

FEATURE

AI分野でのオープンイノベーション推進

no data augmentation.

Args:

class_label = number of class labels

''' '''

```
def __init__(self, class_labels=2):
    super(myVGG_1, self).__init__()
    with self.init_scope():
        self.block1_1 = Block(64, 3)
        self.block1_2 = Block(64, 3)
        self.block2_1 = Block(128, 3)
        self.block2_2 = Block(128, 3)
        self.block3_1 = Block(256, 3)
        self.block3_2 = Block(256, 3)
        self.block3_3 = Block(256, 3)
        self.block4_1 = Block(512, 3)
        self.block4_2 = Block(512, 3)
        self.block4_3 = Block(512, 3)
        self.block5_1 = Block(512, 3)
        self.block5_2 = Block(512, 3)
        self.block5_3 = Block(512, 3)
        self.fc1 = L.Linear(None, 512, nobias=True)
        self.bn_fc1 = L.BatchNormalization(512)
        self.fc2 = L.Linear(None, 100, nobias=True)
        self.bn_fc2 = L.BatchNormalization(100)
        self.fc3 = L.Linear(None, class_labels, nobias=True) #added lay

def __call__(self, x, t=None):
    # 64 channel blocks: 32 to 16
    h = self.block1_1(x)
    h = F.dropout(h, ratio=0.3)
    h = self.block1_2(h)
    h = F.max_pooling_2d(h, ksize=2, stride=2)

    # 128 channel blocks: 16 to 8
    h = self.block2_1(h)
    h = F.dropout(h, ratio=0.4)
    h = self.block2_2(h)
    h = F.max_pooling_2d(h, ksize=2, stride=2)

    # 256 channel blocks: 8 to 4
    h = self.block3_1(h)
    h = F.dropout(h, ratio=0.4)
    h = self.block3_2(h)
    h = F.dropout(h, ratio=0.4)
```

CONTENTS

FEATURE

AI分野でのオープンイノベーション推進

1 INTERVIEW

NICTが培ってきた先端のAI技術を社会へ

佐伯 宜昭

4 社会的課題解決とオープンイノベーションのための
AI技術の研究開発連携の取組

岩爪 道昭

6 サイバーセキュリティ対策の自動化に向けた
機械学習技術の活用

高橋 健志

8 AI技術を脳情報データベースに結び付ける研究基盤の構築
認知機能検査ソフトウェアの社会実装へ向けて

下川 哲也

10 「翻訳バンク」を通じたAIデータテストベッドの構築

足田 啓太

TOPICS

12 NICTの知的財産⑦

「長期的なプライバシー保護とビッグデータ利活用の両立」
が可能な暗号化方式 (サーバ、サービス方法)

—世代を超えた将来までプライバシー保護を可能とする世界初の準同型暗号方式—

13 NICT発のスマホ・タブレットアプリ&翻訳Webサイト

INFORMATION

14 平成31年度 採用情報

14 情報バリアフリーのための情報提供サイト



表紙写真「深層学習のプログラム」 左上写真「AI研究開発用計算機」

表紙写真は、画像認識（画像を見て何の画像か判断する）用の機械学習プログラムの一例。ヒトの脳（特に大脳皮質）の神経ネットワークをヒントに開発されたもので、ニューラルネットワークと呼ばれている。情報を処理する層が何段にも深く重ねられているため、深層学習（ディープニューラルネットワーク）とも言われ、画像や音声の識別では、人間と同等以上の能力を獲得できるようになり近年注目されている。

左上写真は、AI研究開発用計算機の一部。最近のAIの躍進には、良質かつ大量のデータを蓄えるストレージとそれらを高速に解析する演算装置など、ハードウェアの進展によるところも少なくない。

INTERVIEW

NICTが培ってきた先端のAI技術を社会へ



佐伯 宜昭 (さえき たかあき)

知能科学融合研究開発推進センター
副研究開発推進センター長2016年 オープンイノベーション推進本部
ソーシャルイノベーションユニット戦略
的プログラムオフィス統括。2017年4月
から現職。

情報通信の研究開発には多くの方向性があるが、情報の送り手／受け手である人間の支援や、その機能の解析において、様々な場面で重要性を増しているのが、AI（人工知能）の存在である。2017年4月1日に発足した知能科学融合研究開発推進センター（AIS）は、自然言語処理、知識処理、脳情報通信融合研究など、これまでNICTで進められてきたAIに関わる研究開発の知見を集め、社会実装を進めるための組織である。その設立の経緯、目指すものなどについて、副研究開発推進センター長である佐伯宜昭氏にお話を伺った。

■ AI分野のオープンイノベーション拠点

——2017年度、AISが設立されることになった経緯について教えてください。

佐伯 背景としては、大きな2つの流れがあります。ひとつは、2016年4月から始まった、NICTの第4期中長期計画の中で掲げられている、オープンイノベーション機能の強化との関わりです。中長期計画ではNICTの研究分野を「センシング基盤分野」「統合ICT基盤分野」「データ利活用基盤分野」「サイバーセキュリティ分野」「フロンティア研究分野」と、大きく5つにカテゴライズしていますが、AI関連はそのひとつ、データ利活用基盤分野のカテゴリーとして取り上げられており、この分野におけるオープンイノベーション推進拠点が必要であると認識されたわけです。

もうひとつは、国としてもAIの研究開発と産業化は大きな課題であるとして、2016年4月、総務省、文部科学省、経済産業省の3省が中心となり、総理大臣指示を受けた「人工知能技術戦略会議」が立ち上げられています。戦略会議においては、総務省所管のNICT、文科省所管の理化学研究所（理研）、経産省所管の産業技術総

合研究所（産総研）も含めて、これからのAI技術の研究開発目標と産業化のロードマップが議論され、2017年3月末に「人工知能技術戦略」がまとめられています。この戦略を進めていくにあたり、NICTとしても外部との連携を図りまとめていく核となる組織が必要でした。

この2つの流れが合流するかたちとなり、AIS設立に至ったのです。

■ 「連携研究室」「連携推進室」の両輪体制

——AI技術の社会実装に向けて、具体的にどのような活動を行っているのでしょうか。

佐伯 実際に研究を行っている各部門により、研究内容に応じて社会実装への取り組み方は若干の差異があると思っています。

例えば自動翻訳に携わっているグループは、元々、総務省とともに多言語音声翻訳システムの社会実装を目指す「グローバルコミュニケーション計画」を進めています。これは短期的には、2020年の東京オリンピック・パラリンピック競技大会に向け、ある程度の社会実装を達成するというのが、ひとつの大きな目標となっています。そのため、多くの企業と組んで、既に様々なツールなどの開発も行われています。

しかし、より基礎研究に近い部門では、まだ、社会実装に向けての動きがあまり具体的でないところもあります。我々AISの活動の中心となると考えているのは、それらの研究が社会にどれだけのインパクトをもたらす可能性があるかを測り、関心を持ちそうな外部の研究機関や民間企業との仲立ちをしていくことです。

部門ごとに必要なサポートの内容も違ってくると思いますので、今、様々な研究部門と意見交換を重ねているところです。AISの中には民間企業出身の人もいますか

INTERVIEW

NICTが培ってきた先端のAI技術を社会へ

ら、そうした人々の知見も活かしつつ、「このような活用も考えられるのでは」などと、こちらからの提案も行っています。

——そうした活動を進めるにあたっての、現時点でのAISの体制を教えてください。

佐伯 現在AISの研究開発推進センター長は、徳田理事長が兼務しているのですが、その下に私を含め2人の副研究開発推進センター長と、連携研究室、連携推進室という2つのセクションを擁しています。

連携研究室のスタッフは、小金井の本部に常駐する専任の研究者のほか、NICT内でAI関連の研究開発に携わる各拠点／部門の研究者のうち、データの取得や研究成果の利活用などで、オープンイノベーションへの関わりがより深い方々に兼務しても

らって、本部と共同でプロジェクトを行っています。

連携推進室は連携研究室の活動をバックアップするとともに、「人工知能技術戦略会議」、特に産総研や理研におけるカウンターパートとなる部門との連携協議も行っています。例えば2017年5月には、産総研との間で「情報通信分野における連携・協力の推進に関する協定」を締結し、共同研究「特許文献専用のニューラル機械翻訳とそれを可能とするシステム構築に関する研究」を開始したほか、人工知能技術戦略会議や総務省とともに「第2回 次世代の人工知能技術に関する合同シンポジウム」を主催しました。こうした外部との連携に関する合意書（MOU）の締結やAIに関するイベントの企画なども推進室で担当しています。

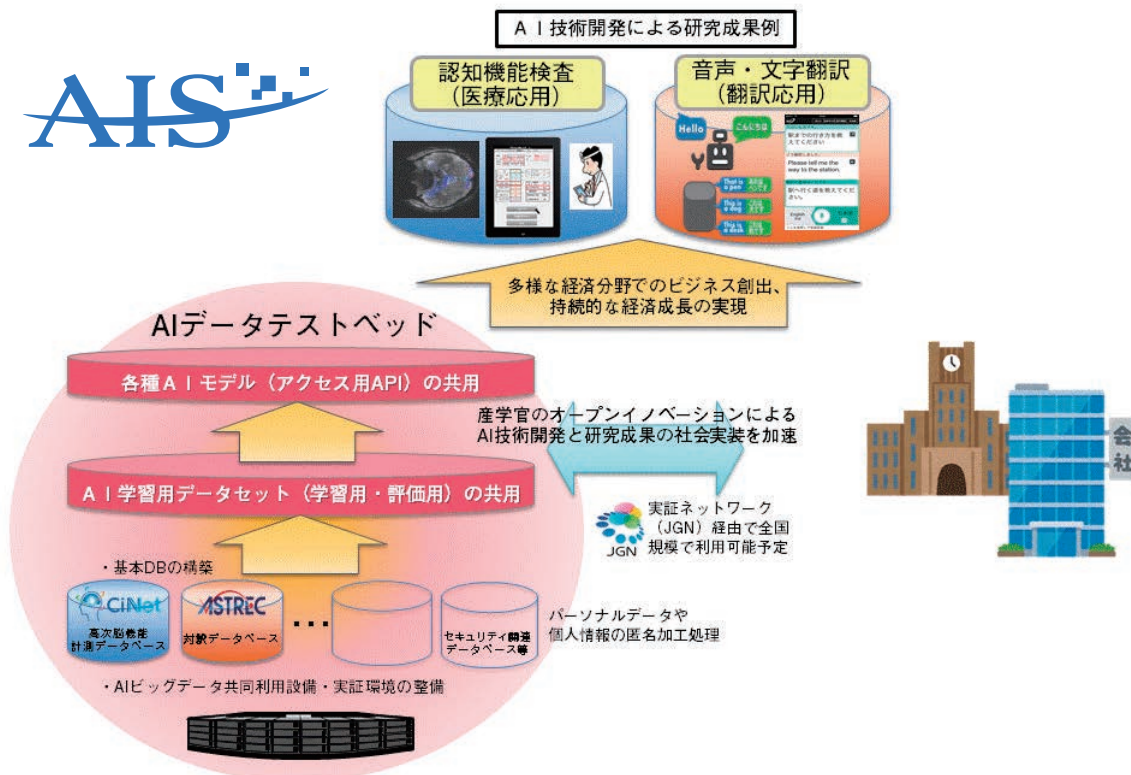
■外部との連携・共有でwin-winの関係を

——AISの具体的な活動として、2017年9月に運用が開始された「翻訳バンク」が挙げられると思いますが、これはどういう位置付けのものなのでしょうか。

佐伯 特に昨今、AIの研究開発においては、計算機資源の整備やアルゴリズムの開発に加えて、基となるデータの集積が重要であると言われています。そこで「AIデータテストベッド」の名称のもと、計算機資源だけでなく、AIに利用できるデータもしっかり集め、整備しようという計画が立てられました。

AIS設立のもうひとつのミッションとして、こうしたAIに関するデータを外部と

『知能科学融合研究開発推進センター(AIS)』のイメージ





AIS主体で運営された「次世代の人工知能技術に関する合同シンポジウム」会場内（2017年5月22日）



閉会の挨拶をされた徳田NICT理事長はAIS研究開発推進センター長を兼任



第2回の今回は大阪大学が会場となった

共有しつつ、それらデータを使ってどうすることが可能かを外部と連携して探っていくことが挙げられます。「翻訳バンク」は、その活動の一環ということになります。

特に機械翻訳においては、翻訳データの量が翻訳精度に密接に関係してくることは研究から分かっていますが、データを買集めるとなると大きなコストがかかります。しかし一方で、民間企業などでは、技術資料などの対訳データが意外に死蔵されているのではないかと推測もあり、そうした埋もれた文書をサルベージし、リサイクルして、翻訳エンジンの精度向上に役立てようというのが「翻訳バンク」の役割です。データの蓄積によってAIの研究にも役立ちますし、社会実装が進んでいけばデータを提供していただいた民間会社などにも恩恵がある。win-winの関係を作れるのではないかと発想で始めているものです。

——これも含め、今後はいろいろと面白い動きが増えていきそうですね。

佐伯 AIS自体は発足して間もなく、「翻訳バンク」を含め、個々の具体的な動きはスモール・スタートですが、最終的には、AIに関わる様々な分野で「データが回るエコシステム」的な動きを巻き起こしていければいいと思っています。

例えば、脳情報通信融合研究センター（CiNet）の手掛ける脳情報データについても、匿名加工処理などを行ったうえでこういったモデルで公開し、社会実装を進めていくか——実際にストーリーを描いたうえで、民間企業と一緒に何かできることはないかなどを探っています。自動翻訳や脳情報以外にも、NICTの持っているAI関連のデータ全般に、こうした動きを広げていきたいですね。

昨今のAIブームの中、様々な機関や企業でその研究開発が進んでいますが、NICTにおいても、特定の分野では世界一の研究と自負できるものも複数あります。そうした研究について内外に認知度を高め、民間企業や他の研究機関に対しても、「こんな分野で社会に実装されていく可能

性がありますよ」「一緒にやっていくことで、これだけ新たな道が拓けてきますよ」と、結びつきを強めていく。AISのミッションとして、そのような役割を果たす組織を実現していきたいと考えています。



社会的課題解決とオープンイノベーションのための AI技術の研究開発連携の取組



岩爪 道昭 (いわづめ みちあき)
知能科学融合研究開発推進センター
連携研究室
室長

大学院修了後、日本学術振興会特別研究員、近畿大学理工学部助手、理化学研究所脳科学総合研究センター研究員を経て、2005年NICT入所。2017年から現職。知識ベースシステム、Web情報処理などの研究に従事。博士（工学）。

知 能科学融合研究開発推進センター（AIS）連携研究室では、NICTが強みとする機械学習技術等の人工知能（AI）技術と脳情報通信技術、セキュリティ技術、リモートセンシング技術等とそれらに由来するビッグデータを利活用し、社会的解決型のAI研究開発を内外の研究機関と連携しながら進めています。

■ 背景

最近、メディアでは、AIスピーカー、自動運転、ロボット、AI将棋などAIの話題を耳にしない日はないほど高い関心が集まっています。この背景のひとつとしては、ディープラーニング（Deep Learning、深層学習）と呼ばれる多層のニューラルネットワークを用いた機械学習技術が、画像認識や音声認識等の分野で従来手法を凌駕する性能を示すようになったことが発端になっています。多層ニューラルネットワ

クの研究は、1960年代には始まっており新しい概念ではありませんが、当時のコンピュータの処理能力では、膨大な時間を要することに加え、多層にすると学習がうまく進まなくなるという技術的問題もあり、現実的ではありませんでした。しかし、近年のICTの発展・普及により膨大なデータを蓄積し、高速に解析する計算機環境が整い、学習方法の改良により数10層から100層を超える多層ニューラルネットワークを学習できるようになってきました。

■ NICTにおけるAI技術の研究開発戦略

現在の機械学習技術で賢いAIシステムを作るには、品質の高い大量のデータが必要になります。ここでいう品質の高さとは、解きたい課題に合致した正解／不正解等のラベルが付与された学習データセットを意味します（図1）。手書き文字認識や一般物体認識のための画像データセット等

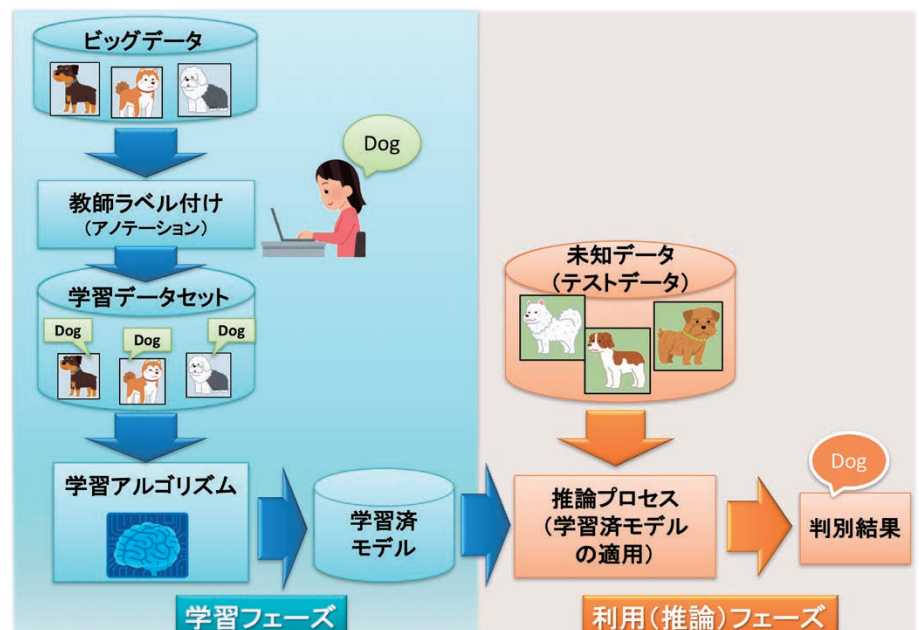


図1 AIシステム開発の流れ

は、インターネット上で公開・共有されて、多くの研究者が利用できるようになっていますが、社会的課題に役立つようなデータセットは、公開されていないか、データそのものが集積化されていないケースも少なくありません。このようなデータセットを整備するためには、現在の技術では、人手でラベルを付与する必要があるため、一般的に多大なコストがかかります。

一方、NICTでは、現在のようなAIブームが到来するはるか前から、データの重要性を認識し、多言語コーパスや大規模Webアーカイブを自ら構築し、音声認識技術、機械翻訳技術、社会知解析技術等、先進的なAI技術の研究開発と社会実装を進めてきました。また、少ない経験からも学ぶことができ、かつ低消費電力なヒトの脳の機能を解明し、次世代の情報通信技術に応用するため、fMRI、脳波計等の高次脳情報のデータベース化にも取り組んでいます。さらに、セキュリティ、リモートセンシング、有無線通信等、ICT基盤技術の開発のために様々なデータを計測・収集しています（図2）。その中には、NICTでしか計測できない貴重なデータも含まれています。我々は、NICTの各研究拠点と連携しながら、NICTが強みとするデータ群に対して機械学習等のAI技術を適用し、実証システムの研究開発、さらには社会実装を加速するため、最先端AIデータテストベッドの整備・利活用を進めています。

■今後の展望

今後AI技術はどのように進展していくのでしょうか？現在のAIは、音声認識、画像の物体認識など単一機能の精度が飛躍的に向上した段階です。生物の系統的進化に例えると、外界の情報を知覚する目や耳などの感覚器を獲得し、認識・識別のための脳が発達し始めた段階と言えるでしょ

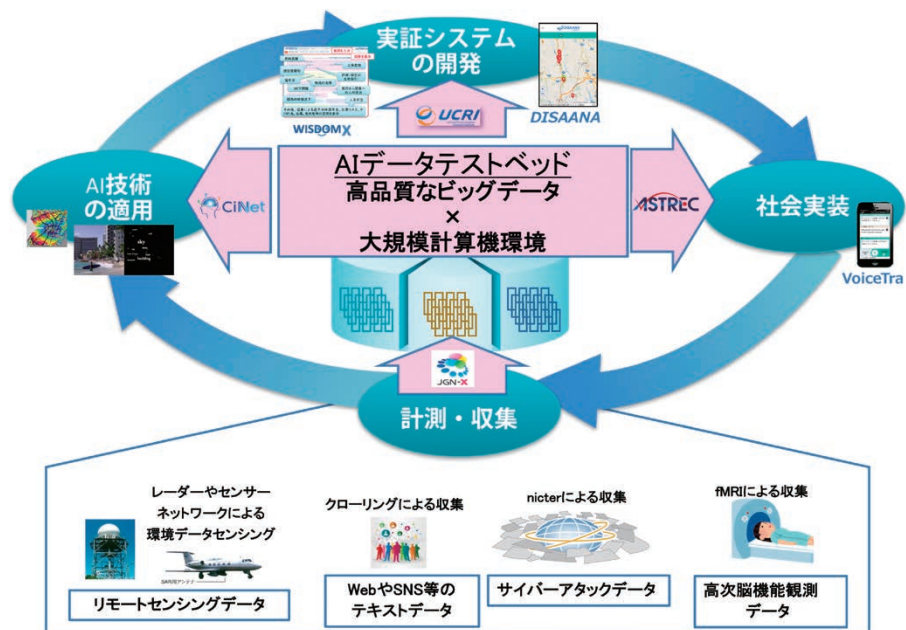


図2 NICTにおけるAI技術の研究開発戦略

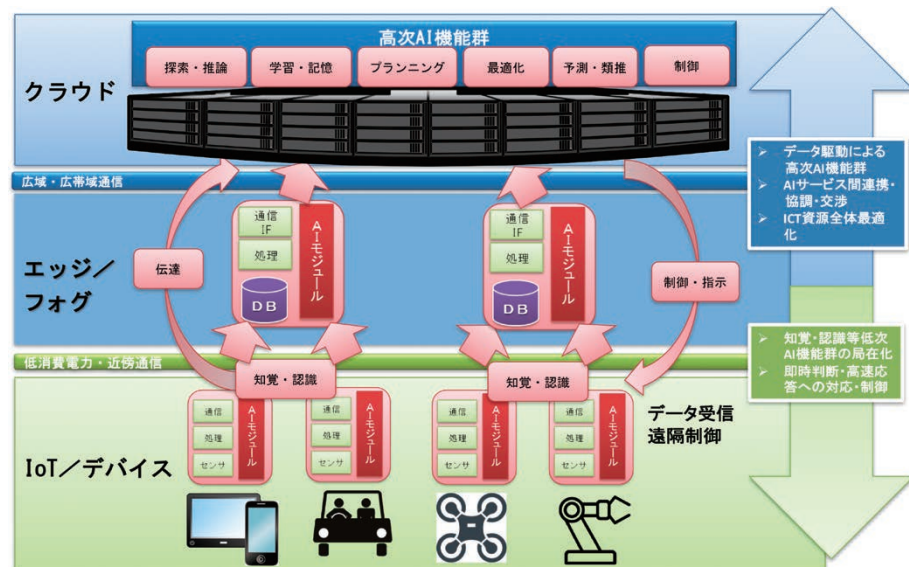


図3 AIとICTの進展イメージ

う。IoTデバイスから膨大な情報が刻一刻と送られてくるようになったとき、ネットワークの通信コストや、レスポンスタイム等の制約からクラウドによる集中処理だけでは対応が困難になると思われます。その対策として、IoTに知覚・認識の機能が埋め込まれ、一次処理を行ったうえで、分節化された情報がエッジ/フォグ側（ネットワークの端末やセンサに近いサーバ/端末

とクラウドの中間的に位置するサーバ）に受け渡され、推論、最適化、制御等のより高次処理が施されるような機能の分化が進むと予想されます（図3）。そのときICT全体のアーキテクチャも変容してくるでしょう。具体的な社会的課題と向き合いながら、人とAIが共生するICTのあるべき姿についても、探究していきたいと考えています。

サイバーセキュリティ対策の自動化に向けた機械学習技術の活用



高橋 健志 (たかはし けんし)

サイバーセキュリティ研究所サイバーセキュリティ研究室 研究マネージャー
／知能科学融合研究開発推進センター
連携研究室主任研究員兼務

タンペレ工科大学、株式会社ローランド・ベルガーを経て、2009年NICT入所。サイバーセキュリティ技術に関する研究開発と標準化活動に従事。博士(国際情報通信学)。

ここ数年、ランサムウェアやIoT機器のボットネット化など、様々なサイバー脅威がインターネット上で猛威を振るっています。セキュリティ人材不足の中でその状況に対応するには、人材育成に加えてセキュリティオペレーションの自動化を推進することが必要になります。我々は、機械学習を中心とした人工知能（AI）技術を活用することにより、その自動化を積極的に推進していきたいと考えています。例えば、大量のマルウェア（不正プログラム）の分類やその機能の類推、インターネット上のマルウェア活動の傾向変化の検知や組織内通信の異常検知、さらには将来の攻撃予測など、様々な技術革新を目指して研究開発に取り組んでいます。

元々、AI技術とサイバーセキュリティ技術は異なるコミュニティで発展してきた経緯があるため、両領域の技術の効果的な融合には時間と労力を要します。ですが、その両者を幅広い領域で融合させることができれば、セキュリティの自動化は大きく進展します。2017年度NICT内に設立された知能科学融合研究開発推進センター（AIS）はその融合を加速し、双方の領域に精通した人材を発掘・育成し、サイバーセキュリティ技術を進化させるのに重要な役割を果たします。

表 NICTが保有するデータ例

カテゴリ	データセットの具体例
ダークネット 関連データ	未使用IPアドレスへの攻撃関連通信データ。pcapファイル、統計情報、悪性ホスト情報、など。
ライブネット 関連データ	NICT内の通信データ。pcapファイル、フローデータ、セキュリティアラートが生成するアラート、など。
マルウェア 関連データ	マルウェア検体、静的・動的解析結果、など。
スパムメール 関連データ	スパム（ダブルバウンス）メールデータ、統計情報、など。
Android APK 関連データ	APKファイル、マーケットのアプリ情報、など。
ブログ・記事	TwitterのTweet、セキュリティベンダーブログ、など。

研究競争力の源泉となるデータセット

サイバーセキュリティ分野において、データセット（蓄積データ）は研究開発の競争力の源泉となります。機械学習は大量のデータを学習・分析する技術ですので、どれだけ多く、良質なデータを収集・蓄積できるかが非常に重要になります。表のとおり、我々は様々なデータを蓄積しており、例えばダークネット（未使用IPアドレス）トラフィックの観測データについては約

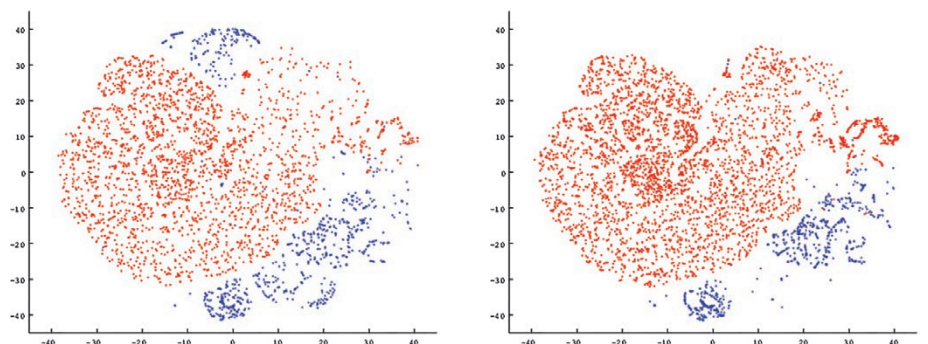


図1 t-SNEにより2次元化した特徴ベクトル。左図は実験開始月、右図はその翌月の状況であり、赤はDDoS攻撃に関連する特徴ベクトルを、青はそれ以外を示している。

10年間、収集・分析・蓄積を継続してきています。

■ 機械学習の活用事例

我々は上述の収集データを活用し、機械学習を用いた分析・自動化技術を検討しています。ここで、そのいくつかを紹介します。

パッカーの特定—多くのマルウェアは、分析を難しくするため、パッカーと呼ばれる難読化ツール（実行形式ファイルをその機能を損なうことなく暗号化するツール）を利用しています。そのためマルウェアの分析をする際には、どのパッカーが利用されているのかを特定する必要があります。我々は、パッカーの復号ルーチンのバイナリ列をサポートベクタマシン（SVM）を用いて分類することで、パッカーの特定を自動化し、99.46%の精度での特定を実現しています^[1]。

DDoSの発生検知—インターネット上のDDoS攻撃（分散型サービス不能攻撃）発生を早期検知するのにも機械学習を活用しています。ダークネットではボックスキャタ*が観測されますが、それを分析することで、DDoSの発生を早期に検知できます。具体的には、ダークネットトラフィックについて、IPアドレスごとに17種類の特徴を定め、その特徴に基づいて機械学習による分類処理を実施することにより、DDoSに関係するボックスキャタか否かを判定する処理を実装しています。なお、インターネット上の実トラフィックを分析する際には、その傾向が刻一刻と変化していくことを考慮し、学習を追加で繰り返していく必要があります。図1は、実験開始月（左図）およびその翌月（右図）の特徴ベクトルをt-SNEと呼ばれる次元圧縮アルゴリズムを用いて2次元化した結果を示しています。この図のうちの赤い部分がDDoS攻撃に関連する特徴ベクトルをプロットしたもので、

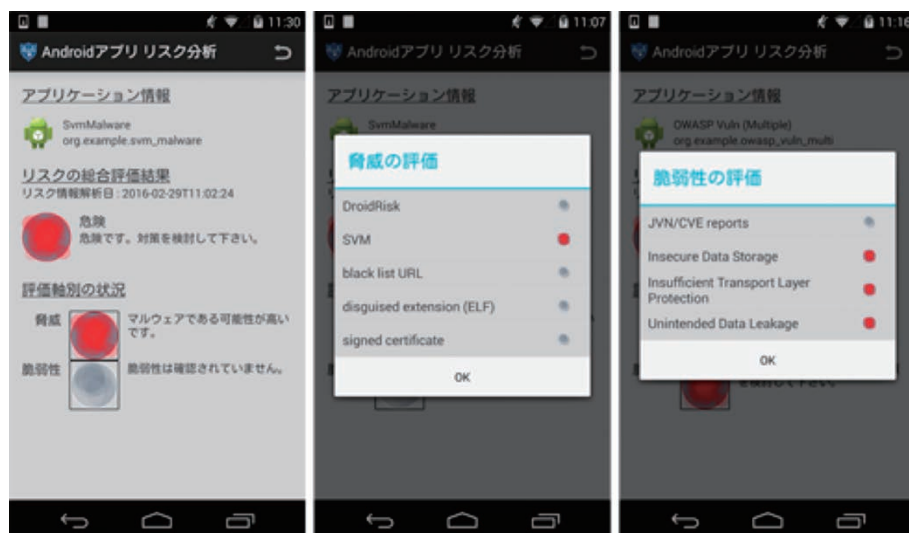


図2 本ツールは脅威と脆弱性の両面からリスクを評価する。左図が総合評価、中央図が脅威の評価、右図が脆弱性の評価を示している。

それ以外が青い部分です。この赤と青の領域が少しずつ変化している様子が分かります^[2]。

Androidアプリ分析—AndroidアプリはGoogle Playなどの各種オンラインマーケットにて配布されていますが、これらの中にはマルウェアが混在しています。そのため、アプリがマルウェアであるかどうかを判定する必要があります。我々は、Androidアプリから特徴抽出し、同時にオンラインマーケットに存在しているアプリ説明文とカテゴリ情報を収集し、その特徴と収集情報をSVMに入力することで、マルウェアか否かを94.07%の精度で判定できるAndroidアプリ分析技術を開発しました^[3]。図2のとおり、そのマルウェアか否かの判定結果を可視化するツールも作成しています。本ツールでは、脅威と脆弱性の両視点からアラート（赤信号）の有無を可視化していますが、その中の脅威情報の分析にもSVMを利用しています。

■ 今後の課題と当面の方向性

セキュリティ分野での機械学習の適用範囲は今後急速に拡大していくことが予想さ

れます。我々は、競争力の源泉となるデータを継続的に収集・蓄積していくのと同時に、解決すべき課題を再定義し、それに適用可能な機械学習技術自体の高度化にも取り組んでいきます。また、機械学習技術を実際のセキュリティオペレーションの現場に適用するためには、リアルタイム性の担保など、より一層の技術革新が必要です。東京2020オリンピック・パラリンピック競技大会に向けて、セキュリティオペレーションの自動化を進めるべく、研究開発を加速していきたいと考えています。

References

- [1] R. Isawa et al., "An Accurate Packer Identification Method using Support Vector Machine," IEICE Trans. Fundamentals, Vol.E97-A, No.1, pp.253-263, 2014.
- [2] I. Skrjanc et al., "Evolving Cauchy Possibilistic Clustering and Its Application to Large-Scale Cyberattack Monitoring," IEEE Symposium Series on Computational Intelligence, Hawaii, 2017.
- [3] T. Takahashi et al., "The Usability of Metadata for Android Application Analysis," The 23rd International Conference on Neural Information Processing, Kyoto, 2016.

* 送信元IPアドレスが詐称されたDoS攻撃（SYN-flood攻撃）を受けているサーバからの応答（SYN-ACK）パケットのこと

AI技術を脳情報データベースに結び付ける研究基盤の構築

認知機能検査ソフトウェアの社会実装へ向けて



下川 哲也 (しもかわ てつや)

知能科学融合研究開発推進センター
連携研究室主任研究員／脳情報通信
融合研究センター脳情報工学研究室
主任研究員兼務

大学院修了後、大阪大学助手、特任准教授を経て、2010年NICTに入所。ヒトの脳のネットワーク解析に関する研究に従事。博士（工学）。

かつてのAI研究では、人間の専門家（エキスパート）の意思決定方法を定式化したエキスパートシステム開発が盛んで、医療応用なども検討されました。しかし、専門家の意思決定が多分に主観的なため、専門家の知識を定式化することが大変すぎて衰退していった歴史があります。今回の我々の試みは、医者の主観的な意思決定だけでなく、検査結果などの客観的な医学データそのものから情報を抽出し、医学の専門家と情報工学の専門家がタッグを組んでAIに使えるデータベースをきちんと作る、というのがポイントとなります。本稿では、知能科学融合研究開発センター（AIS）のミッションである具体的な社会実装へ向けた試みとともに、精神疾患の脳情報データベースとAI技術を結び付ける研究基盤についてご紹介します。

■背景

国立研究開発法人情報通信研究機構（NICT）脳情報通信融合研究センター（CiNet）では、脳科学と情報通信技術の融合を目指して研究を進めてきました。そして2017年4月にAISが設立されました。

なぜ情報通信の研究機関が脳を研究して

いるのか？と不思議に思われる方もおられるかもしれませんが。実は脳というシステムは、10の10乗個もの神経細胞が集まっており、その複雑さは現存するどの人工物も太刀打ちできないほどです。その上、消費エネルギーは大体10～20ワットとされています。そこで、脳に触発された新しい情報通信技術の開発を目指すべく、脳の研究、特に脳ネットワークの研究を進めてきました。

■データベースコアと解析コアを結ぶAISの役割

我々は、これまでCiNetで蓄積してきた脳情報データや解析技術に加えて、大阪大学大学院連合小児発達研究科の橋本亮太准教授（2017年10月からAISに招へい）が代表を務める研究グループCOCORO^{*1}の所有する膨大な統合失調症関連の脳情報データベース構築を進めています。

今回のような、医学の分野と情報工学の分野という異なる2つの分野が融合して研究を進める場合、必ずある問題が顕在化します。それを表したのが図1です。医学部を中心とした医療機関を「データベースコ

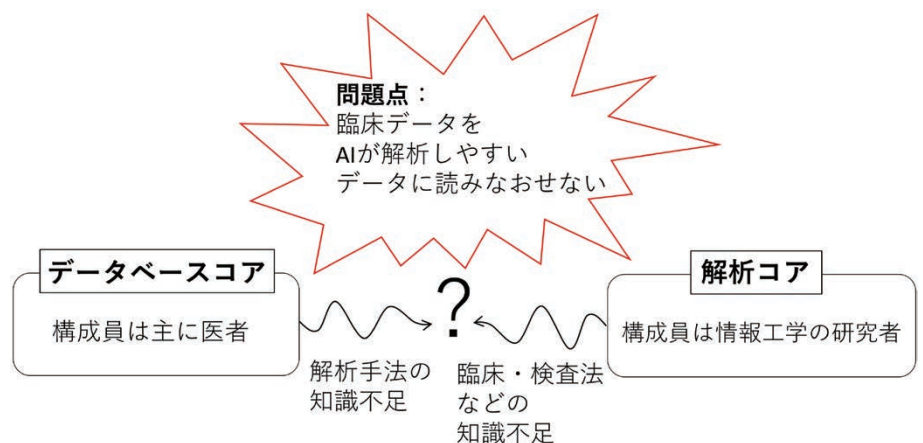


図1 医学分野（データベースコア）と情報工学分野（解析コア）が融合研究を進める場合に顕在化する問題

*1 COCORO (Cognitive Genetics Collaborative Research Organization- 認知ゲノム共同研究機構)、
<http://www.sp-web.sakura.ne.jp/lab/cocoro.html>

*2 Sumiyoshi et al., Psychiatry Research, vol. 245, pp. 371-378, 2016.

*3 伊藤弘人、福田敬、岩成秀夫、西田淳志、奥村泰之（2011）「平成22年度障害者総合福祉推進事業（精神疾患の社会的コストの推計）報告書」、学校法人順天堂、pp.1-39.

*4 NICT委託研究 未来を創る新たなネットワーク基盤技術に関する研究開発（平成28～32年度）「時空間ダイナミクスの記述を可能とする拡張Bow-Tie構造に基づく進化発展可能なネットワーク化情報処理基盤アーキテクチャの創出」（受託者：大阪大学）

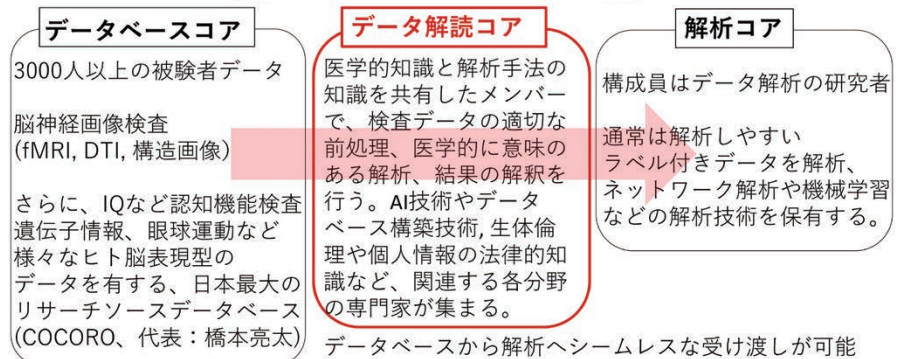


図2 データベースコアと解析コアを結ぶ「データ解読コア」AISの役割

ア」、そしてネットワーク解析やAI解析の知識を有する研究機関を「解析コア」とします。データベースコア側は解析手法の知識が不足しており、逆に解析コア側は臨床や検査方法などの知識が不足していることから、なかなか臨床データを解析しやすいデータに読み直せないという問題が生じるのです。

そこで今回、AISの設立とともに、この2つのコアをつなぐ「データ解読コア」として、AISを中心としたCiNet、大阪大学の三者による共同研究グループを作っていくことになりました(図2)。

■認知機能障害計測のソフトウェア開発

では具体的にこの3つのコアで構成されるプロジェクトでどのような研究が可能となるのか、具体的に出来つつある成果の例として「認知機能障害計測のソフトウェア開発」を挙げます。我々は主にfMRI装置を用いて脳機能画像を計測し、解析を行っています。その際に重要となるのが被験者自身の属性です。もちろん性別や年齢もありますが、それ以上に臨床的にも重要になるのがIQなどの認知機能レベルの検査です。この認知機能レベルの検査は代表的なWAIS-IIIという方法などを用いると、検査に2時間も要するのに対し、大阪大学の橋本准教授らのグループでは、これまで計測した精神疾患患者・健常者合わせて400人近くのデータを基に、どの検査項目が必須なのかを絞り込み、検査時間を15分程度に縮めることに成功しました^{*2}。現在AISでは、機械学習などのAI技術を用いて更なる解析を行い、より精度の高い、認知機能レベルの短縮版検査方法の開発を進めつつ、最終的な社会実装として、タブレットで実行可能なアプリケーション開発を進めています(図3)。

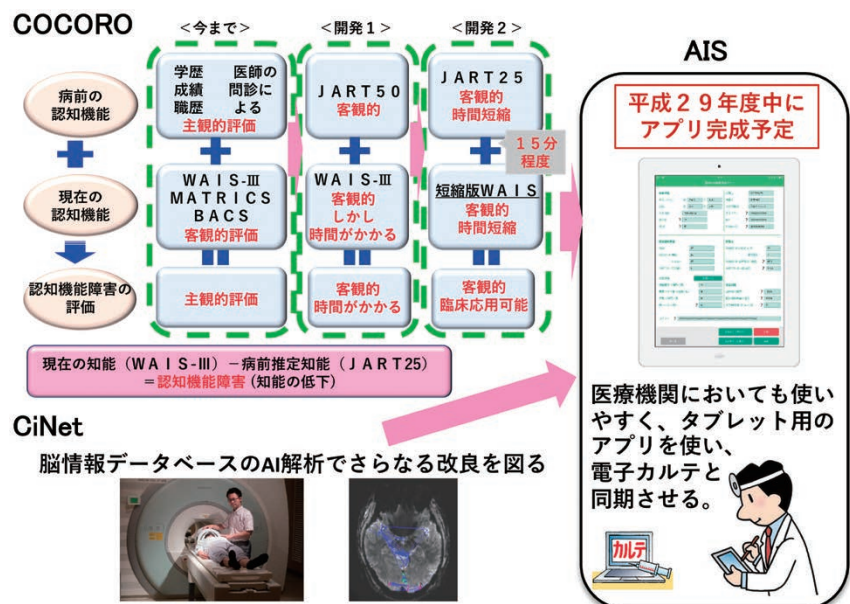


図3 検査データベース解析に基づく短縮版認知機能障害計測のソフトウェア開発

■今後の展望

厚生労働省の調査によると我が国の精神疾患患者の総数は200万人以上と推定され、その社会的コストは8兆円以上と推計されています^{*3}。我々の取組は、医学的知見とICTを融合することで、従来、早期発見・診断が困難であった精神疾患の脳バイオマーカーの発見を支援し、新たな診断・治療法の確立、創薬に貢献することが期待されます。

本研究は医学的重要性だけにとどまりません。情報通信技術開発においても大きな期待が寄せられています。脳が脳としてあ

るために何が必要なのかを知るには、健常者の例だけでなく、何らかの理由で脳機能に疾患が生じてしまった例を見ることによって、初めてわかります。現在、NICTの委託研究^{*4}と連携して、大阪大学大学院情報科学研究科の研究グループと定期的に情報交換をしつつ、脳情報データベースの解析結果の抽象モデル化、そして情報通信技術への応用を目指して研究を進めています。今後もこうした医療分野や情報通信の分野への応用、社会実装を更に進めるべく、健常者と患者の脳情報データベース構築と研究体制の更なる整備、そして研究開発を引き続き行ってまいります。

「翻訳バンク」を通じたAIデータテストベッドの構築



疋田 啓太 (ひきた けいた)
知能科学融合研究開発推進センター
連携推進室
主任

法科大学院修了後、2011年NICT入所。知的財産推進室、法務・コンプライアンス室、総務省出向等を経て現職。法務博士。

AIS（知能科学融合研究開発推進センター）では、AIデータテストベッドの構築の一環として、ASTREC（先進的音声翻訳研究開発推進センター）と共に「翻訳バンク」を運用しています。本稿では、「翻訳バンク」の運用を開始した経緯、取組の内容及び今後の展望を紹介します。

■ AIデータテストベッド構築に向けた取組

AISでは、AI技術に関するオープンイノベーション型の戦略的な研究開発を推進するため、AIデータテストベッドと称してAIに利用できるデータの集積に取り組んでいます。ここでは、その取組のひとつである自動翻訳技術の研究開発に利用できる翻訳データを集積する活動「翻訳バンク」について説明します。

■ 自動翻訳技術の現状

NICTでは、2020年までに多言語音声翻訳技術の社会実装を目指すグローバルコミュニケーション計画^{*1}の下、AI技術で多用される深い階層構造を持つニューラルネットワークを用いた自動翻訳技術（以下、ニューラル翻訳）の研究開発に取り組んで

います。ニューラル翻訳による自動翻訳の精度向上のためには、NICTが研究開発において取り組んでいるニューラルネットワークの改良に加え、翻訳データが多ければ多いほど翻訳精度が向上することがわかっていることから、翻訳データの量の確保が重要です（図1参照）。例えば、NICTでは特許文献の自動翻訳に関して特許庁と協力^{*2}しており、翻訳データの提供を受けています。ニューラル翻訳では翻訳データが多い分野の翻訳精度が高くなるため、その翻訳データによって改良されたニューラル翻訳エンジン（「みんなの自動翻訳」webサイトで公開中^{*3}）は、特許やマニュアルなど長文を正確に翻訳する精度が特に高いものとなっております。

■ 翻訳データの不足

翻訳データとは、様々な言語による原文とそれが翻訳された訳文の対を集めたものです。現在NICTは、特許庁をはじめ、地方自治体、企業、各種団体などからも翻訳データの提供を受けていますが、NICTが研究開発に利用している翻訳データは主にwebページのクロールによって集めたものです。

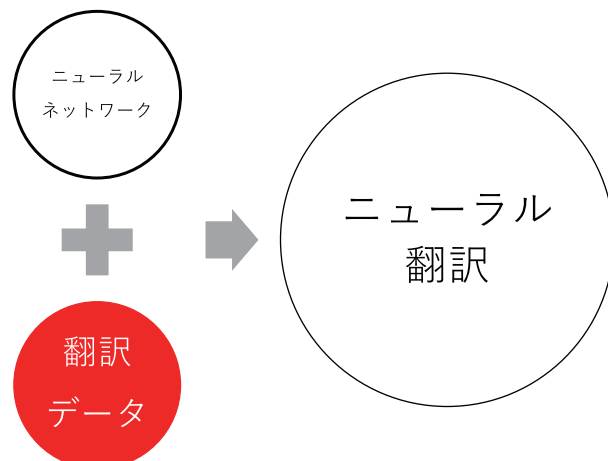


図1 ニューラル翻訳のイメージ

*1 http://www.soumu.go.jp/main_content/000285578.pdf

*2 <https://www.nict.go.jp/info/topics/2016/04/160401-1.html>

*3 <https://mt-auto-minhon-mlt.ucrj.gn-x.jp>

*4 <https://www.nict.go.jp/press/2017/09/08-1.html>

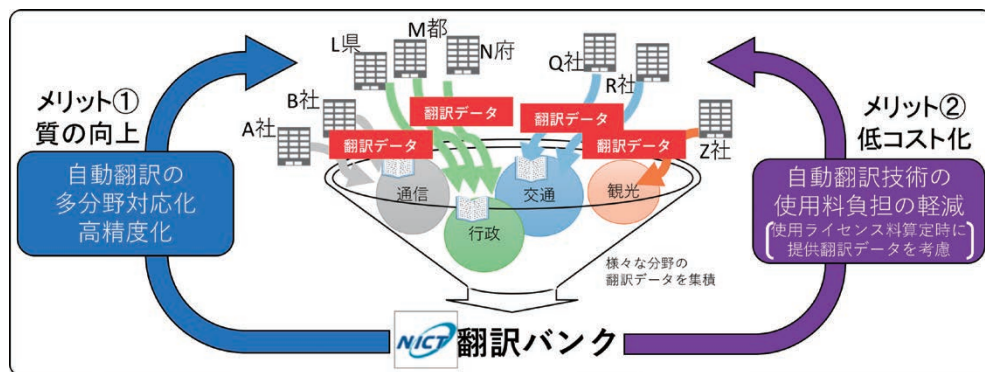


図2 翻訳バンクのコンセプト



図3 自動翻訳のエコシステム

しかし、web ページに掲載されている翻訳データは社会にある翻訳データのごく一部に過ぎません。企業などには技術資料やマニュアルなどの翻訳データが眠っているのではないのでしょうか。今後日本の自動翻訳技術が世界に対抗していくためには現在の翻訳データ量では足りないため、様々な分野の翻訳データを更に集積しなければなりません。それには翻訳データを積極的に提供いただける企業などを増やすことが必要です。また、そもそも自動翻訳の精度向上には翻訳データが必要だということ自体を、社会に周知することも必要です。

■「翻訳バンク」の運用開始

そこで、NICTは総務省と共に「翻訳バンク」の運用を開始^{*4}しました。今後はオールジャパン体制で翻訳データをNICTに集積し、日本の翻訳技術の多分野化・高精度化に取り組んでいきたいと考えています。

皆さまからNICTに翻訳データを提供していただく方法としては、次の3つがあります。

1. 「みんなの自動翻訳」webサイトの利用
webサイト上に対訳集を登録すれば、NICTの自動翻訳技術の研究開発目的にすぐに貢献します。
2. NICTとの翻訳データの提供に関する契約の締結
現在29組織から翻訳データの提供があります（2017年8月末現在）。
3. NICTとの自動翻訳技術使用ライセンス契約に伴う翻訳データの提供
翻訳データをNICTに提供するメリットがより明確化されるように、NICTの自動翻訳技術の使用ライセンス契約の際に翻訳データを提供いただく仕組みを新たに用意しました。これにより、提供が見込まれる翻訳データをライセンス料の算定の際に勘案して負担を軽減することが可能です（図2参照）。

提供いただいた翻訳データは、AIデータテストベッドの一環として他の翻訳データと共に集積され、自動翻訳の精度向上に役立ちます。そして、精度が向上した自動翻訳技術を実際に使っていただくことにより更に翻訳データが増え、その結果自動翻

訳の精度が更に向上することになります。社会全体で自動翻訳を育てながら高精度自動翻訳を活用していくというエコシステムが生まれると考えています（図3参照）。

■オールジャパンからオールアジアへ

今後は「翻訳バンク」の周知活動に取り組み、当面の目標としては100万文×100社＝1億文の翻訳データの集積を目指しています。また、ニューラル翻訳は言語に依存せず、同様に「翻訳バンク」も言語に依存しないので、どの言語でも高精度翻訳を実現できます。そこで、将来的には「翻訳バンク」をアジア各国の公的研究機関と共に運用することにより、オールアジア体制で翻訳データの集積に取り組むことができれば、高精度翻訳技術が更に発展し、日本だけではなくアジアの経済・社会の活性化にも貢献できると考えています。

このような取組により、様々な分野で高精度翻訳を実現することで「言葉の壁」をなくし、日本を『世界で最も多言語コミュニケーションが容易な国』にしたい、と外国語学習が苦手の私は思っています。



「長期的なプライバシー保護とビッグデータ利活用の両立」 が可能な暗号化方式(サーバ、サービス方法)

— 世代を超えた将来までプライバシー保護を可能とする世界初の準同型暗号方式 —

医療・金融など社会の様々な分野でパーソナルデータの利活用に貢献するためのプライバシー保護技術が求められています。改正個人情報保護法でも、ビッグデータの利活用とプライバシー保護を両立するための施策が打ち出されています。本発明により、データを暗号化したままセキュリティレベルの向上と様々な演算の両方が行えるため、ビッグデータ利活用と長期的なプライバシー保護との両立が可能となります。

■技術の概要・適用分野等

データを暗号化してしまうとそのまま複雑な演算処理ができず、一度ダウンロードし復号(平文に戻す)してから計算をする必要があります。この方法では復号した際のデータ漏洩の危険性や、更にセキュリティレベルの高いシステム導入のコストが問題となります。また、データを暗号化したままセキュリティレベルを更新する既存方式では、データ形式を変換・分割しなければならないため、2度目の暗号化以後のデータ処理が不可能でした。

このような技術的課題のため既存技術では、疾病、治療や遺伝子情報など、個人に関するプライバシー情報を安全に保管できるのは数年から数十年とされてきました。

本発明では、暗号化と復号処理のプロセスに、量子コンピュータを使っても解読困難な暗号方式のひとつである「格子暗号」の技術を活用し、また更新処理には、新しい暗号文のあて先を復号せずに変更可能とする「代理再暗号化」技術を用いています。実証実験では暗号化線形回帰計算が従来より100倍高速化され、暗号化されたまま安全な状態でのセキュリティレベルの向上が可能という成果が得られています。

このような成果を基に、疾病、治療や遺伝子情報などを本人だけでなく、世代を超えて安全に長期的に扱うことが可能となりました。その結果、例えば医療カルテ情報から個人情報に触れることなく、長期にわたる疾病や治療の統計データだけを抽出・計算することなどができます。

■利用・応用・連携先の探索

本技術の実社会での応用を進めるため、匿名加工さ

れた公開の医療データを用いて運用を行うための検証が行われています。有用性が認められれば医療情報システムや遺伝子保存サービスなどでの導入が期待されています。また、金融工学やマーケティングリサーチの分野での活用も検討されています。

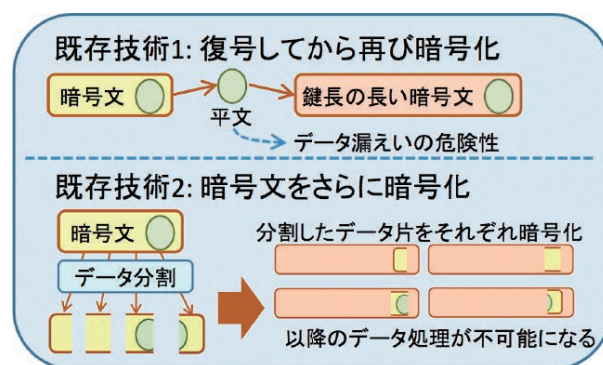


図1 従来の鍵長の変更方法

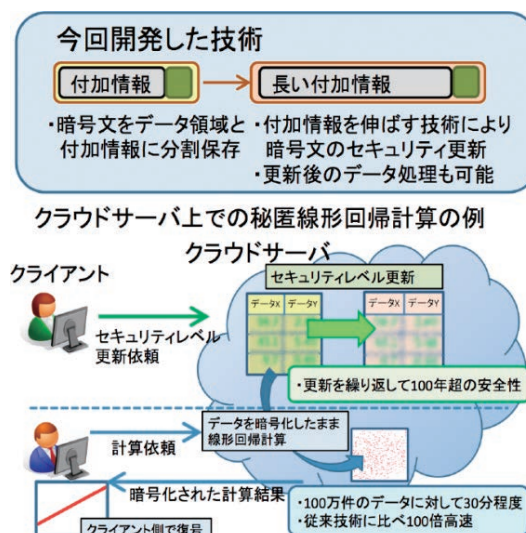


図2 今回開発した鍵長が変更できる準同型暗号方式

〈特許情報〉

公開番号: 特開2016-131271

発明の名称: サーバ、サービス方法

〈連絡先(問合せ等)〉

イノベーション推進部門 知財活用推進室

E-mail: ippo@ml.nict.go.jp

TEL: 042-327-6950 FAX: 042-327-6659

NICT発のスマホ・タブレットアプリ&翻訳Webサイト

NICTでは「多言語音声翻訳技術」の研究開発成果や、本号で紹介した「翻訳バンク」等のプロジェクトの一環として多様なアプリの提供やサイトの運営を行っています。いずれのアプリやサイトも無償（通信料は利用者負担）で利用可能ですので、様々なシーンでご活用ください。

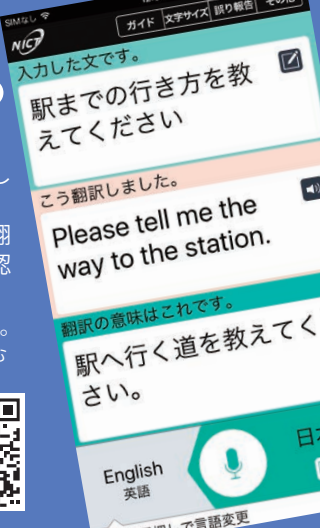
多言語音声翻訳アプリ

ボイストラ

VoiceTra®

話しかけると外国語に翻訳してくれる音声翻訳アプリです。見やすい画面で操作も簡単、翻訳結果が正しいかどうか確認できます。翻訳できる言語は31言語*です。
*中国語、ポルトガル語の方言を含む

iOS 8.0以降 Android 4.1以降



サービス提供：(株) フィート

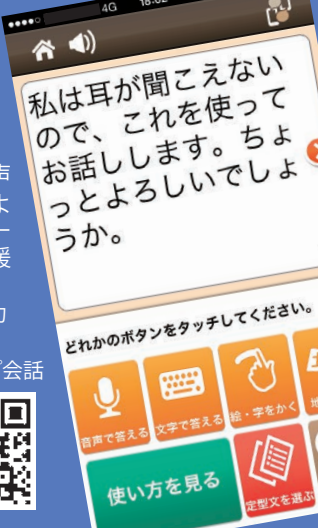
聴障者コミュニケーションアプリ

こえとら

NICTの音声認識技術や音声合成技術を活用することにより、聴障者と健聴者とのスムーズなコミュニケーションを支援するアプリです。

〈特徴〉・音声入力可能・音声出力
・簡単操作・定型文登録と利用
・絵文字・複数台端末でグループ会話

iOS 8.0以降 Android 4.0以降



サービス提供：(株) フィート

声がそのまま文字になる 指で書く会話ができる

スピーチキャンパス

SpeechCanvas®

NICTの音声認識技術を使って、聴覚障害者との会話を強力にサポートするアプリです。話した言葉が次々と画面上にふりがな付きで文字となり、画面を指でなぞれば絵や文字が書けます。操作がシンプルでわかりやすいので、どなたでも簡単に使えます。インターネットが繋がらなくても音声認識してくれるので、いつでもどこでも安心です。

iOS 7.0以降 Android 4.0以降

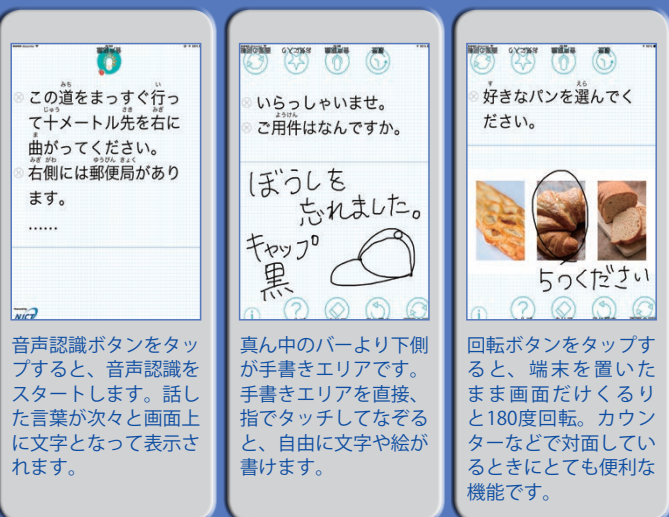


サービス提供：(株) フィート

音声認識ボタンをタップすると、音声認識をスタートします。話した言葉が次々と画面上に文字となって表示されます。

真ん中のバーより下側が手書きエリアです。手書きエリアを直接、指でタッチしてなぞると、自由に文字や絵が書けます。

回転ボタンをタップすると、端末を置いたまま画面だけくりりと180度回転。カウンターなどで対面しているときにとても便利な機能です。



自動翻訳をみんなで育てるサイト

テキストトラ

みんなの自動翻訳@TexTra®

本サイトでは対訳や用語をユーザーが追加することにより、自分好みの自動翻訳エンジンが作成できます。登録(無料)を行うと、あらかじめ登録された自動翻訳を試したり、自動でファイルを翻訳できたり、サイト上で自分で翻訳することができます。

<https://mt-auto-minhon-mlt.ucrj.jgn-x.jp/>



新しい形の翻訳支援・共有サイト

みんなの翻訳

世界中の文書をみんなで協力して翻訳するサイト。誰もが無料で登録し、利用でき、コンピュータや時事、文化、料理、百科事典など様々な記事が翻訳され公開されています。翻訳活動を行う多彩なNGO/NPOも本サイトのサービスを使っています。

<http://trans-aid.jp/>



