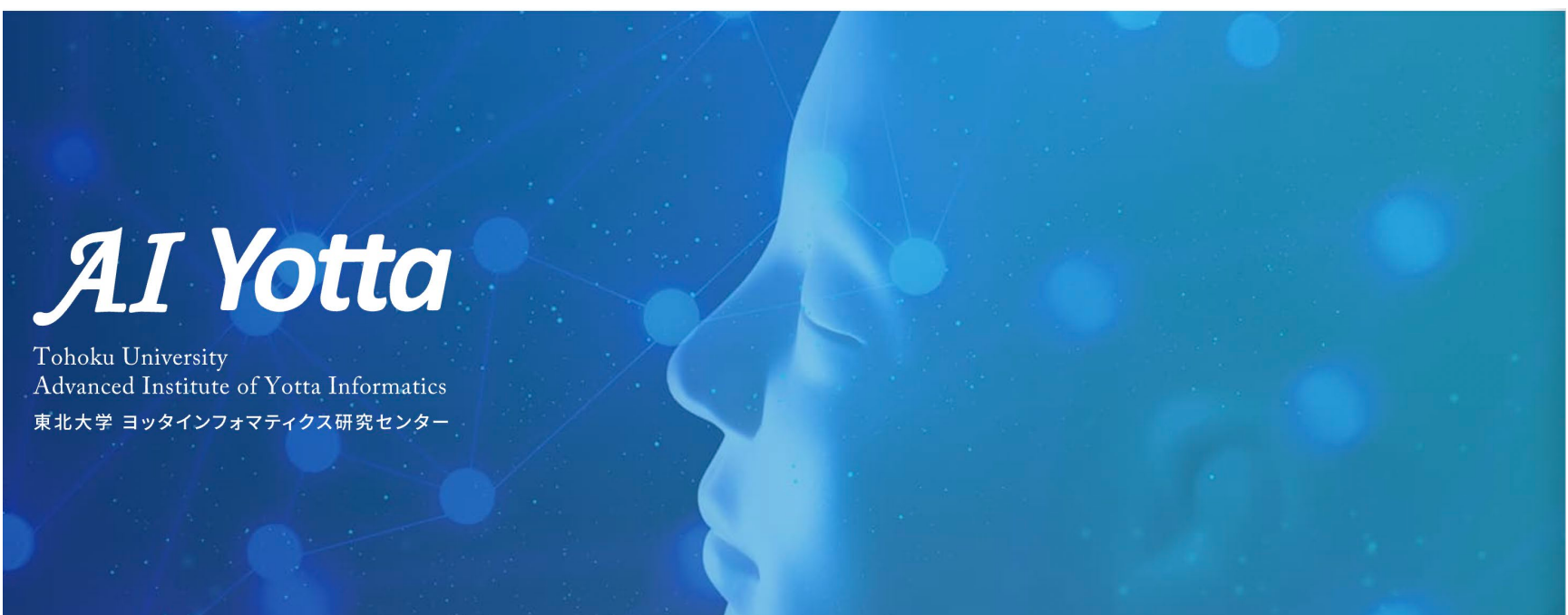


The top half of the image features a blue background with a network of white dots and lines, and a faint profile of a human head. The text 'AI Yotta' is written in a large, white, stylized font.

AI Yotta

Tohoku University
Advanced Institute of Yotta Informatics

東北大学 ヨッタインフォマティクス研究センター

ヨッタインフォマティクス研究センターの挑戦 総合知による情報の価値研究 ～情報生成時の優先づけシステムの構築～

総合知による情報の価値研究

課題：技術的制約により膨大なデータが生成直後に消失している。データの消失は、知の蓄積、創造に対して大きな損失となる。そのため、情報の価値づけ及び取捨選択がなされなければならない。

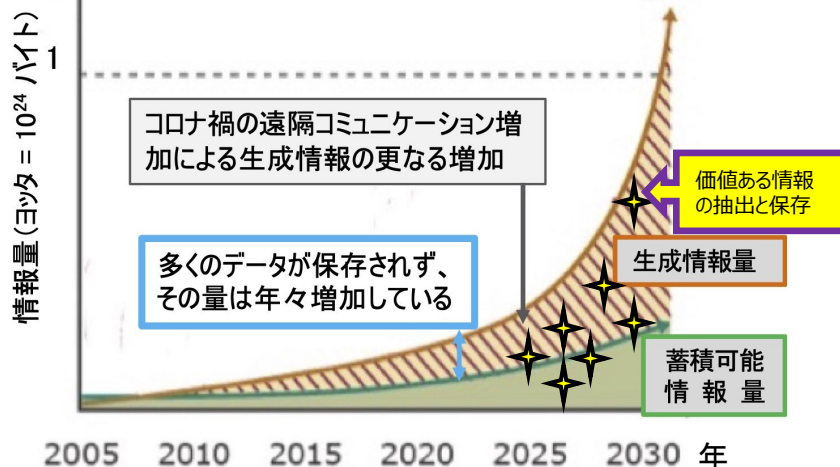
目的：情報の価値づけと取捨選択に関する人文社会科学により培われてきた総合知を利用した新しい学問体系の確立と、その知識を応用し、これまで自然科学の分野で培われてきたデータ科学に基づくシステムの構築を目指す。

価値創造の源泉となる「知」の創造が求められるSociety 5.0の実現に貢献（第6期科学技術・イノベーション基本計画）

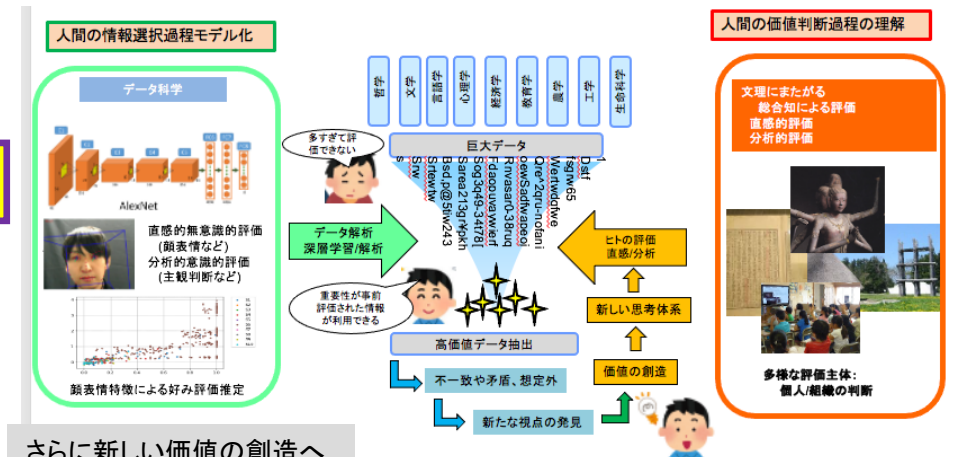
波及効果：文理にわたる研究者の連携により、歴史的視点及び未来予測も含めた総合知に基づき、現在のみならず将来の多様なデータの価値を予測することができるようになる。また、様々な分野のデータについてそれぞれのデータの重要性を予測し、その重要性に基づいたデータの取捨選択と長期保存を可能とすることができるようになる。さらに、取捨選択をデータ生成をエッジで自動的に実現することで、社会全体においてデータの計算、伝送、蓄積コストの大幅な削減に貢献する。

組織：これまでのヨッタインフォマティクス研究センターの取組によって、連携を深めた、哲学、文学、言語学、心理学、美学、経済学、教育学、生命科学、農学、工学、情報科学など文理にわたる多岐の分野教員による組織を構築し、未踏スケールデータアナリティクスセンターやサイバーサイエンスセンターとも連携し、広範囲のデータへのデータ科学の成果を適用することができるものとする。

2030年には生成情報量は1ヨッタバイトを超える速度で増加し、評価されことなく消失するデータは増加の一途



ヨッタスケールデータの有効活用には、生成時の価値評価、優先付けが不可欠
情報価値評価の学理とそれに基づくエッジコンピューティングシステムにより課題解決



さらに新しい価値の創造へ

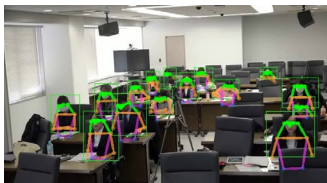
多岐にわたる分野でデータの価値評価研究を推進:支援プロジェクトの成果事例

教育学

学生の姿勢/体動から授業改善に利用可能な価値の高い指標を抽出する手法を提案。

授業への参加度の指標として、動画から得られる生徒の姿勢が利用可能かどうかをした。その結果、頭の向きの標準検討偏差が最後に大きくなることを示され、体の姿勢が注意散漫の度合いを反映している可能性が示唆された。

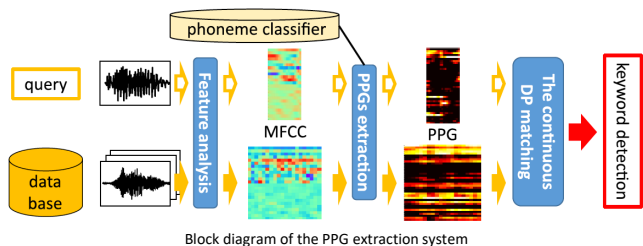
Hatori et al., "Body posture analysis for the classification of classroom scenes"



Example output. The green rectangles show the locations of the detected persons. The dots indicate the estimated keypoints for each person.

言語学

少数言語の会話データから、言語間共通特徴を利用した有益な情報を抽出するための試み。絶滅に瀕している言語文化の承継に必要な高い情報を価値の高さとし優先づけが可能。



Block diagram of the PPG extraction system

ゼロ資源言語の音声データベースからキーワード検出を行う手法として、query-by-sample 音声単語検出手法を提案する。提案手法は、リソースが豊富な複数言語で学習した phonetic

rogram (PPG)を用い、多言 PPGを組み合わせで会話の表行。本手法をカクテル適用し、従来の音声特徴量つく手法よりも良い結果がれることを示した。

zuki et al., "Spoken Term Detection of Resource Language using iogram of Multiple Languages"

文字学

生活空間にある様々な文字の重要性の主観評価要因を調査し、AIによる重要性推定を実現。

本研究では、風景画像中のテキストの重要度を推定する課題に取り組んだ。本研究では、テキストを含む風景画像データセットを構築し、主観評価によるテキストの重要度を割り当てる。我々の提案手法により、既存手法よりも高い精度で重要度を推定できることを示す。

Odaira et al., "Importance Estimation for Scene Texts Using Visual Features"



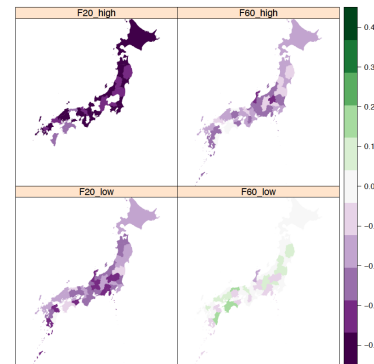
Example of importance. The calculated importance of the text surrounded by the blue rectangle is shown below each image.

経済学

主観的幸福度を高めるために必要な要因を特定できる。個票からなる経済マイクロデータの収集において、幸福度の観点からデータの優先づけが可能となる。

COVID-19の発生をまたぐ2019年から2021年にかけて、日本全国の約26,000人を対象に幸福度調査を実施した。このデータに個人要因と空間的要因の効果を記述する動的回帰モデルを適用し分析を行った。その結果、若い女性が最も深刻な被害を受けたことや、特に低所得者層では決定的な被害が今なお続いていることがわかった。

Sato et al., "Space time analysis of subjective well-being in the COVID-19 outbreak in Japan"



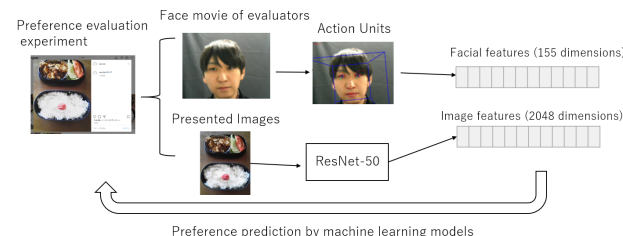
Change of happiness at the prefectural level for high- and low-income female of 20s and 60s in the first period (Dec. 2019 - Sep. 2020).

心理学

好みなど主観的評価を、顔表情などから推定することで、人の評価を推定するシステムが構築できる。保持したい写真であるか否かを判断し、各個人が感じる重要性に沿った優先順位づけを可能とする。

本研究では、機械学習モデルを利用し、自発的な顔表情から人間の画像に対する好み推定を行った。その結果、顔表情特徴からの予測は、人間の評価者の予測を上回ることを示した。さらに、予測に重要な顔表情特徴は画像カテゴリによって異なることを示し、画像カテゴリ間の多様な認知メカニズムに関する知見を得た。

Sato et al., "Prediction of image preferences from spontaneous facial expressions"



Schematic diagram of the image preference evaluation experiment and the analysis of image preference estimation.

多岐にわたる分野でデータの価値評価研究を推進：その他の支援プロジェクト

文学

○画像認識技術による古典籍データの活用

機械学習を用いた画像認識技術により古典籍データの認識や検索を実現する。

(代表：大町(東北大学工学研究科)/分担:佐倉(文学研究科))

考古学

○データ科学による考古学データの活用支援

考古資料を対象とした大容量データに基づく認知的解釈

(代表：鹿又(東北大学文学研究科)/分担:野村(工学研究科))

哲学

○人間中心のAIについての考察

認知的不正義をAI倫理原則の議論と結びつけることによって人間中心のAIを理解する。

(代表：二瓶（松山大学経済学部）)

神経科学

○大規模神経活動データからの有価値情報の抽出

脳の広範囲長時間計測データに基づく脳内情報処理機構の解明

(代表：筒井（東北大学生命科学研究科）/分担塩入（電気通信研究所）他)

○大規模神経活動データの効率的利用

脳ニューロン構造・機能の高精細3次元可視化解析のためのイン・ストレージ・コンピューティング型情報質解析プラットフォームの研究開発

(代表：田中（東北大学電気通信研究所）)

食品科学

○食のライフログから生活の質向上に資する情報に価値をおく

大学生から始まる健康寿命全うに向けた食生活確立への最適な情報探索プロジェクト

(代表：井元（東北大学農学研究科）/分担：佐藤（ヨッタインフォマティクス研究センター）他)

水産科学

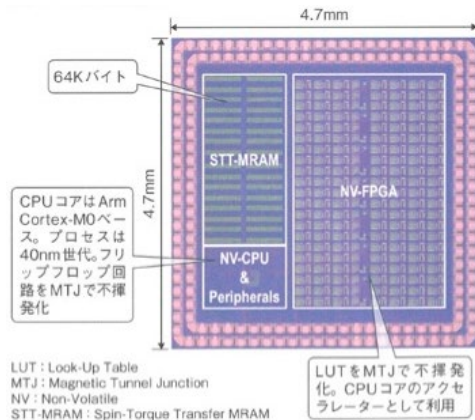
○遺伝情報解析の水産業への活用

ヒラメの遺伝的多様性解析の効率化

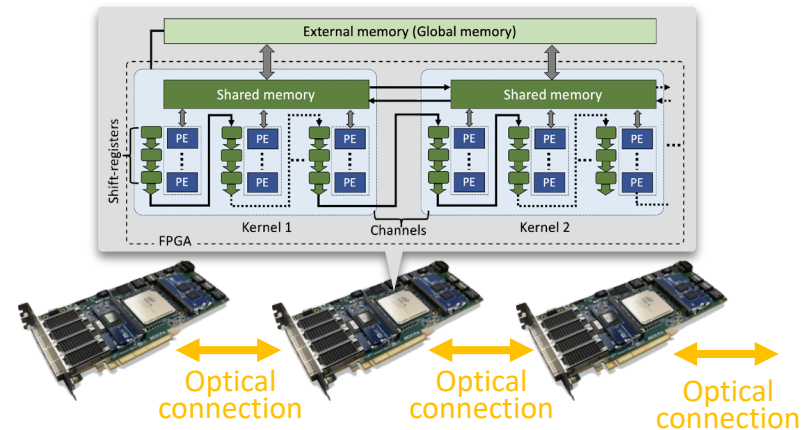
(代表：大林武（東北大学情報科学研究科）/分担：池田（農学研究科）)

価値評価アルゴリズムのエッジコンピューティングへの展開に向けた実績

膨大なデータの価値を評価するためには、機械学習に代表されるAIアルゴリズムを極めて高速かつ省電力に実行するAIハードウェアの研究開発も並行して推進する必要がある。本プロジェクトでは、特にその有望な技術としてスピントロニクス等の新概念デバイスを駆使した新たなAIハードウェアアーキテクチャ^{1),2)}、量子アニーラーハードウェアの研究開発などに取り組み^{3),4)}、端末や分散配置された計算機レベルで価値評価を可能とするエッジコンピューティングの実現を目指している。また、アルゴリズムレベルにおいても視覚や聴覚といった人間の情報処理過程の高速低消費電力演算を実現する確率的演算手法を開発する^{5),6)}など、価値評価—アルゴリズム—ハードウェアの融合を進めている。

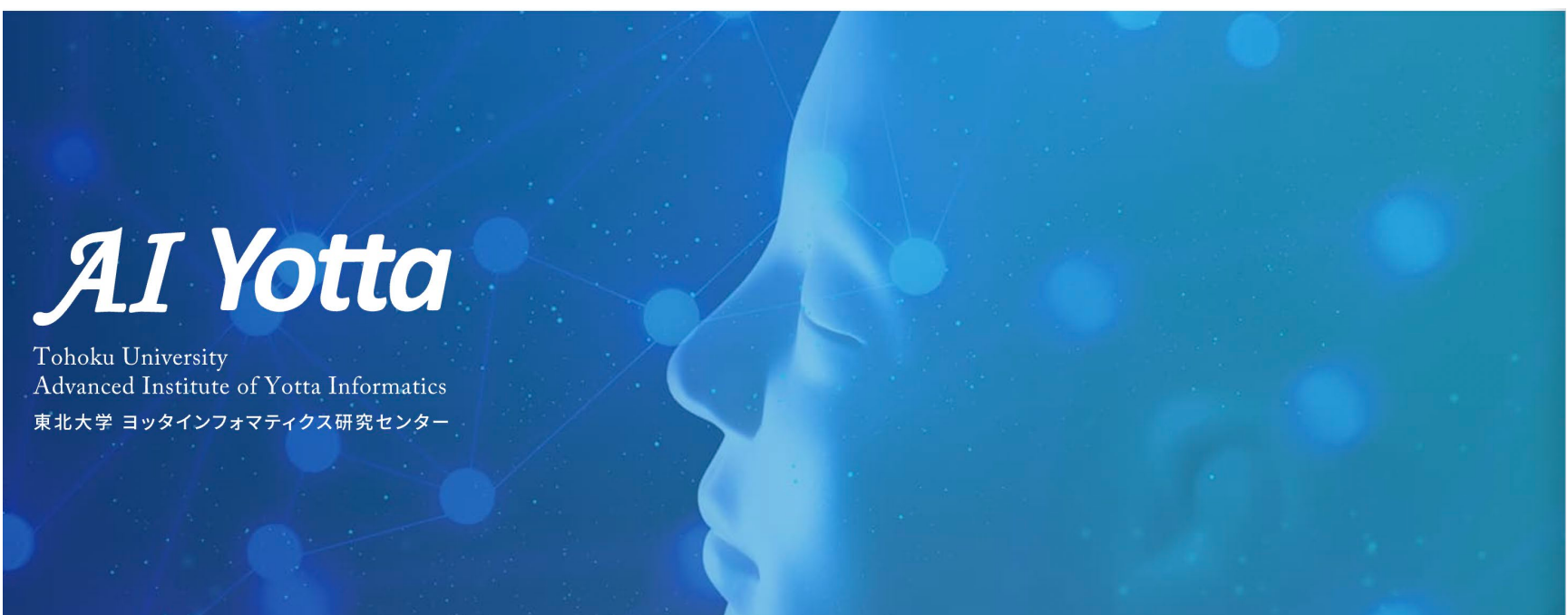


エッジAI向け高性能 & 超低電力スピンマイコンLSI,



量子annealer応用向けFPGAアーキテクチャ

- 1) 遠藤哲郎（工学研究科）スピントロニクスが切り拓く革新的AIプロセッサ 第81回応用物理学会秋季学術講演会（招待講演）
- 2) 羽生貴弘（電気通信研究所）“Challenge of MTJ-Based Nonvolatile Logic-in-Memory Circuits and Their Applications,” 16th IEEE/ACM International Symposium on Nanoscale Architectures, Nov. 2021 (Keynote speech)
- 3) N. Onizawa, K. Katsuki, D. Shin, W. Gross, and T. Hanyu, IEEE Transactions on Neural Networks and Learning Systems, 2022
- 4) 張山昌論（情報科学研究科）A GPU-based Quantum Annealing Simulator for Fully-connected Ising Models Utilizing Spatial and Temporal Parallelism, IEEE Access, vol. 8, pp. 67929-679.
- 5) N. Onizawa, D. Katagiri, K. Matsumiya, W. J. Gross and T. Hanyu, IEEE Journal on Emerging and Selected Topics in Circuits and Systems 2018 Vol. 8 Issue 3 pp. 444-453, 2018
- 6) N. Onizawa, S. Koshita, S. Sakamoto, M. Kawamata and T. Hanyu IEEE Global Conference on Signal and Information Processing (Globalsip 2017) pp. 1315-1319, 2017

The image features a blue background with a network of white dots and lines, and a faint profile of a human head. The text 'AI Yotta' is prominently displayed in a white, stylized font.

AI Yotta

Tohoku University
Advanced Institute of Yotta Informatics

東北大学 ヨッタインフォマティクス研究センター

Challenge of AI Yotta

Investigation of methods to estimate
information values with integrated wisdom
～ System to prioritize data at the time of generation ～

Investigation of methods to estimate information values with integrated wisdom

Background

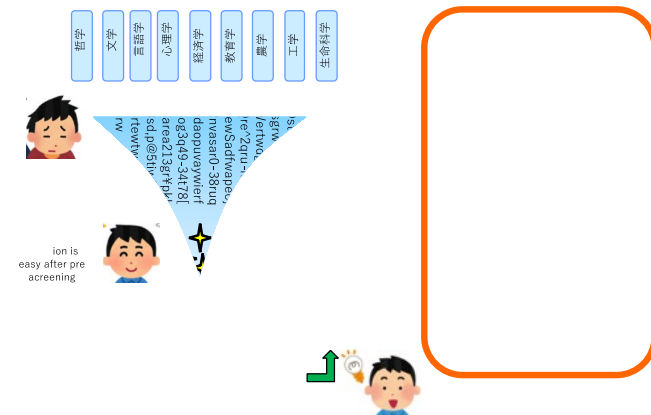
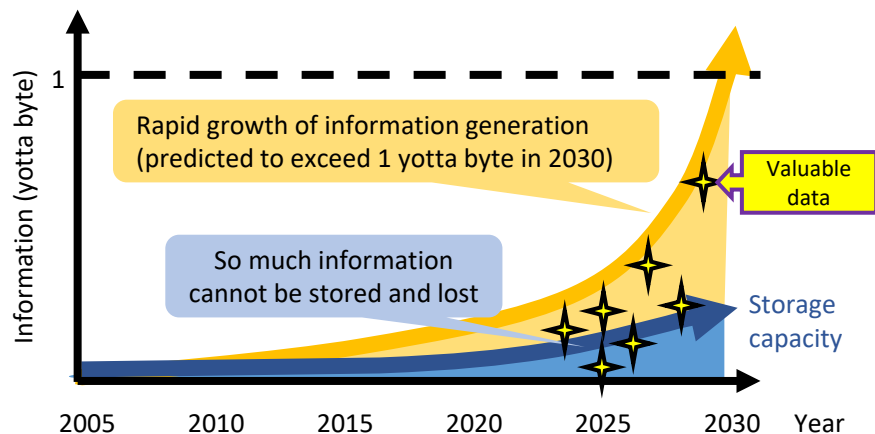
Due to technical restrictions, a huge amount of data is lost immediately after generation. The loss of data is great for knowledge accumulation of human society. To minimize the loss, science and technology to value information to prioritize.

Purpose

To establish a new research field where knowledge cultivated by all academic fields including the humanities and social sciences is combined with data sciences for valuing information as assets, contributing to the realization of Society 5.0 (6th Science and Technology / Innovation Basic Plan).

Goal

To achieve predictions of the value of diverse data through collaboration of researchers across different fields. High value data are prioritized for selection or store, which is automatically realized at the time of data generation by edge computation. This approach contributes to a significant reduction in data calculation, transmission, and storage costs for each person and each organization so that for the society as a whole.

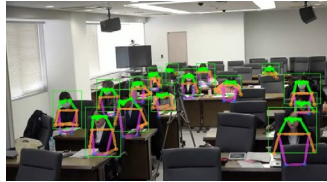


Studies of data value in a variety of research fields #1

Pedagogy

A method is proposed to extract high-value indicators from student behavior that can be used to improve teaching.

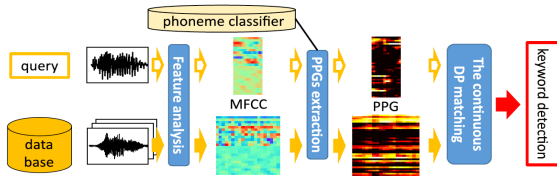
It is investigated whether the body posture of students can be used as an indicator of classroom engagement. A comparison between clusters at the beginning and end of the presentation showed that the standard deviation of head direction becomes large at the end, suggesting that body posture might reflect the degree of distraction. Hatori et al., "Body posture analysis for the classification of classroom scenes"



Example output. The green rectangles show the locations of the detected persons. The dots indicate the estimated keypoints for each person.

Linguistics

The proposed system tried to extract important information from the speech of endangered languages using phonetic features that are universal among languages. It can estimate the priority of the speech signal of any language even when the development of a speech recognizer is impossible.



Block diagram of the PPG extraction system

This paper proposed a query-by-example spoken term detection method from any language. This method utilizes the phonetic posteriorgrams (PPG) trained by multiple resource-rich languages to remove the speaker dependency of the speech while keeping the diversity of the phonemic diversity. The proposed method was applied to Kaqchikel speech, and the result was better than conventional speech features.

Mochizuki et al., "Spoken Term Detection of Zero-Resource Language using Posteriorgram of Multiple Languages"

Grammatology

The subjective factor of importance of various texts in the environment was investigated, and the importance estimation by AI was realized.

We addressed the challenge of estimating the importance of texts in scene images. We constructed scene image datasets, including texts, and assigned an importance to each text via subjective evaluation. The proposed method can estimate the importance with a higher accuracy than the existing method. Odaïra et al., "Importance Estimation for Scene Texts Using Visual Features"



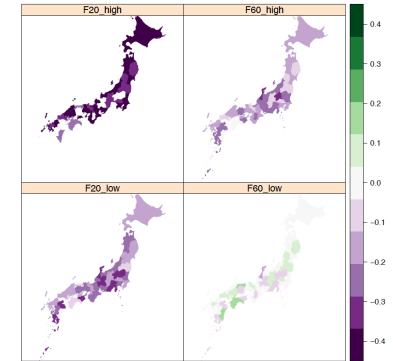
Example of importance. The calculated importance of the text surrounded by the blue rectangle is shown below each image.

Economics

This research aims to identify the factors to improve subjective well-being. Micro datasets in Economics, big data collecting individual behaviours, can be arranged to extract significant parts in terms of subjective well-being.

We conducted happiness surveys for around 26,000 respondents all over Japan on the five periods in 2019-2021, including before and after the COVID-19 outbreak. A dynamic regression model analysis clarifies factors affecting subjective well-being in relation with COVID-19. We find that the young females were most seriously damaged by the pandemic and the damages are still continuing especially for the low income group.

Sato et al., "Space time analysis of subjective well-being in the COVID-19 outbreak in Japan"

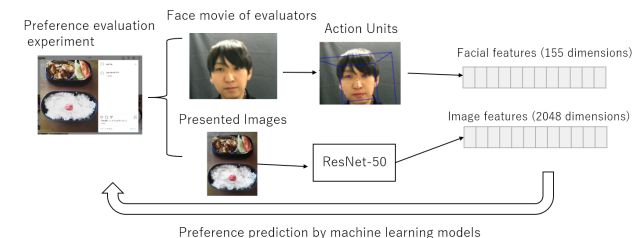


Change of happiness at the prefectural level for high- and low-income female of 20s and 60s in the first period (Dec.2019 - Sep. 2020).

Psychology

We can make a system to estimate human subjective evaluation such as image preference from their facial expressions. This enables us to automatically select images to keep, thus to prioritize information according to individual's perspective of importance.

We can make a system to estimate human subjective evaluation such as image preference from their facial expressions. This enables us to automatically select images to keep, thus to prioritize information according to individual's perspective of importance. Sato et al., "Prediction of image preferences from spontaneous facial expressions"



Schematic diagram of the image preference evaluation experiment and the analysis of image preference estimation.

Studies of data value in a variety of research fields #2

Literature

- Utilization of historical documents by image recognition technology

Realization of recognition and retrieval of historical documents by image recognition technology using machine learning.
(Tohoku University, Omachi & Sakura)

Archeology

- New utilization of archaeological data with data science

Cognitive interpretation based on archaeological big data

(Tohoku University, Kanomata & Nomura)

Philosophy

- Considerations of human-centered AI

Epistemic injustice as a philosophical conception for considering fairness and diversity in human-centered AI principles
(Masuyama University, Nihei)

Neurosciences

- Extraction of valued data from huge data of neural responses.

Investigation of brain information processing by measuring neural responses from while brain for long period of time.

(Tohoku University, Tsutsui, et al)

- Effective use of huge data in neurosciences

Computational Storage Platform for Brain Neural Structure Analytics

(Tohoku University, Tanaka)

Food sciences

- Extract value of information to improving quality of life from food lifelogs

Eating Habits and Psychological Aspects of College Students

(Tohoku University, Imoto et al)

Fisheries science

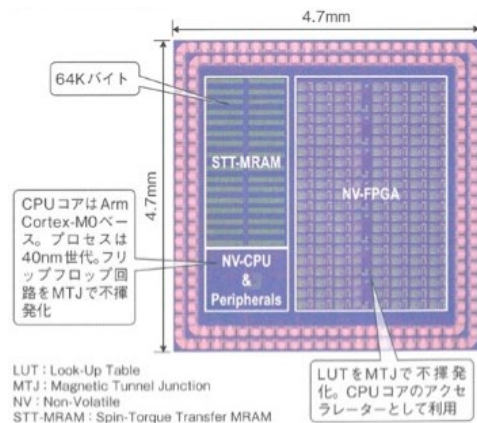
- Applying genetic engineering to fisheries science by data science

Kinship Analysis on wild populations of olive flounder

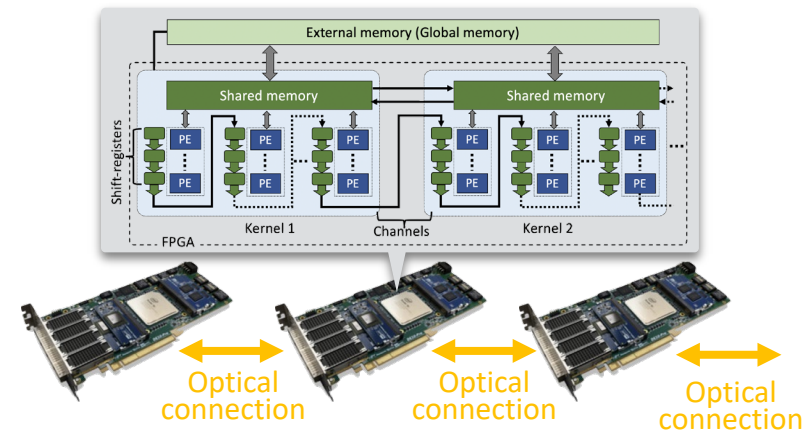
(Tohoku University, Obayashi & Ikeda)

Toward edge computation for data prioritization at the time of data generation

In order to evaluate the value of huge amounts of data at the time of data generation, it is necessary to promote research and development of AI hardware that executes AI algorithms such as machine learning at extremely high speed and with low power consumption. In this project, we are working on new AI hardware based on new concepts such as spintronics devices ^{1), 2)}, and quantum annealer hardwares ^{3), 4)}, which work on terminal/edge devices even for complicated data processing mimicking human judgments. We are also working on applying the AI hardware technologies to models of human vision and audition ^{5), 6)}.



Energy-efficient spintronics-based MCU* LSI for edge AI



FPGA architecture for quantum annealer applications

- 1) T. Endoh, "STT and SOT MRAM technologies and its applications from IoT to AI System," 66th International Electron Devices Meeting (IEDM), 2020 (Invited)
- 2) T. Hanyu, "Challenge of MTJ-Based Nonvolatile Logic-in-Memory Circuits and Their Applications," 16th IEEE/ACM International Symposium on Nanoscale Architectures, Nov. 2021 (Keynote speech)
- 3) N. Onizawa, K. Katsuki, D. Shin, W. Gross, and T. Hanyu, IEEE Transactions on Neural Networks and Learning Systems, 2022
- 4) H. M. Waidyasooriya, and M. Hariyama, "A GPU-based Quantum Annealing Simulator for Fully-connected Ising Models Utilizing Spatial and Temporal Parallelism," IEEE Access, vol. 8, pp. 67929-679, 2020
- 5) N. Onizawa, D. Katagiri, K. Matsumiya, W. J. Gross and T. Hanyu, IEEE Journal on Emerging and Selected Topics in Circuits and Systems 2018 Vol. 8 Issue 3 pp. 444-453, 2018
- 6) N. Onizawa, S. Koshita, S. Sakamoto, M. Kawamata and T. Hanyu IEEE Global Conference on Signal and Information Processing (Globalsip 2017) pp. 1315-1319, 2017