、先生が、アメリカと比べて日本は圧倒的に少ない。現実、産業界の中ではそうだと思います。若い人のレベルで考えたときに、違う話もあって、例えば先日、我々の学会のACMで、プログラミングコンテストの対抗戦を大学別でやると、東京大学は4位に入っております。その上は、アメリカではなくてモスクワでした。次が大体、北京大とか清

華大が来て、次に日本が来て、欧米は意外と順位が低いようです。

それはこの大会だけなのかなと思うと、若い子たちがよく参加するコンテストがあります。Top Coder とか CodeForces とか、あの手を見ても、日本は、国別で言うとロシア、中国と一緒に、大体 3 番目ぐらいには来ています。どのコンテストを見てもほぼそんな感じですが。グーグルの人工知能コンテストになってくると、やっぱりアメリカも強いんですけども、その辺のギャップとこれとを考えると、若い子の育成とかじゃなくて、最終的にその後に、就職した後の人の流動性において、アメリカに人が集中しているのではないかと思います。なので、日本にも結構、プログラミングできる子たちはいるけれども、最終的に日本の会社に入っちゃうと、例えば役所の行政システムの更新みたいな仕事をさせられたりとかで、活かされていない気がします。そういう社会的な構造の問題なのかなという気もしますが、AI は私とはちょっと分野が違うので、

先生から見ると、その辺はいかがでしょうか。

【委員】 日本人が潜在的に優秀だというのは多分、世界的に言うと、それはおっしゃるとおり、間違っていないと私も感じているところなんですけど、今、例えば我々のコンピューターサイエンスの周りを見ますと、例えば学部生の 9 割以上は修士に進学しますが、修士から博士に行く人は多分、1 割ぐらいしかいない状況なんですね。一方、アメリカは、先ほど見たように、大学院に何千人という単位で応募が来る世界ですので、全然そこは差がありまして、そう見たときに、ちょっと逆説的な言い方をすると、日本は修士のレベルが多分、世界トップクラスなんですね。修士の人でも論文を書いたことがあるとかという状況ですので、ミニ博士みたいな学位をとって、みんな企業に行ってしまうんですね。

一方、アメリカの修士課程というと、ただ授業を受けるだけですので、研究をやったことがないような人たちばかりですので、いまいちなんですね、レベルとしては。なので、そうすると、ドクターに行って、研究を初めてやろうという人がたくさんいるのも自然なのかなという気がしますので、この状況を改善するには、日本の修士のレベルを下げればいいんじゃないかと逆に思ったり

もするんですね。研究はドクターに行かないとやらせませんというふうに変えれば、みんなドクターに行くかもしれないという気もしますので、優秀な人がちょっと研究をして満足して、企業に行ってしまうということで、高度な研究に携わるきっかけを失っているのが一つの問題かなと感じていますね。

【委員】 ありがとうございます。

【委員】 先生、お願いします。

【委員】 いいですか。どうもありがとうございます。

途中から伺ったので、違うようなことを言うかもしれないけど、今のお二人の話について言うと、日本の理系の研究者養成という意味で言うと、やっぱり圧倒的に、ドクターも取得するというのが、企業社会の中で、産業社会の中で、非常にネガティブですよね。アメリカはそっちが強いから、修士がちょっとぐらい勉強ばかりして、レクチャリングだけでも伸びていくじゃないですか。そこじゃないかと思いますね。

一個伺おうと思ったのは、気象庁所管のある種の予報、予測、業務的に言うと、いわゆる大気現象の世界と、それから、この間の土砂崩れがあったみたいに、非常にアンノウンファクターが、つまり、地面の下にアンノウンファクターがいっぱいあり過ぎて、大気現象ほどにはファクターをはかれない。だから、どこで何が起こるかわからない。雨が多くなれば、きっと起こりそうなんだけど、起こらないことも多いし、たまたまあその耶馬溪で起こった。隣だってほぼ同じ状況なんだけど、起こらないという、ランダム性が非常に多い世界と、ちょっと違うじゃないですか。

そうすると、伺いたいのは、人工知能的な技術というのが、どっちに有効なのかということをもまず知りたいんですね。少なくとも大気現象のほうについては、人工知能的なことを使わないでも、相当な進展がしているじゃないですか。そうすると、素人的に言うと一番わけわからないでいる、土砂崩れであるとか、噴火であるとか、噴火はわけがわからないと言っては失礼か。すみません。そういうところへ応用というのはどんな可能性があるのか、伺いたいなと思うんですけれども。

【委員】 ありがとうございます。非常に難しいお話ではあるんですが、基本的に、今の人工知能の技術は全部、統計学に基づいていますので、過去に起こったパターンに基づいて、何かを予測するだけなんですよね。ですので、全く起こらないことを予測するというのは、やっぱり原理的に無理ですよね。

アンノウンファクターが多いのは、多かれ少なかれ、全ての問題で、原理としては多分、同じだと思いますので、あまりにもアンノウンのものが多過ぎるときには、例えばシミュレーションとかでそこを補完して、補ってやるとかすることによって、ある程度、精度が改善されることもありますし、もっと違うところからデータを持ってくるようなことも、今はやったりしますので、そういう工夫をちょっとずつ積み重ねていけば、精度はちょっとずつ改善していくかなという気はしますので、そういう意味では、突発的に起こるようなものを当てたいというのは、なかなか普通の意味では難しいなというのが正直なところですね。

今日、前半の部分で異常検知の話をしたんですが、異常検知もある意味、正常か異常かを分ける分類問題だと捉えることも当然できるわけなんですけど、あまりにも異常があらわれないので、あるいはバリエーションが多過ぎるので、分類問題として解かずに、わざわざ異常を検知するという別の問題で解こうとしているんですね。

ですので、数学的にちょっと違うアプローチをとることによって、異常を見つけやすくするという工夫をしていますので、そういう感じのアプローチはそういう面でできてくるのかなという気はしますね。

【委員】 何で今のところを伺ったかということ、耶馬溪についても、世間でコメントされるのは、要するに、地表部分でのいろいろな計測をたくさんやっていけば、きっとわかるようになるんじゃないかみたいな言い方をする人もいて、実際にそういうものを取りつけようという動きもあるんだけど、そんなものじゃないぐらい、地面の下の地盤レベルでの、岩盤レベルでの局所的な様相なんていうのは、表層だけはかってもだめじゃないですか。

言いたいのは何かということ、今、世間の中では、AIが登場すれば何でもで

きるんじゃないかみたいな、迷信みたいな世界があるでしょう。あれに対して、やっぱり先生のようなご専門の方が、ある種の警鐘と、ある種のできるところとそうでもないところというのは、はっきりおっしゃっていただくことが、無駄な誘導がなく、大事なところに注力できるんじゃないかと思って申し上げました。

どうもありがとうございました。

【委員】 最初の 委員からの質問に関係して、私も十分理解していませんが、現在のところ集中豪雨などは、数値予報モデルを使ってもなかなか予測は難しい面がありますが、例えば、過去に、環境場の温度、湿度、風などの色々なパラメータがこういう条件のときに、集中豪雨が起きたみたいなことはわかっている。そうしたときに、将来を予測したデータから、過去の発生条件に基づいて、この場合には起きそうだ、この場合には起きないという線を引きような使い方には利用できるものなんではないでしょうか。

【委員】 一応、数学的にはできますと言うべきではありますが、そこは、データが十分とれるかどうかというのは当然、常に問題になってきますので、今日の後半のお話は、ある意味でそういうところに、もしかしたら使える可能性があるという気はしてはしまして、ほんとうに起こった場合のデータというのはほとんどとれないと思いますので、そういうのがなくても、何かはたくさん集められる可能性がありますので、そのたくさんを使って、うまく学習できないかというのが後半の研究の趣旨であったんですね。

基本的に、全部統計学に基づいていますので、ルート n 分の 1 というのが出てきたのが全ての場所できいていますので、 n が小さいときには、誤差は絶対大きいんですね、何も知識がない場合は。ですので、何かを使って n を大きくしないといけないというところで、ほんとうにちゃんとラベルのついた、答えのついた正しいデータをたくさん集められれば、もちろん一番簡単なわけなんですけど、それが現実的ではない場合には、何か、そうじゃないけれども、たくさん集められるものを持ってきて、その n に対して、ルート n 分の 1 の誤差になれば、すごく精度がよくなるわけなんですね。

そういうところを、我々としては工夫しようとしていまして、そういう方法の一つが近い将来、どこかに使えると非常にありがたいなと思っているところですので、ぜひ、またいずれ議論させていただければと思います。

【委員】 ありがとうございます。

ほかにご質問はございますか。

先生、お願いします。

【委員】 今日、素人にも直感的にわかりやすい、易しいスライドをたくさん用意いただいて、ありがとうございました。

2つありまして、1つは、せっかくテクノロジーの話をお願いするのに、ちょっと理解できていないのは、20ページの新手法2で、3、4、5というか、後ろもおおむね部分的な、ある種欠損しているけど、一部のデータがあるだけでも判別機能、精度を高められるというのは、直感的にはわかるんですね。

20ページの正例だけがある場合が、「教師付き学習と同じ収束率を達成可能」という意味は、正例、負例、両方そろっている学習と同じぐらいの精度が得られるというふうにも見えるので、そういう意味なんでしょうか。そうすると、すごくそれは興味深いというか、その理屈がもう少しだけ知りたいのが1点目です。

もう1点は、先ほどの議論にも関わるんですけども、確かに系列相関なんかある場合には、機械学習的がやりにくいというのはあるんでしょうけど、でも、その部分も将来的には、技術的に進歩していくというか、開発されていくという理解をしておいていいのか。あるいは、今の気象だと、アンサンブルをとったりして、入力値をいっぱい変えて、それも計算機の速度が上がっているから、ものすごい数の入力値に対してのものすごい数の出力値を出しますよね。

この出力値レベルで言うと、ある意味で、系列相関というのは集約されて、関数によって、シミュレーションによって、そういうデータを今度は機械学習にかけて、ある種、セカンドオピニオンじゃないんだけど、違う判断材料にしていくみたいな、こういうのは今すぐにでもできそうな感じなんですけど。活用方法としては、例えば短時間予測の数時間とか、半日だとか、そういう中で

活用していくという可能性、このあたりはどうお考えかとお聞きしたいと思います。

以上です。

【委員】 ありがとうございます。まず、前半の正例とラベルの件なんですが、結論としては、ラベルを全部つけたときと同じぐらいの精度で出せますということを示してしまっていて、実は場合によっては、こっちのほうがよくなることもあるんですね。負例が少ないときよりも、ラベルなしがたくさんあるときのほうが、実は精度がよくなったりしますので、そういう意味では、赤のデータは全くなくても、黒がたくさんあるんだったら、それで十分精度が出ますということを書いてしまっていて、この直感的なイメージとしましては、黒のデータは赤と青がまざっていますので、黒を赤だと思って、青と黒を分けているイメージなんですね。

でも、そのままやると、黒のデータの中に、赤だけじゃなくて、青もまざっていますので、ちょっとバイアスが乗ってしまうので、そのバイアスをうまく補正しているんですね。

ですので、ラベルなしデータをたくさん集めれば、比率は一定だとすると、赤のデータも原理的にはたくさん増えていくことになりますので、それで精度が上がるといようなイメージなんですね。青はノイズはないですが、黒はノイズがたくさん乗っていて、赤のものが青にひっくり返っちゃっているというイメージで、問題を解いていることになっています。ですので、すごく精度は出るということですね。

後半の系列相関の話なんですが、これも、機械学習の面も当然、何とかそういうものを扱おうということで、いろいろ研究はされているんですが、うまく何かが独立になるように前処理をしてから学習しようというのが、時系列モデルの研究とかは大体そんな感じで、残差だけは独立性を仮定してもそんなに問題ないだろうということで、残差に関して、iid(独立に同一な分布に従う)としてやっていくというような話もありますし、あと、もうちょっと本気で、背後にある動的なシステムをちゃんとモデル化して学習しようという話も、今、

我々の理研のセンターとかでもやっています、そっちのアプローチのほうが、ある意味、本格派ですので、自然科学のデータ解析にはもしかしたら向いているかもしれないんですが、現状ではちょっと次元が高いと難しいというのもありまして、そこはほんとうに徐々に時間をかけて改善していかないといけない分野かなという気はするんですね。

ですので、今、どのアプローチが何に対して有効かというのは、まだ結論は出ていないところだと思うんですが、いろいろな新しいアイデアも出てきているところですので、そういうものをちょっとずつ試しながら、どれかが何かの課題の精度向上にちょっとでも役立てばということで、あらゆることをやっているという状況ですので、ちゃんとしたお答えを持ち合わせていない状況であります、どれかの技術が使えるようになっていくんじゃないかと期待して研究をしています。

【委員】 ありがとうございます。

【委員】 委員、お願いします。

【委員】 27ページの人材育成についてお伺いします。

文系でもビッグデータが扱えるような教育をしようと、幾つかの大学で動いています。気象データを活用するというときには、ここの人材育成も非常に期待したいところだと思います。

実際、理系ですとかなり人気が上がってというのはわかるんですが、文系の学生さん、受験生としては、数学が得意ではないけれども、こういったところで活躍したいと思って出てこられるというのはどんなぐあいか、ご存じでしょうか。

【委員】 受験生はわかりませんが、例えば、私は今、東大にいて、1年生は文系の学生もいますので、たまに話す機会はあったりするんですが、潜在的に興味のある学生は結構いるかなと感じているんですね。そのときに、技術のほんとうに細かい中身をわかってもらおうとも我々は思っていないところはあるんですが、それがどういうふうに使われていて、どういう原理で動いているのかというのを、式を使わずに理解することはできますよね。その部分ぐら

いまで興味を持ってもらえれば、ある意味、十分なのかなと思っているんですね。

さっきの魔法のAIの箱があっみたいイメージで思っている人が、わりと企業の経営者とかには多くて、我々もいろいろな企業の方の訪問を受けているんですが、ほんとうに、我々のビジネスの悩みはこの魔法のAIが解いてくれると思って来られる方もいらっしゃるんですが、そういう人がほんのちょっとだけでも勉強してきてくれれば、それは全く意味がない議論をしているんだというのがわかるかと思うんですね。

それは文系の学生さんでも全然できる話で、やっている人も既に結構いると思いますし。

【委員】 そうすると、そういったところで人が育ってくると、何でもできちゃうという過剰な期待はなくて、むしろ過剰な期待をしがちな一般の方と理工系の方の間をつないで、どこができて、どこが必要で、どういう基本原理で、どういう社会をつくっていくかといったところを固める人材になっていく……。

【委員】 まさにおっしゃるとおりですね。

大分前になるんですが、私、一橋大学の商学部で講義をしばらくしていたことがあったんですが、10年ぐらい前から5年ぐらい前までやっていたんですが、そのときに文系の学生の前で、人工知能の話とかを当時からしていたんですが、理系の学生にそういう話をすると、おもしろいとか、将来期待できるとかという感想が来るんですが、文系の学生の多くは、怖いという感想が来るんですね。それはやっぱり中身を全く知らないので、できている表面だけを見ると、いろいろな不思議なことができてしまうので、何か怖いと思われてしまうんですね。

そこがちょっと危険なことかなと思っていまして、中身を見ればそんな大したことをやっていないわけですので、それが人間を襲うことなんてあり得ないというのは一目瞭然なんですけど、今、世の中、シンギュラリティーで、人間がコンピューターに支配されるぞみたいなことがほんとうに議論されているのが、研究者から見るとばかげている話なので、そんな起こり得るわけないことを、

何でこんなみんな真剣に議論しているんだろうと。

最初はそれを私も無視していたんですが、みんながあまりにも言うので、あまり無視していると、こっちが逆に悪いと思われてしまうので、そうじゃないんですというのをちゃんと言うようにしているんですが、そのギャップが、技術をやっているほうから見ると理解できないくらい、ギャップがあるんですよ。

ですので、そこをほんのちょっと知っていただくだけで変わると思いますし、我々ももちろん、そういう努力はこれからしていかないといけないかと思うんですが、もうちょっと、それこそ高校生とかのレベルでちゃんと技術のリテラシーを学んでいただければ、問題なくなるのかなという気はしますね。

【委員】 ありがとうございます。

【委員】 非常に新しいお話を伺ったので、議論が盛り上がりましたが、本題のほうもございますので、とりあえずここで次の話に進めさせていただきたいと思います。

委員、どうもありがとうございました。

それでは、「気象業務に関わる技術開発」について、事務局のほうで議論の材料を用意していただいていますので、こちらをご説明いただいて、議論に移ってまいりたいと思います。

気象庁より資料の説明をお願いいたします。

【企画課長】 では、企画課の森から、また説明させていただきます。

資料4で、めくっていただくと、まず、「技術開発に関するこれまでのご意見」がございます。これは先ほど申し上げたものとほとんど重なっています。

重なっていないようなものだけ申し上げますと、観測・予測精度の向上のところ、気象庁の役目として、予測精度の向上が一番重要であるといったところが1点、その後、目標を設けるべきとか、気象と地震・火山では精度が違ったりするのでというところで、この辺は先ほどと同じです。

最先端技術の活用のところでは、一番頭の、スマホ・自動車センサー等のソーシャルなデータをもっと活用すべき。信頼性に疑問のあるデータ・情報が広

まらないようにする取組（品質管理）というのが必要ですね。その下に2ポツ目から6ポツ目まで、この辺は先ほどご紹介していただいたものと一緒かと思えます。

産・学・官が一体となった効果的な取組の推進といったところでは、気象庁の担う部分の明確化と他機関との連携が大事である。連携といった観点で、気象と別のものを組み合わせた融合研究・共同研究が他の分野以上に重要。オールジャパンにとどまらず、国際的な取組が必要ですねといったお話をこれまで、第1回目にいただいたものを含めて頂戴していると考えています。

次のページですが、一番上に、先ほど3つ掲げた、防災、日々の生活、経済活動に寄与したいというところがあって、そのために、下のほうに、技術開発の観点、それから、気象情報・データ利活用の観点の相乗効果で実現していきたいというところがありまして、技術開発のほうは今回ご議論いただいて、右側のほう、ちょっとグレーにしてありますが、ここは次回、ご議論いただきたいと考えているところだという趣旨でございます。

左側のところ、気象データ・情報の観測・予測精度向上といったものに向けて、目指すべき水準、それから、到達するための取組も考えていく必要があるというのが、その箱の中で考えてありまして、ただ、気象というものと地震・津波・火山というのは、ご意見を踏まえて、分けて考えていくんだらうなど。

それを推進するIoTやAI等の新たな最先端技術を取り込んでいくのが大事であるというのが下のほうです。

左側のちょっと赤いところでありまして、ただ、ニーズを見ながらやっていくというのがありますし、ニーズも変化していくし、技術も進展するものがあるので、PDCAサイクルを回していくということで、目標も再設定していくのだらうというところです。

その次ですが、では技術的な水準はというところについてですけれども、先ほども言った防災、日々の生活、経済活動に、ニーズに応じた精度向上という、赤字のところは全て、終わりがほとんど「精度向上」となっていますが、つまり、どこまで精度を向上させるんですかとありますけれども、技術水準とニー

ズというものを勘案していくんだらうなといったところでございます。

ここから、気象と地震・津波・火山を分けて考えたいというところで、気象分野に関する展望で、気象といっても、これはリアルタイムというか、実況のところから100年先、温暖化にかかわるところまで、幅広く利用されていくという基盤情報（ソフトインフラ）となっている。上のほうの青い字のところですが。

2番目では、ただ、ニーズというのはどんどんパーソナライズ化していくところで、高度化及びわかりやすさというのが求められる。

常に高精度・高品質な気象情報を社会に提供していくことが極めて大事であるというところで、下の、短いタイムスケールから長いタイムスケールまでいろいろなところで、つまり、ナウキャストから温暖化まで、それぞれのところの高度化をどういうふうに検討していくかという話になる。

右側ですけれども、技術といったところで、どう支えていくかとありますけど、短いところでは、まず、実況監視から観測網のところ、それから、他機関データも導入をしていくか、活用させていただくことはあるだろう。もちろん気象庁が活用するだけじゃなくて、気象業務としていろいろな、社会として活用していただくということ。

予測のところでは、数値予報技術というのがありますけれども、そういったところで、まさに技術の向上という中で、これは実況のほうもそうですけど、AI等をどう活用していけるかというところはあるんでしょうねと。長いところでは、長期の予測となると、いろいろな要素を全部組み合わせないといけないので、「地球システムモデル」とここでは書かせていただきましたけど、そういったものを目指す。いろいろな機関との連携を踏まえてやっていくということであろうということです。

短いほうから、これは現在から1時間程度、ちょっと漫画チックになっていますけど、今、左のところがあって、大雨とか降ってきて、この後どうなっていくのかとよく知りたい。それから、ピンポイントの情報が欲しいといったものがあるのかと。

2030年に目指すべき目標としては、こちらに書いてあるような、解析データやナウキャストの充実、精度向上を図っていきたいというところです。

下に、気象庁の取組とありまして、「具体は次のページ」と書いてありますけれど、次のページというのはこういったところで、いっぱい書いてありますけれども、今後のところとして、下に、実現に向けた基盤観測の充実・高度化、それから下の右側で、いろいろな主体でなされている観測データをお互いに有効に使っていく。結局、ビッグデータとなるわけですが、それをどのように活かしていくかということかと思えます。

そういったものを活用しつつ、上のほうにありますとおり、2030年のところ、豪雨、雷、突風等の激しい現象について、「シビアストームアラート」として出せないかなとか、それから、今でも面的情報というのは、予測とありますけれども、推計とか、こういったものの予想を増やしていくといったところもあるだろうなと考えています。

データ利活用のところは、先ほど申し上げたとおり、詳細は次回、ご審議いただきたいと考えています。

それから、半日から3日程度のところ。先ほど大雨というところの、そういうものが起こるといふのをどのくらい前から見通すことができるかというところです。そういったところをできるだけ早く見通せるようにしたいというところで、早目早目の避難に資するような情報をどうやって出せるかといったところです。

これは例として、線状降水帯と、それから、台風とかに伴い降るときの大規模水害時の、そういったものに資する情報を早いうちから出していきたいというところがございます。

上の線状降水帯については、12時間前ぐらいから線状降水帯の発生を予測したり、さらに言うと、それがそこにずっととどまるのかどうかといったものについて、取り組んでいきたいというところです。

下の台風については、進路予報をさらによくしていくというところもありますけど、流域の総雨量の予測も精度を上げていきたいといったお話でございます。

す。

次に、1・2週間から数カ月というところで、その下に気候と書いてありますが、だんだん気候的な方向になっていくところですけど、そういった例示というか、1・2週間、数カ月したところでの見通しで、自分たちのところがどうなるのかについての情報を高度化していきたいといった話です。それがこれですね。

もう一つ、続けまして、これはほんとうに長くなりますけど、数十年から100年後、これは温暖化のところになってきますけど、こうなってくると、一人一人の住民がというニーズではなかなかぱっと出てこないところを書くので、これは「自治体等」と書いてありますが、自分の街が将来どうなるのか。それから、100年ごとまでいかなくても、もうちょっと目先の近い先、大雨だとか増えるんだろうか、どうなるんだろうかといったものを教えてほしいなといったニーズ、つまり、自治体等における適応策といったものに貢献していく必要があるんだろうといったところですよ。

その次で、これは先ほどの、1・2週間、1カ月のところですけども、2週間先、1カ月先までの予測情報を高度化したいという話。

それから、もう一つ、さらに長いところですけども、その辺についても、めり張りのある予報というのは、なかなかそこまで先になると出していない状態ですけど、もうちょっとその精度を上げていきたいと書かせていただいているところであります。

地震・津波・火山については、最近も、地震であるとか火山噴火といったもの、草津白根であるとか霧島のほうとか、噴火等あるところですが、そういう極めて重要な情報ですが、なかなか、先ほど申し上げたとおり、予測の困難性もある。そういったところで、地下で発生する現象を確実に捉えて、観測・監視していく体制の充実を通じて、的確に評価して、今後の推移・見通しについて情報提供していくのは大事だろうといったところですよ。

これは字でいっぱい書いてありますが、この後に行ったほうがわかりやすくなると思います。

地震分野では、緊急地震速報、世界最高のI o Tとっていただいたところもありますけど、自分のいる場所がどんな揺れになるのかというのがあるといいなど。それから、大きな地震があったとして、その後がどうなっていくのかについての情報が欲しいなといったところです。

現在が左側、右が2030年と書いてありますけど、2030年のところでは、面的な揺れの広がりといった予測を提供するとともに、震度だけではなくて、高層ビル、高層階が大きく揺れる、長周期地震動階級といったものもあわせて情報を提供していきたいということです。

それから、地震活動の推移、見通しについても、より評価をできるようにしていきたいといったところです。最近では南海トラフの地震が注目されているところでもあります。

津波分野ですが、今、津波警報のところ、データベースを用いて、即時的というか、3分程度以内とかで出せるようにしているところですけど、いつまでその警報が続くんだろうとか、もうこれ以上大きい津波は来ないのだろうかといったところで、つまり、今後の津波の全体像という、言い方が適切かとありますけど、そういった情報が出せないか。その次のページでは、2030年のところで、右下のほうにあるとおり、第1報というのは、引き続きデータベースで出していくんですが、その後、どういうふうに減衰していくんだろうかといったところについては、シミュレーションに当たって考慮するデータとありますけど、こういったものを活用しつつ、全体としてどうなっていくのか、いつぐらいに解除できそうなのかといった見通しまで出していきたいというお話であります。

それから、火山のほうでは、上に厳しい顔をした専門家がいらっしゃいますけど、噴火警報等的確な発表のためには、観測体制の充実と研究成果と、監視・評価体制を強化していくことが必要じゃないですかと。

それから、灰のほうは、噴出した後については、基本的に上空の風に流されていくわけですけど、そういったところの精度を見ていくところはもうちょっと高度化できないかというところがあります。

右側のところになって、火山そのものについては、火山帯の内部構造、つまり、マグマだまりがどこにあるかとかそういった話になるかと思えますけれども、そういった知見を高めていきたいなど。それにより、評価をよりの確にできるようにしたい。

下は、降灰のところ、これは噴煙現象の全体像をリアルタイムで把握して、高度化、先ほど、上空の風で流れたり、そもそもどれだけの灰が出たかというのも関係しますので、その辺との兼ね合いも必要だということなのです。

これらを踏まえたところで、最先端技術の活用で、まず、IoT・ビッグデータの活用について、ビッグデータの、社会のさまざまなところで活用できるソフトインフラだと考えています。さらに、近年、社会における気象観測データの流通も広がっているところであります。そういったものをうまく組み合わせることによって、革新的な付加価値をもたらす可能性があると考えています。

今後に向けて取り組むべき事項としては、産学官のさまざまな主体で行っているデータを、どういうふうに集めるというだけじゃなくて、どういうふうに共有して活用するか。それは気象庁として活用させていただく場合もありますし、気象業務として、社会全体として活用していただいているということもあるだろうと考えています。そののところでは、ご意見にもありましたけれども、品質管理が必要である。例えば品質管理として、ベストエフォート型、品質保証型をうまく使い分ける必要もあるんじゃないかということなのです。

下に書いてあるところでは、「データ品質の見える化」等、データの流通・利用拡大に資する環境整備が必要と書いています。産学官の役割分担がありますけど、官の役割として、こういった環境整備は大事になるのかもしれない。

ちなみに、けさの朝日新聞で、膨大な情報を使いやすいように気象庁も利用環境を整えるべきだといった記事が出ているところでございます。

それから、最先端技術のAI、先ほど先生から貴重なプレゼンをいただいたところであります。識別とか予測といったところを分けて考えていくべきだというご意見もいただいているところですが、こういった中で、AIのところ、先ほど先生から、nが増えないとだめだとありましたけど、nをどう

やって増やすかというのもありますけど、工夫をしつつ、こういった形で生かしていけたらよろしいかなと考えているところであります。

こういった中では、やはり専門的知見が非常に重要だなというふうに、先ほどもご教示いただいたところなので、さまざまな機関との連携をしながら、知見を積み重ねることが大事だということを、今後の取り組むべき事項に上げさせていただいているところです。

その他、先ほど、ここで申し上げましたけど、社会ニーズの把握、目標設定、不断の改善（P D C A）といったところをうまく活用していったって、2030年に向かって活用を進めていきたい。

その中で、経済活動という観点では、昨年設立されました、越塚先生に会長をしていただいている、気象ビジネス推進コンソーシアムといったもの、それから、防災では、今後、災害が起こったところで、そのときの防災対応として、どこがよかったのか、どこが反省点かというのを、「振り返り」もありますので、そういったものを用いて、自治体さんのニーズを把握しながら進めていく、こういうふうな不断の改善を進めたいというところです。

それから、連携については、わりと気象・地震・火山分野は、予測の困難性の一方で、他機関との連携というのは進んでいる分野であるというところがあります。その他、AI等のところで、どういうふうに産学官が連携していくかといったところもあると思います。

今後に向けて取り組むべき事項で、下にありますとおり、オールジャパンにとどまらず、国際的な取組も大事であろうといったところです。さまざまな連携をすることによって、気象と別なものとリンクさせることによって、潜在性のあるビッグデータを社会に貢献する情報として活用していけるのではないかと考えているところでございます。

以上でございます。

【委員】 どうもありがとうございました。

それでは、これから、「気象業務に関わる技術開発」に関するご意見をいただければと思います。どなたからでも結構ですので、よろしく願いいたします。

では、委員、お願いします。

【委員】 最初によろしいですか。僕はコンピューター屋なので、あまり気象の予測精度の向上とか、そっちのほうのご意見はできないんですけども、そういう意味で、IoTとかのデータの共有とかそういうところに関して言うと、今日の先生のお話からも、データに関しては明らかなように、nを増やしていくということですね。だから、なるべく、より多くのデータをより多くの人に提供して、より多くの人にさわっていただいたり、やっぱりより多くのデータが必要だと非常に思いますので、そういった活動が必要ではないか。

だから、それを支えるようなシステムであるとか、プラットフォームであるとか、また、気象庁とか気象会社だけでは、やはりカバーできない、欠損しているような、また、nを増やすという意味では、そこを、民間の気象以外の業務で得られている情報も、いかに活かしていくかは、今までの話ではありますけれども、まさに今日の先生のお話で、そういうことをどんどん進めていく必要があると思いました。

ただ、そのときに、データでクオリティーの話も一言させていただきます。今日、まさに先生のお話を聞いて思ったことは、データは2通りあって、人に直接見せるデータと、AIとかでも、シミュレーションでも、統計でもいいんですけども、プログラムに食わせて計算させるデータがあって、これらは違う気がしていて、それで、今まで、人に見せるデータ、例えば、気温がこれから何度になりますとか、これからどうなりますとか、人に見せるものは、オープンにすると、精度が社会に与える影響は非常にダイレクトだと考えられるわけですね。

計算に使われたというのは、おそらく今までは、気象庁内部の利用であるとか、場合によっては、専門家間で共有していればよいものだったと思いますが、これからのことを考えると、それも多分、オープンにしていかなければいけない。ただ、計算のためにオープンにする基準と人に見せるデータをオープンにする基準が同じだと、多分、やっていられない感じがします。例えば食べ物でも、食べるトマトとトマトジュースにするトマト、あと、ブドウでも、食

べるブドウとワインにするブドウは、品種も違えば、育て方もつくり方も違うので、データもそうかなと思います。

今、人に見せるデータに関しては、制約とか規制がありますが、それと同じことを計算させるデータに適用すると、かなりデメリットが大きいのかなと感じて、その辺の制度整備が必要と思ったのが第1点です。

もう1点は、人に見せるほうのデータですけれども、人に見せるので、これも今日のまとめですと、まさに4ページ目にありますように、技術とリテラシーの2つの相乗効果部分が非常に小さいのですが、ここは、僕は結構大きいと思っています。単に人に見せること、気象関係でよくお伺いするのは、例えば、天気予報のテレビの番組があったときに、ここ数十年、それほど大きく変わった見せ方もしていません。だから、今、例えばリテラシーの向上ということであると、リテラシー向上の活動そのものに、技術を寄与させなければいけないと思っています。

例えば今の技術水準でできる世界最高の天気予報コンテンツはどのようなものかは、あまりやられていません。災害の情報を出すにしても、天気を伝えるにせよ、今は、結構古典的で、テクノロジーがそのリテラシー向上に生かされていません。というのも、僕自身がそういうユーザーインターフェースとかビジュアルライゼーションをもともと専門にしていたので申し上げていますが、見せ方自体も技術なので、もう少しリテラシー向上というか、国民にどういうふうに気象とか災害の情報を伝えるかというところにも世界最高の技術を投入して、見やすい、わかりやすい情報の提供の仕方の研究も必要と思いました。

以上です。

【委員】 どうもありがとうございます。

委員、お願いします。

【委員】 産学官連携で、資料ですと後半のほうに当たるんですけども、こちらで少し、私のほうからご意見をお伝えします。

特に、先端の気象データを活用してどのような産学官連携ができるかというところが、これから新しくやりたいところかと思います。イメージ的には、気

象の科学技術と、掛けるエックスといいますか、いろいろな物流ですとか、農業ですとか、医療ですとかという別の分野、それから、データ関係の方々の組み合わせになるかと思えます。別の言い方をすると、気象庁の技術者の方たちが一つ。大学や研究法人の研究者がもう一方にいる。そして産業界や自治体という組み合わせになると思えます。

通常、文部科学省や経済産業省のような研究開発支援の事業をたくさんお持ちのところだと、こういったテーマでやってくださいと予算を出して、参加していただくというのがしやすいと思うんですが、なかなか気象庁はそれが難しいということで、どうしたらいいんだろうということを少し考えてまいりました。

1つは、気象ビジネス推進コンソーシアムのように、実際、顔を合わせる場をつくってしまうことが考えられます。その中からおそらく、これだったら一緒に組んで進められるよねという動きが出てきて、そうしますと、学術研究者の方が研究代表者になって、文科省や経産省のプロジェクトに応募するというのが比較的やりやすいですし、それを狙いたいところかなと思えます。

そのほかに、比較的できるかなと思ったものとしては、シンポジウムみたいな形です。新しい研究ネタを探している学術の研究者の方たちに、気象庁のこんなデータを使って、どうですかとアピールするような場というのは、一回のシンポジウムでもそれなりに効果があるかなと思いました。

それから、データ関係だと、先生が前にもおっしゃったように、随分いろいろなコンテストが出てきているので、そういった場に持っていく題材にする。ビジネスプランのコンテストも随分最近、イノベーションの関係で増えています。社会実装に向けた文部科学省のセンター・オブ・イノベーション事業などもありますし、あと、若手研究者、学生を育てるために、若手にビジネスプランを出させるという、EDGE-NEXTのプログラムなどもあるので、そういったところにもどんどん、実は気象庁はこんなおもしろいものを持っていると、持って前にどんどん出ていく。そうすると、それだったら我々が使って、ちょっとおもしろいことをやりたいねという動きが出てくるかなと思って

います。

今ある国のプログラムで、これは多分、気象庁でもご担当者の方はご存じかなと思うんですが、念のため、情報共有のためにご紹介します。

1つは、内閣府オープンイノベーションチャレンジというものです。これは国の機関のニーズに対して、中小とかベンチャー企業の技術を発掘して事業化を後押しするもので、2017年スタートです。最初の年は、警察庁、消防庁、海上保安庁のニーズに対してで、どんなことができますか、研究を支援しますよというプログラムです。一例ですと、海上保安庁から、船舶のメンテナンス作業軽減を何とか進めたいねというニーズを出したことに對して、筑波大学発ベンチャーが、水中ドローンによって船底点検を簡単にするというのが出ている。こういったところで気象庁のお話を持ってくると、結構反応があるのではないかなと想像しました。

もう一つが、文部科学省系の科学技術振興機構で力を入れて、やはり17年度からスタートしている未来社会創造事業です。これは、社会産業ニーズに對して研究開発を行うものです。JSTとしては、学術的な研究成果をもとに豊かさを生み出すという方向性と、将来をこういう社会にするんだという目的を定めて、バックキャストするという、大きな2つに事業をまとめていこうとされていて、この事業はその2つ目のほうの柱と位置づけられています。

これは、文科省がこういうのでやってくださいと設定するのではなくて、ユニークなものです。文部科学省としては大きな4領域を提示しますが、その中で、どんな研究テーマで考えられそうですかねというのを公募するという新しいスタイルをとっています。つまり、例えば気象庁のこういったデータを使って、新しいデータと気象の研究をしてみませんかというテーマをここに設定するというのも可能だということです。

まず、いろいろな研究テーマのニーズを社会から集めます。こういうことをやってほしいんだけどと企業など産業界から来て、そのテーマを定めた後に、これについて、研究開発課題を応募してくださいという公募をとります。2つ目の段階で、具体的に研究者の人たちが応募してくるわけですがけれども、もと

もと社会での実用化、実装を掲げているということで、自然、産学連携につながるようなものが多いと聞いています。

JSTの担当者にお伺いしますと、実際、気象研ですとか、日本気象協会とも話をしたと聞いているので、ぜひ、これは活用できるんじゃないかなと思っています。

こんなテーマをお願いしますという形で、気象庁側から手を挙げることもできますし、あと、審査側のオブザーバーに気象庁側が入るということで、テーマをどういうふうに固めていこうかというベクトルづくりに協力できるんじゃないかというふうに伺っております。いろいろなところから出てくるもの、社会のいろいろな声を集めるということなので、テーマは随時に集めていって、1月ぐらいに締め切って、次の新年度に投げかけるというふうに伺っています。特別な大きな研究開発予算をこちらが用意しなくても、可能かと思います。自然に産学官の動きが出てくればいいなという、企業さんのお金が引き出せればという後押しが今、政府の動きでもありますので、狙えるかなと思っています。

東京のほうの中央だけではなくて、地域でもこれは可能だと思っています。地域の気象台さんがするシンポジウムも、年間たくさんされているということなんですが、そういったものですとか、コンテストみたいなものもそれぞれの地域で可能だと思っています。それから文部科学省でも、地域活性化のための地域産学連携というプログラムもありますし、そういったものをうまく組み合わせることで、広がりが出てくるんじゃないかなと。

気象庁のデータを使って、ほかの人と組むとこんなことができるんだという事例が少し出てくると、人気が高まるんじゃないかなというふうに思っています。

以上です。

【委員】 大変具体的に、やれそうなことをご提案いただき、ありがとうございました。今後、検討してもらえればと思います。

先生、お願いします。

【委員】 どうもありがとうございました。

3点ほどになりますけれども、1点目は、簡単なことで、先ほどの資料1の6ページには、目指すべき気象業務のあり方ということで、方向性がありました。(再整理)というスライドなんですけれども。その中で、一人一人の生命・財産、これが守られるという言葉があって、もちろん言うまでもなく、これは原点に相当する。それに貢献するという、気象業務というのはおそらくそういうことだったと思って、議論もされてきたと思います。これと、今の資料の4ページの2030年における目指すべき気象業務に向けた取組の3つの柱があって、これも今までの議論でこういうことを、方向性としては、ポジティブなところでは非常にわかるんです。けれども、今言った、生命・財産を守るみたいな、一番ベースになって、時代が変わってもずっと担わなければいけないような基本的な役割のところ、何となく見えなくなっているところがありまして、両方の資料にちょっと整合性がない。そのあたりについてはどうなんだろうかというのが、質問というか、意見ですね。簡単に変えてもらえばいいだけですけれども。

2点目は、先ほどの 先生のご指摘にもかかわるんですけれども、今回の議論では、特に気象と、要するに、先ほど 先生おっしゃったみたいに、地面の中と空の上というのはかなり違うぞということ。一方で、この何十年の議論もあって、国民的にも、地面の中を予測するのはまず無理だねというところが、何となく了解されている部分もあります。少なくとも、ほぼ間違いなく困難であるというのはわかっているわけです。一方で、気象のほうは、頻度高く結果がわかるということで、当たった、外れたがしょっちゅうわかっているわけですから、それに対しての精度向上をしていくという、目標設定も非常にしやすい分野だと思います。今回は特に、技術的にジャンプするので、そういうものも活用しながら、2030年に向けて飛躍的に精度を高めていこうというような、目標が設定しやすい分野だと思いますので、そういうことを考えて意見を申し上げたいと思います。

今日の資料の最後のところにPDCAが出てきますよね。24ページにPDCAの取組があって、PDCAというのもよく知られた手法だし、当然、こう

いう考え方は持つべきなんだけれども、これを何のためにやるかという、多分、長期の目標もあって、それを短期間で評価しながら回していくわけですから、見直しが必要であればしていくという、これはいわば計画とか、少なくとも戦略だとか、何らか具体的な行政計画のようなものがはっきりとあって、それに向けて取り組んでいく。そのような建付けとしてはよくわかるんです。けれども、そういうことがはっきりしないままやっていくというのは何だろうかという、クエスチョンが逆につきます。これは最初の回にも申し上げたんだけれども、ぜひ、技術にかかわる気象庁の将来計画というのか、戦略計画というのか、そのような形づくりをしていただいて、それに対してP D C Aがうまく回っていくということで、しかもビジネス等々の新しいこともどんどんやっていくわけです。一方で、普遍的というかな、国民に対して生命・財産を守っていくみたいな基本的なところもきっちりと進めますよ、というメッセージがちゃんと伝わるような形づくりに、やっぱり向かっていくべきじゃないかなと思います。改めてもう一回申し上げましたけど、たまたまP D C Aが出てきているものだから、そういうふうに申し上げました。これが2点目です。

3点目は、先ほど、産学官連携だとかいろいろお話があって、私もぜひ進めていただきたいという思いです。コンテストがありました、今回議論になっている予測精度を高めるという部分に関しての新しいコンペティションとかコンテストだとか、そういうものを産の連携も含めて、官学が中心かもしれませんが、競争環境をつくってポジティブにやっていくことができる就非常に広がるし、効果もあるのではないかなと思います。ちょっと加えさせていただきました。

以上です。どうもありがとうございます。

【委員】 どうもありがとうございました。

では、委員、お願いします。

【委員】 先生、難しいお話でしたけれども、わかりやすく、ありがとうございました。

お聞きしていて、最後、気象庁の方策の23ページに、AIの活用について

というところがあります。例えば気象衛星画像の解析とか、それから、いろいろデータ、余震分布のデータの画像になりますけれども、生データじゃなくて、そうした画像などから何か読み取れないのかとか期待を持ちます。例えば火山の先生でも、火山が噴火したときの噴煙の様子の変化から何かを読み取ろうという取り組みも以前からあるんですけれども、例えばそういうものを、AIから何か読み解いていけるとか、いろいろな可能性が感じられます。ここにまさにそういうイメージで、写真が3枚ぐらい載っているんですけれども、先生のレクチャーをもとに、もっとこういうところを、まだ、実現は10年後難しいかもしれないんですけれども、夢を持った提言というか、書き方を若干付加してもいいかなと思います。せっかくの提言なので、先生のご助言もいただきながら、膨らませて書いていただいたらいいなと、お聞きしながら思いました。

また、感想で申しわけないんですけど、産官学というお話が先ほどから出てきているんですが、お聞きしていて、これはもう産官学という言葉ではおさまらないなと。ビッグデータもそうですし、あるいは、民間のアマチュアの人たちが持っているデータとか、携帯、スマホなんかのデータもそうですけど、アマチュアの観測者、あるいは民間企業の持つデータなんかもどんどん取り入れていくべきフェーズかなと思って、これは産官学に変わる、何かいい言葉はないのかなとちょっと思いました。

申しわけございません、もう一つだけ。今日の技術についての話とちょっとずれるかもしれないんですけれども、分科会はあと2回あって、この後も気象データとかそういうお話が中心なので、あえてソフトのお話にちょっと触れさせていただきます。先週、宮崎の硫黄山が噴火して、火山の予測は非常に難しいんですけれども、今回はいろいろな意味で、すごくうまくいったケースかなと思うんですね。

霧島連山は火山活動が活発になっていて、新燃がああいう形で激しく噴火をして、硫黄山、ちょっと離れたところも、地殻変動のデータとか地震のデータなどでずっとウオッチしてきて、一旦落ちついて噴火警戒レベルを下げて、またおかしいぞと言って上げて、ウオッチしているところで、地殻変動のデータ

なんかも捉えられている中で、来るぞ来るぞというところで噴いた。

その後、噴火速報も、草津白根と違って的確に、迅速に出ましたし、映像できっちり捉えられていて、すごくうまくいったケースでした。申し上げたいのは、さらにそこから先なんですけれども、直前予知というのはできなかったわけなんですけれども、構えているところに噴火して、速報がすぐ出て、それで、近くの1キロほどのところに宿舎とかホテル、国民宿舎とかいろいろなビジターセンターみたいなものとかがあったり、数は多くなかったんですけれども、登山者が、散策している人が比較的近くにいたりして、その人たちに情報を伝えるまでの連携が、連絡が非常にスムーズで、避難もスムーズにいった。

警察がヘリコプターを飛ばして伝えたり、防災行政無線をやったり、エリアメールをやったり、やっぱりこれは、気象庁を核とした情報を発信して、それを防災に活かしていくという、活火山法が御嶽山以降、変わりましたけれども、それが非常にうまく機能したケースだなと考えています。その方向性の正しさというか、そういう学習をしてきたケースかなと思います。その背景には、やはり気象庁を核として火山防災協議会というのをつくって、火山ごとに話し合っているというところが功を奏している。例えば警察がすぐヘリを飛ばすなんていうのは、そういうところの話し合いでできるわけですね。

前回は申し上げているんですけど、これは火山に限らず、気象災害、土砂災害もそうですし、河川の氾濫なんかでも、災害ごとに地域の協議会をつくって情報共有や防災対応にあたる取組みが進められています。こうした場が非常に大事で、ここはこれまで申し上げていないところなんですけれども、レジユメの最後のほうにも、防災のそういう取り組み、レビューをするところがあるんですけれども、地域ごとに防災の横のつながりというのをもっと強くしていくことができないか。

理想を言えば、例えば気象庁も、県庁とか、河川国道事務所とか、全部一緒のところに集めて、監視をして、何かあったときに備える。それで市町村にアドバイスをするというところが理想だと思います。なかなかすぐさまそういうことには行かないので、今もやっていることですけれども、各組織の連携を強

めて、例えばテレビ会議システムとか、データを共有する仕組みは今、すごく発達しているわけで、そういうものを思い切って取り組んでみるという、モデル的な事業はできないのかなと。全国47都道府県がある中で、熱心な都道府県がありますから、気象庁が音頭を取って、火山でもそうですし、気象災害でも、土砂災害、河川災害でもそうなんですけれども、そういうモデル的な事業を今後10年、少し意識して取り組んでいくことはできないかと。

行く行くは、それこそ国レベルでの防災組織、縦割りの組織を再編するという必要もあるでしょうし、それを先駆けて都道府県、自治体レベルで、防災に関して一堂に集まって、顔の見える関係、あるいは声の聞こえる関係で、予防的な、防災に資する情報を発信していけるようなトライを何か期待したいなど。その提言の中に、そういう側面ももう少し強めて書き込むことはできないかなと、そんなふうに感じました。

長くなりましてすみません。

【委員】 どうもありがとうございました。

【委員】 もう一回、一言だけいいですか。

【委員】 今のお話しに関係しますでしょうか。では、お願いします。

【委員】 今の さんからの話で、同感なんですけど、だけどやっぱり、何というのかな、この国は特にそうなんだけれども、何か始めようとするときに、よかった経験ばかり言うんですよね。これはよかったですね、これはうまくいきましたね、こんなことをやるといいですねとばかり言うんです。

僕は、さっきの発言とも関係するんだけど、それだけではやっぱり気象庁という見識を示せないと思っているんです。というのは、気象の話じゃないですけども、2010年ですけど、公共事業の事業評価というのをやっているわけですよね、今もずっと。もう20年やっていますけれども、そこで、民主党政権ですけども、精度を上げるために、より厳正にやるために、今までは5年置きにやってきた再評価のやり方を、3年に短くしましょう、これは大変いいことだと言って、やったわけですよ。膨大な作業量が増えて、そのために膨大なお金を使って、何が起こったかということ、5年を3年にしてみても、変化

なんて何もありませんから、無駄に発注して、無駄に作業量だけ増えたというのがここまでの経験です。

だから、それをもう一回改めようという方向にありますが、言いたいのは何かというと、やっぱりできそうなことで、つまり原理的にもできそうなことで、技術によって負うべきことは、ぜひこれは期待を持って言うべきだと思うんだけど、原理的におかしいようなことを、あたかも技術が頑張ればできそうなように言うというのは、明らかにミスリードで、例えばさっきの地震予知にしたって、寺田寅彦の非常に原理的な数学を使った論文によると、最初からできないと言っているんですよね、巨大地震の予知なんて。

あの当時からわかっている。だけど、ああいうところに走っちゃったということを見ると、ぜひ、さんがおっしゃったような面と同時に、もうちょっと良心的に、技術と社会というものを捉える、そのところを言うということもやっぱり責任だというふうに強調させていただきたいと思います。

以上です。

【委員】 どうもありがとうございます。確かに、技術的にどこまでやれるかをきっちりと説明するということは大変重要で必要だと思いますけれども、防災につなげる上では、連携をよくしていくということは大変重要だと思いますので……。

【委員】 それもそうだし、技術的にできないことははっきりありますからね。

【委員】 はい。それはちゃんと伝えるということで、今も、気象庁もいろいろ努力されていると思うのですが、ご提案があったようなことも検討していただけばということだと思います。

委員、お願いします。

【委員】 今の委員のご意見、大変賛成です。

時間がないので、1つだけ考えを述べさせていただきますと、気象庁の気象予報というのは、やはり予報方程式をよくして、解像度を上げて、予報をよくしていくというのが大本山だと思います。一方で、AIを使いまして、データ

活用や情報伝達のあたりで、非常に見通しがよくなるということも、時間スケールごとに上手にまとめられて、わかりやすくなっていると思います。

1つ感じましたのは、予報方程式とか物理過程というものをよくしていくこと自体にもAIというものを今後もっと活用できるということが、もう少し前に出てきてもよいのではないかと。そのあたりを今後の展望として考えられてもいいんじゃないかと考えました。

以上です。

【委員】 貴重なご意見ありがとうございました。

【委員】 1つだけよろしいですか。夢がある話と 委員がおっしゃっていて、私もそこは大事かなと思っているんですが、少子化でどんどん人が減っていて、どこの分野も人がいないと言っている状況ですので、2030年の予測を考えたときに、2030年に予測業務に携わっている人がちゃんといないといけないというのは、結構深刻な問題だと思うんですね。

そうすると、2018年の今に、中学生とか小学生、高校生とか、それぐらいの世代が、気象予測は技術的におもしろいものだと感じてもらうのは結構重要かなという気がするんですね。ですので、こういう提言とは少し趣旨がずれるかもしれませんが、若い世代に気象予測の重要さとかおもしろさをしっかり伝えていくというのも、提言としては必要なのかなと感じました。

【委員】 どうもありがとうございました。

私の不手際で時間が押してまいりました。いつも最後にまとめをさせていただいていますが、時間がありません省略させていただき、最後に、その他の議題として、顕著な災害を起こした自然現象の命名について、気象庁のほうから説明があるということですので、お願いします。

【企画課長】 今、新野分科会長からご紹介いただいた経緯については、1回目の分科会で、気象庁として、基準といったものについて、あと、命名の仕方について、検討をしているところですよというご紹介をさせていただいたところですよ。

本件につきましては、今回、この場ではありませんけど、気象庁で現在考え

ている案を、各委員の方にご説明させていただいて、既にご意見を頂戴したところでございます。

気象庁としましては、他の学者の方や報道機関の方からも意見をいただいているところでございますので、それをあわせて、引き続き検討させていただいて、次回の分科会で、どのような検討をした、どのようになったのかをご報告させていただきたいと考えております。

以上でございます。

【委員】 どうもありがとうございました。

それでは、大体時間になりましたので、本日はここまでということにさせていただきたいと思っております。次回は、2030年に目指すべき気象業務に向けた取組のうちで、残っております気象情報・データの利活用推進を議論するとともに、提言の骨子についてもご意見をいただく予定にしております。関連する資料の準備を、事務局のほうでよろしく申し上げます。

ということで、本日予定しておりました議事は以上でございます。

それでは、これで第26回気象分科会を終了したいと思っておりますが、最後に、事務局のほうから何かございましたら、お願いいたします。

【総務課長】 新野分科会長、また、委員の皆様、ご審議大変ありがとうございました。

2点ほど連絡させていただきます。

1点目は、議事録、会議資料の扱いでございますが、議事録につきましては、委員の皆様には後日、議事録の案を送付しまして、ご同意をいただいた上で、本日の会議資料とともに公開したいと思っております。

2点目です。次回の予定でございますが、第27回気象分科会は、資料にもありますとおり、7月2日、15時30分からを予定しております。お忙しい中、大変恐縮でございますが、委員の皆様のご出席をお願いしたいと思っております。

本日はどうもありがとうございました。

了