

第2回CREST「人工知能」研究領域 成果展開シンポジウム 質疑

項番	講演	発表者	質問	回答
1	深層学習のためのCo-Designフレームワーク	篠田 浩一	attention based のCNNが全盛時代、ノード間通信が局所性を持たなくなるので、高速化できないのではないのでしょうか？	現在の省資源なプラットフォームのためにチューンしたモデルはそのままでは超並列で高速化するのは難しいと思います。超並列を前提としたモデル化を考える必要があります。
2	深層学習のためのCo-Designフレームワーク	篠田 浩一	150K nodesのような超並列環境だとモデルパラメータのstalenessの扱いが益々重要になるが、どのように対処することが可能なのか？	stalenessに対してより頑健なアルゴリズムを開発する必要があると思っています。
3	深層学習のためのCo-Designフレームワーク	篠田 浩一	応用毎に edge computing の割り振り、圧縮率と精度のバランスを adaptive に調整する必要がある。どのように開発支援する想定か？	応用を絞った開発を行うとともに、幅広い応用に使えるシミュレーションツールを開発したいと思っています。
4	深層学習のためのCo-Designフレームワーク	篠田 浩一	速すぎて応用がないと説明されましたが、詳しく教えていただきたい。	その後研究に進展があり、高解像度画像における大量の物体検出に活用できると考えています。
5	異種サービスをつなぐプライバシー保護を考慮したAIマーケティング技術の実現	原 隆浩	Amazonのオススメモデルを、他のEコマースで活用するのなことでしょうか？	Amazonには限りませんが、イメージは近いと思います。さまざまな業種のEコマースの購買行動や実空間の位置情報など、異なるドメインの情報を横断利用して、より詳細かつ細粒度にユーザ（顧客）の行動を予測することを目指しています。
6	異種サービスをつなぐプライバシー保護を考慮したAIマーケティング技術の実現	原 隆浩	ペルソナモデルをどうして時間変化するグラフとしてモデル化したんでしょう（扱いやすかったから？）	本質的にそうなることが理由ですが、実際には対象とするタスクに応じて、各ドメインのペルソナモデルにおいて個々のユーザ（ペルソナ）を分散表現（ベクトル）としてモデル化するなど、より適切な形式に変換（表現）して機械学習モデルの入力とします。 なお時間変化するグラフは決して扱いやすいことはなく、むしろビッグデータ処理の観点からは、非常に扱いにくい（処理に時間を要する）データ構造です。そのため、本研究では、研究課題1-2として、複雑な形式のデータを効率的に処理するための技術群の開発を行います。
7	異種サービスをつなぐプライバシー保護を考慮したAIマーケティング技術の実現	原 隆浩	個人を特定しない、とは同じ人かどうかはわかる仮名レベルでしょうか。同一類似ユーザを同定するとのことですが、そこから個人を特定可能な情報がどの程度漏れているのでしょうか。 また、誰が同定権限を持ち、誰が（一般ユーザ？）が持たないのか気になりました。	あるドメインのユーザ（ペルソナ）の情報は、そのドメインでのモデリング技術に従った表現（例えば、特徴ベクトル（分散表現）など）となります。ドメインをまたがったペルソナマッピングは、実際のIDなどではなく元ドメインでの表現をターゲットドメインの表現に変換する処理となります。つまり、ドメインをまたがって交換される情報は個人特定可能なIDではなく、解釈不能なベクトルなどの表現となります。この変換後のベクトルは、ターゲットドメインでのそのユーザの特徴ベクトルと完全一致することも現実的には起こりません。そのため、このマッピング（同一性・類似性判定）処理によって個人情報が漏洩することはありません。この同定処理は、基本的に各ドメインのサービスプロバイダが実施します。ただし、ターゲットドメインと元ドメインとの間で許諾済みのブリッジユーザ（両ドメインのID情報を提供したユーザ）の情報をを用いてマッピング規則を学習しますので、任意のドメインのサービスプロバイダが任意のドメインでマッピング処理（同定処理）を実施できるわけではありません。
8	異種サービスをつなぐプライバシー保護を考慮したAIマーケティング技術の実現	原 隆浩	移動のモチベーションとなる情報というのは、混雑度や広告のような情報を想定されているのでしょうか。時間や目的地のスコアによってかなり異なるように思いますが。	混雑度を含めて、居住地の属性や回遊先の施設の属性などの環境属性を、移動のモチベーションとなる情報として想定しています。 これらの環境属性に基づいてユーザの行動パターンを分析し、ペルソナモデルの構築のための一つのアプローチとして利用することを検討しています。これにより、ユーザの特徴および環境に応じて、異なった行動予測モデルを提供し、レコメンデーション（広告という形もあります）による行動変容を促すことを考えています。

第2回CREST「人工知能」研究領域 成果展開シンポジウム 質疑

項番	講演	発表者	質問	回答
9	異種サービスをつなぐプライバシー保護を考慮したAIマーケティング技術の実現	原 隆浩	ペルソナモデルを精緻化すると個人が特定できてしまうということはありませんか？ 精緻化 vs. セキュリティの評価が重要かと。	※二つ前のご質問に対する回答と同様ですが、改めて記載しております。 あるドメインのユーザ（ペルソナ）の情報は、そのドメインでのモデリング技術に従った表現（例えば、特徴ベクトル（分散表現）など）となります。ドメインをまたがったペルソナマッチングは、実際のIDなどではなく元ドメインでの表現をターゲットドメインの表現に変換する処理となります。そのため、ドメインをまたがって交換される情報は個人特定可能なIDではなく、解釈不能なベクトルなどの表現となります。
10	プライバシー保護データ解析技術の社会実装	花岡 悟一郎	コメント： 分散秘匿計算はAIと無関係と思われがちですが、この分野は非常に重要です。COMPETITIVE（競争）する組織間でCOOPERATIVE（協働）する造語として、COOPETITIVE Learningという言葉があります。	貴重なコメントをいただき、ありがとうございます。我々も、秘匿計算を通じて組織間にまたがるデータの相互活用が活発となることを期待しています。
11	プライバシー保護データ解析技術の社会実装	花岡 悟一郎	秘匿計算で情報漏洩が防げるのはわかるのですが、自分の銀行以外に自分の情報が使われるのは、プライバシー的には問題にならないのでしょうか？利益vsプライバシーや個人情報の所有権の問題。	各銀行の事業内容やお客様との約款で定められている範囲で銀行様自身が利用や目的を定められるものと考えております。本チームではそれを技術的に支援するという立場です。
12	プライバシー保護データ解析技術の社会実装	花岡 悟一郎	銀行のディープラーニングの暗号化の話で、ディープ以外の手法で使いたい人もいます。どのようなクラスのアルゴリズムであれば扱えるという基準はありますか？また、ライブラリとして公開する予定はありますか？	現時点ではニューラルネットベースの機械学習のみとなっています。ライブラリ公開については検討したいと思います。
13	プライバシー保護データ解析技術の社会実装	花岡 悟一郎	例えば、線形回帰やロジスティック回帰はディープ系のフレームワークで実装できますが、これらのアルゴリズムも使えと考えるとよさそうですね。	基本は、Downpour SGDにおいて勾配情報を暗号化して送る方式ですので、できるのではと思います。
14	プライバシー保護データ解析技術の社会実装	花岡 悟一郎	データ自体は銀行から出さずに、中央サーバーに送っているのは学習した結果のパラメータを暗号化したものだけ、ということでしょうか？	金融機関からサーバにアップロードするのは「暗号化された勾配情報」のみ、ダウンロードするのは「暗号化された重み」のみです。
15	プライバシー保護データ解析技術の社会実装	花岡 悟一郎	銀行の共同実験に、今から銀行が参加することは可能でしょうか？	今からでも可能です。銀行によってご事情はいろいろあるかと思いますが、まずはご相談下さいませ。
16	プライバシー保護データ解析技術の社会実装	花岡 悟一郎	アップロードする「暗号化された勾配情報」は個人情報やプライバシーデータには該当しないということでしょうか？	個人情報に該当するかどうかはまだ一般には定まっていませんが、該当しないとの見解に対する支持も強いです。また、我々としてもそのような見解がより一般的になっていくものと期待しています。
17	ビッグデータ解析によるサイバー脅威の検知と対策のアシスト	関谷 勇司	応用的な観点では手段は問わず精度が出れば良いもの、技術的には「画像化する」必然性というか design policy が知りたい。	画像化するというデザインは、通信の中身を通常の、たとえば src/dst IP、src/dst port、Protocol といった情報に基づいてベクトル化したデータセットを用いAutoEncoder 等で判定した場合に、あまり精度があがらずに画像化した後に既存の画像処理の手法を適用した方が精度が高かった、という背景があります。通信のデータセットから異常な挙動を機械学習で発見するための知見がほぼ存在せず、既存の手法より精度が高く発見できるというのが正直なところ。もちろん通信挙動としてのパラメータをそのまま解析する手法も試行しており、半教師学習を用いた異常発見にも取り組んでおります。
18	ビッグデータ解析によるサイバー脅威の検知と対策のアシスト	関谷 勇司	SYNデータの画像化は AutoMLのようなアプローチを使うことに違和感があります。下位レイヤーの構成が大事では？	下位レイヤーの構成は重要と考えています。IP パケットのヘッダ情報をそのまま多次元配列として利用しているわけではなく、検知したい攻撃にあわせて適切に用いるフィールドを選択し、密度情報等を計算して利用しております。本質的に検知のためには画像化する必要は無く、CNN を用いて学習と判定を行っております。CNN を適用するにあたって、既存のライブラリや手法を効率的に再利用するという点、ならびにネットワークの業界で標準的に用いられるトラフィックの可視化を行うために画像化したというのが実情となっております。

第2回CREST「人工知能」研究領域 成果展開シンポジウム 質疑

項番	講演	発表者	質問	回答
19	fishtechによるサステイナブル漁業モデルの構築	飯山将晃	FishTech興味深いです。ここ2-3年、日本で、陸上だとゲリラ豪雨などが発生するようになり、いわば気候帯が変わってしまったのでは（温帯→亜熱帯）という印象を受けます。 1. 海の水温も傾向として変化は生じているのでしょうか 2. もしそうなら、過去情報を教師データとして使えないと考えますがいかがでしょうか	1. 海洋環境については温暖化に伴うシグナルよりも海洋内部の変動が大きくて（年々とか10年規模とかエルニーニョとか）、顕著な温暖化傾向はほとんど見えていません。ただ、漁獲の傾向が変わっているという漁業者からの現場の声はあります。 2. 上の要因もあり、「毎年ここに行けば獲れる」という経験知が使えなくなっています。過去情報から環境要因をうまく抜き出すことにより、この漁場の変化にも対応した漁場推定技術が開発できるのではないかと考えてます。
20	fishtechによるサステイナブル漁業モデルの構築	飯山将晃	遠洋漁業と比べて、沿岸漁業の難しさはどの辺にあるのでしょうか？	遠洋と比べて沿岸の方が海況予測が難しい（その結果漁場予測も難しい）要因として、河川の流入まで計算する必要があること、海底地形の影響、遠洋よりもより高精細な予報が必要なこと、などが挙げられます。
21	fishtechによるサステイナブル漁業モデルの構築	飯山将晃	漁業と一口に言っても魚だけでなく、貝や海藻と言ったものもあります。その辺りは如何でしょうか？また、沿岸の表面水温となると海藻をターゲットにする方が魚よりも漁業者に喜ばれそうな気がします（いわゆる磯焼けの問題）。	貝・海藻の生産量は天然資源よりも養殖がはるかに多く、加速フェーズでのダイレクトな対象ではありませんが、今後対象には入っていくと思います。 磯焼けは要因が複雑で、河川とか、環境の変動とか、（護岸工事とかの）人為的な影響とか、生態系のバランスの変化とか、研究レベルの問題でもあり、海洋環境だけでどうこうできる範疇を超えています。赤潮も同様ですが、この問題は機械学習でも（教師データが少なく）難しい領域です。
22	AI技術を用いた法的文書作成支援	角田篤泰	対人構造（誰がどの権利を持っているか）などは単純な形態素解析や編集距離だけで参照しようとするとしばしば誤った結果が出るがあると思いますが、そのあたりはどのように解決されていますか。	対人構造の問題は、契約書のような具体的な個人や法人などの間での問題ですが、スモールフェーズでは例規が対象でしたので、この問題はありませんでした。また、対人構造の問題も、関係する法律関係（契約書など）が文書になっているのであれば、それらをまとめて1つの文書として統合することで当面は対応しようと考えています。
23	AI技術を用いた法的文書作成支援	角田篤泰	最近のBERTなどpretrainingの方法(Similarity Measure using Pretrained BERT Model)を使って、文書間距離はやってみるべきかと。	はい、現在実験中です。ただし、本研究では、意味内容以上に、字面上の多くの部分の一致が決定的に重要な応用ですので、少なくとも、ミクロな条文単位での文書の比較では、現行方式の方が優位だと予測しています。
24	オンデバイス学習技術とその応用	松谷 宏紀	オンデバイス学習（逐次学習）の場合に、ゆっくり徐々に変化していくような異常も検知できるのでしょうか？	動作中に「推論のみ」と「推論+逐次学習」をいつでも切り替え可能です。また、両者のアンサンブルも可能です。装置の設定を変えたとき、もしくは、部品を交換したときだけ一定期間逐次学習する（あとは推論のみ）という使い方もいらっしやいます。
25	オンデバイス学習技術とその応用	松谷 宏紀	The future of machine learning is tiny Pete Warden https://petewarden.com/2018/06/11/why-the-future-of-machine-learning-is-tiny/ ボタン電池でDNNを1年間動かすというもの。	この資料、面白いですね！参考にさせていただきます。
26	オンデバイス学習技術とその応用	松谷 宏紀	オンデバイスラーニングとはエッジ側の深層学習AIのことでしょうか？そうであれば、ニューロンの素子数はいくつぐらい確保できますか？かなり簡単な判断しかできないのではありませんか？	「エッジ側AI」と呼ばれている技術の多くは、推論に特化しています。一方、我々のオンデバイス学習はエッジで「逐次学習」します。この点が特殊です。ニューロン数は容易に増やせます。制約は、計算量もしくは面積コストです。
27	精神医学×メディア解析技術による心の病の定量化・早期発見と社会サービスの創出	佐藤真一	面談のテキスト自体から自然言語処理で判断するだけでなく、入力として音声そのもの(ピッチやトーン)などの情報も精神疾患診断に効きますでしょうか？直感的には内容のようですが、話し方にも内面を大きく表すように思います。	おっしゃる通り、対話のテキストだけではなく、ピッチやトーンなどの音響情報も含み解析を行うと有効であることは知られており、そうした研究や製品も存在します。一方、テキスト解析のみで判断できるとすると、アプリ化しやすいなどのメリットがあるかと考えています。
28	てんかん診断支援 脳波判読プラットフォーム	田中 聡久	臨床に使われている脳波ビューアーって世の中に何種類くらいあるのでしょうか？ 医局の数だけ種類があるのでしょうか？	基本的には脳波測定機器とセットと考えていただければ構いません。つまり、脳波計メーカーの数だけです。日本で主に使われている脳波計は2社程度のものです。
29	てんかん診断支援 脳波判読プラットフォーム	田中 聡久	多くの健常者の脳波を事前学習(pretraining)して、少ないてんかん脳波だけをアノテーションして識別するアプローチは使えない？	そのようなアプローチも考えていますが、今の所患者さんの脳波を使うのが近道と考えます。病歴の無い健常者の脳波を測定する手間とコストがかなり大きいからです。

第2回CREST「人工知能」研究領域 成果展開シンポジウム 質疑

項番	講演	発表者	質問	回答
30	てんかん診断支援 脳波判読プラットフォーム	田中 聡久	頭蓋内脳波をベースにしておられるようですが、非侵襲的な計測による脳波をベースにした場合の診断支援AIに取り組みられるご予定はありますか？簡易に日常環境下で計測できる脳波をもとに診断ができると、脳波データのビッグデータ化が期待できるのではないかと感じました。	前半にお見せした例は非侵襲計測です。簡易に日常計測できる脳波デバイスがあれば確かにビッグデータ化できますが、現在市場に出ているものは臨床に耐えうるクオリティの脳波を記録できません・・・
31	人工知能技術を用いた革新的アプタマー創薬システムの開発	浜田道昭	標的に結合しない配列情報は評価できないという事でしょうか？	SELEX実験は、標的に結合する配列を濃縮する実験となります。しかしながら、講演でお話した深層学習を用いた予測モデルにおいては、仮想的に配列に変異等を導入した場合にどのように活性が変化するかなども評価することが可能です。
32	3D画像認識AIによる革新的癌診断支援システムの構築	諸岡健一	実際のがん標本が少ないですね。世の中には、がん画像（立体）の生成器と正常・ガンの判定きのGANのアプローチはありそうに思うのですが、どうでしょう？	確かに1枚の画像を対象とした細胞の識別やGANはありますが、多重焦点画像列を対象としたものは我々オリジナルのアプローチです。