NTT

corevo

NTT Communication Science Laboratories

Open House 2020

on the web

6/4〈THU〉 From 12:00~

Access from: http://www.kecl.ntt.co.jp/openhouse/2020/index_en.html

Science of Machine Learning

01. People on the WWW, give us your computation each!
    - Generating datasets using people and information on the WWW

02. Presenting a quick solution to system failures
    - Generating recovery-command sequences by neural networks

03. Refinement of spatially aggregated data
    - Multivariate Gaussian processes for spatially aggregated data

04. Fast inference of accurate anomaly detectors
    - Transfer anomaly detection for unseen datasets

05. Anomaly detection with low false positive rate
    - Semi-supervised learning for maximizing partial AUC

06. Is the data really biased?
    - Testing combinatorial correlation by decision diagrams

Science of Communication and Computation

07. What happens if every player rushes selfishly?
    - Equilibrium computation of congestion games

08. Handle a huge quantum world through a tiny window
    - Investigation of the ability of indirect quantum controls

09. Tuning machine translation with small tuning data
    - Domain adaptation with JParaCrawl, a large parallel corpus

10. Assessing children's emotional development
    - Investigating developmental changes via multiple cues

11. Creating a personalized picture book
    - Support for parent-child picture book interaction

12. How many words do you know?
    - Vocabulary size test, Reiwa edition

13. Kyomachi Seika will guide you!
    - Training the role-play AI with community cooperation

14. What does he/she think in this situation?
    - Sentiment text generation based on personality

Science of Media Information

15. Can you guess the age from this voice?
    - Deep speaker attribute estimation with speaker clustering

16. More wireless microphones are available in a room
    - BRAVE: Bit-error-robust low-delay audio and voice encoding

17. Pay attention to the speaker you want to listen to (II)
    - Neural selective hearing with audio-visual speaker clues

18. Controlling voice expression using face expression
    - Crossmodal voice expression control

19. Learning to search like human
    - Adaptive spotting for efficient object search

20. Deep learning without data aggregation from nodes
    - Asynchronous consensus algorithm on distributed networks

21. Cardiac model that makes it heart
    - Gaussian process with physical laws for 3D cardiac modeling

22. Listening carefully to your heart beat
    - Cardiohemodynamical analysis based on stethoscopic sounds

Science of Human

23. Make natural-looking illusions by perceptual model
    - Adaptive motion retargeting for illusion-based projection AR

24. Tiny eye movements reflect cognitive states
    - Relation of eye-movement dynamics with cognition and pupil

25. Haptic metameric textures
    - Direct control of perceived texture of 3D printed stimuli

26. What causes emotional change?
    - Monitoring emotion in experimental settings and daily life

27. Special cognitive abilities of esports experts
    - Performance, physiological state, and brain activity

28. Realizing a harmony in rugby scrum
    - Easy assessment of player coordination with wearable sensors

29. What is a "straight" ball?
    - Physical and perceptual attributes of a pitched ball

30. Body representation for quick and skillful action
    - Uncertainty of hand-state estimate regulates stretch reflex

31. Unconscious is smarter than conscious
    - Environmental dependency in visuomotor responses

01 People on the WWW, give us your computation each!

Generating datasets using people and information on the WWW

Abstract

Although contents on the WWW are potentially valuable data as training data for machine learning, they are difficult to use in their current state. Our approach, Browser-based Human Computation (BbHC), offers a cost-effective way to extract desirable data from web contents. BbHC enables people to label various web contents through the web browsers they normally use for web browsing. To accelerate the labeling of data without the inducement of monetary rewards, browser extensions based on BbHC motivate people to continuously engage in labeling tasks through various human computation techniques. We implemented systems based on BbHC to explain how it works. Matome supporter helps us to collect labeled images to create an image classifier. Text monster reduces the cost of annotating word familiarity values for updating a word familiarity database. Multi-voice labeler's purpose is to collect writings with speaker information for natural language processing research.

Deep learning requires much labeled data

Samples to be labeled

- Need budget, if you want to buy them

Labeling task

- Time-consuming, need efforts of many people

Building human computation space on the WWW!

World Wide Web（WWW）

Candidates of samples

- A wide variety of data is stored

Potential workers for labeling tasks

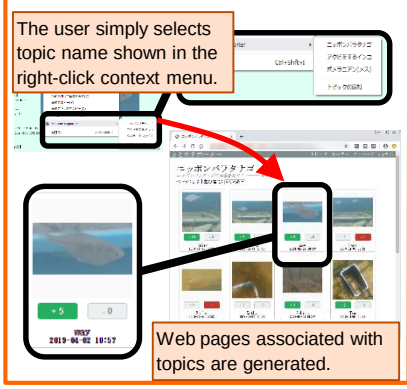
- Many people spends much time on the WWW

Browser-based Human Computation (BbHC)

To collect desirable data from web contents, web browser extensions offer labeling interfaces to users and motivate users to engage in labeling tasks.

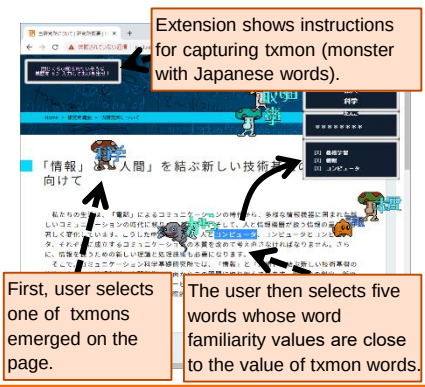
Matome supporter

Users can easily collect images to build web pages that show a collection of images. Collected images are used to update datasets for image classification.



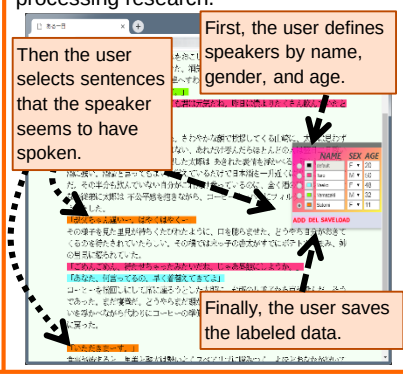
Text monster

Users can enjoy collecting Japanese words which are personified as monsters. The game results are used to update word-familiarity database.



Multi-voice labeler

Users annotate speaker labels to web contents so that smartphones can read them with appropriate voices. Such labels can be used for natural language processing research.



References

[1] Y. Shirai, Y. Kishino, Y. Yanagisawa, S. Mizutani, T. Suyama, "Building human computation space on the www: labeling web contents through web browsers," *Proc. The seventh AAAI Conference on Human Computation and Crowdsourcing（HCOMP2019）*, 2019.

Contact

Yoshinari Shirai Email: cs-openhouse-ml@hco.ntt.co.jp

Learning and Intelligent Systems Research Group, Innovative Communication Laboratory

オープンハウス 2020

02 Presenting a quick solution to system failures

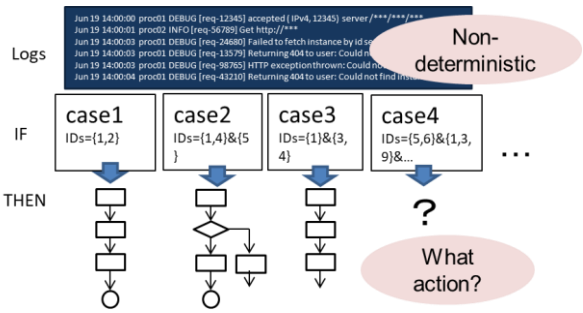
Generating recovery-command sequences by neural networks

Abstract

We propose a method for automatically generating recovery-command sequences, which is intended to support quick recovery actions by system operators and to achieve automatic recovery from ICT (information and communication technology)-system failures. Our method is based on Seq2Seq (sequence-to-sequence), a neural network model usually used to solve translation tasks in the field of natural language processing. This model can learn complex relationships between logs obtained from equipment and recovery commands that operators executed in the past. When a new failure occurs, our method estimates plausible commands that recover from the failure on the basis of collected logs. Our method also evaluates the confidence score of the estimated recovery-command sequences. Operators can use this confidence score as a criterion to determine whether the estimated recovery-command sequence should be executed.

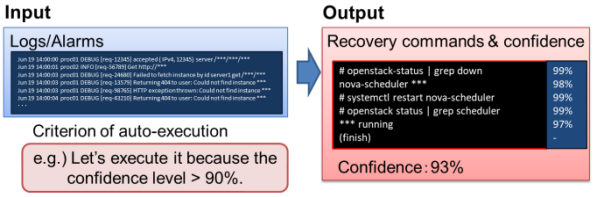
Problem: Construction of Recovery rules

- Automated recovery requires predefined rules associating logs/alarms with recovery actions.
- Operators spend a lot of time making rules and action sequences



Technology: Automatic generation of recovery commands

- Input: Logs/Alarms, Output: Recovery commands
  → Reducing operation cost
- Confidence score of estimated commands
  → Supporting operators' judgement of command execution



Method: Log–Command transformation by Seq2Seq

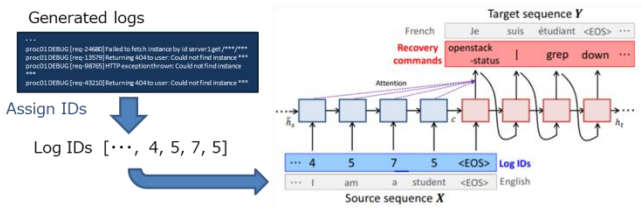
We constructed a neural network model that converts log/alarm sequences into the corresponding recovery–command sequences on the basis of historical data by using Seq2Seq , which can learn the relationship between multiple sequences.

Input: Log IDs

- Using a log templater [2] to assign IDs to logs

Output: Words included in recovery commands

- Considering an action such as "Pressing Enter key" as a word
- Evaluating the likelihood of the output sequence as the confidence score of estimation



References

[1] H. Ikeuchi, A. Watanabe, T. Hirao, M. Morishita, M. Nishino, Y. Matsuo, K. Watanabe, "Recovery command generation towards automatic recovery in ICT systems by Seq2Seq learning," *Proc. of IEEE/IFIP Network Operations and Management Symposium (NOMS)*, 2020, to appear.

[2] T. Kimura, A. Watanabe, T. Toyono, K. Ishibashi, "Proactive failure detection learning generation patterns of large-scale network logs," *IEICE Transactions on Communications*, Vol. E102-B, No2, pp. 306–316, 2019.

Contact

Hiroki Ikeuchi Email: cs-openhouse-ml@hco.ntt.co.jp

NTT Network technology laboratories

オープンハウス 2020

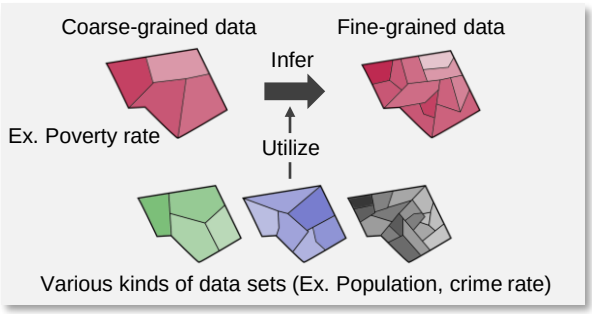
03 Refinement of spatially aggregated data

Multivariate Gaussian processes for spatially aggregated data

Abstract

Spatial data collected from cities are often aggregated into geographical partitions (e.g., districts). We propose a probabilistic model for refining coarse-grained aggregated data by utilizing multiple aggregated data sets with various granularities. Our model is based on multivariate Gaussian processes (GPs), in which dependences between data sets are established by linearly mixing some independent latent GPs. We newly introduce an observation model with spatial aggregation processes, which allows us to use multiple aggregated data sets for the refinement task even if they have various granularities. Our model can be used for predicting data values with arbitrary fine granularity; it is useful for finding key pin-point regions (e.g., poverty area) in a city, efficiently. In the future, we will extend the model to handle data gathered from multiple cities simultaneously.

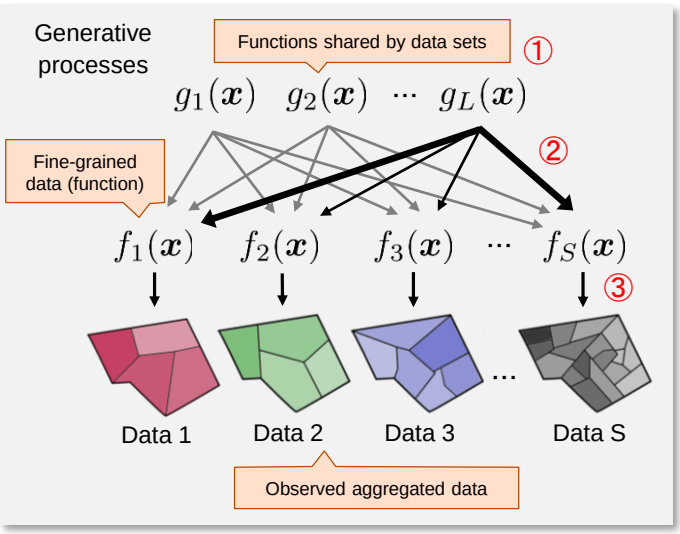
Problem : refinement of spatially aggregated data



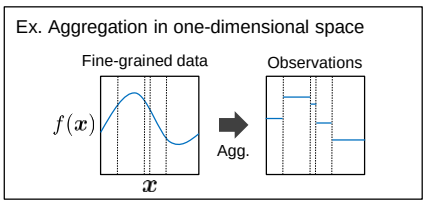
- **Task**
  Predicting fine-grained data by utilizing multiple aggregated data sets with various granularities.
- **Idea**
  Interpolating coarse-grained data by using fine-grained data that have spatial correlation similar to target data.
- **Difficulty**
  It is not straightforward to evaluate the similarity between aggregated data sets whose granularity is different.

Proposal : spatially aggregated Gaussian processes

We design generative processes of multiple aggregated data sets and train the model from observation data.



- **Point ① : Spatial interpolation**
  Assume the underlying smooth functions (i.e., GPs).
- **Point ② : Dependences between data sets**
  Share the functions $g_l(x)$ among multiple data sets on the basis of the similarity of spatial patterns.
- **Point ③ : Spatial aggregation processes**
  Average fine-grained data values over regions



References

[1] Y. Tanaka, T. Tanaka, T. Iwata, T. Kurashima, M. Okawa, Y. Akagi, H. Toda, “Spatially aggregated Gaussian processes with multivariate areal outputs,” *Proc. 33rd Conference on Neural Information Processing Systems (NeurIPS)*, pp. 3000-3010, 2019.

Contact

Yusuke Tanaka Email: cs-openhouse-ml@hco.ntt.co.jp
Service Evolution Laboratory

オープンハウス 2020

04 Fast inference of accurate anomaly detectors

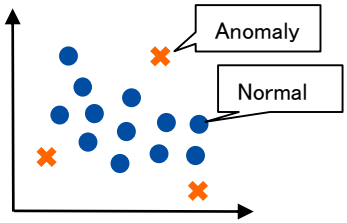
Transfer anomaly detection for unseen datasets

Abstract

We propose a method to improve the anomaly detection performance on target datasets by transferring knowledge on related datasets. Although anomaly labels are valuable to learn anomaly detectors, they are difficult to obtain due to their rarity. To alleviate this problem, we use anomalous and normal instances in the related datasets as well as target normal instances. Our method can infer the anomaly detectors for target datasets without re-training by introducing a novel permutation-invariant neural network. This neural network takes the set of normal instances as an input and infers the dataset-specific anomaly detector from the set. By learning with multiple related datasets, our neural network can learn the latent relationship between the anomaly detector for each dataset and the set of normal instances in the dataset. When target normal instances can be used during training, our method can also use them for training in a unified framework.

What's is Anomaly Detection?

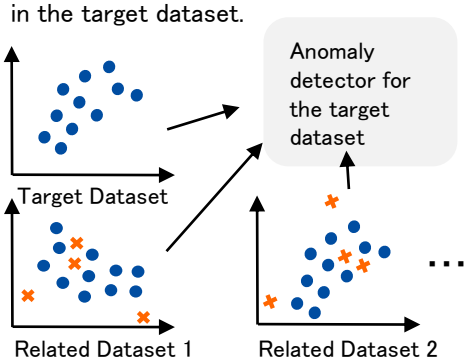
A task to detect anomalous instances in a dataset.



- We can detect anomalies accurately by using normal/anomalous data.
- However, it is difficult to collect anomalies due to their rarity.

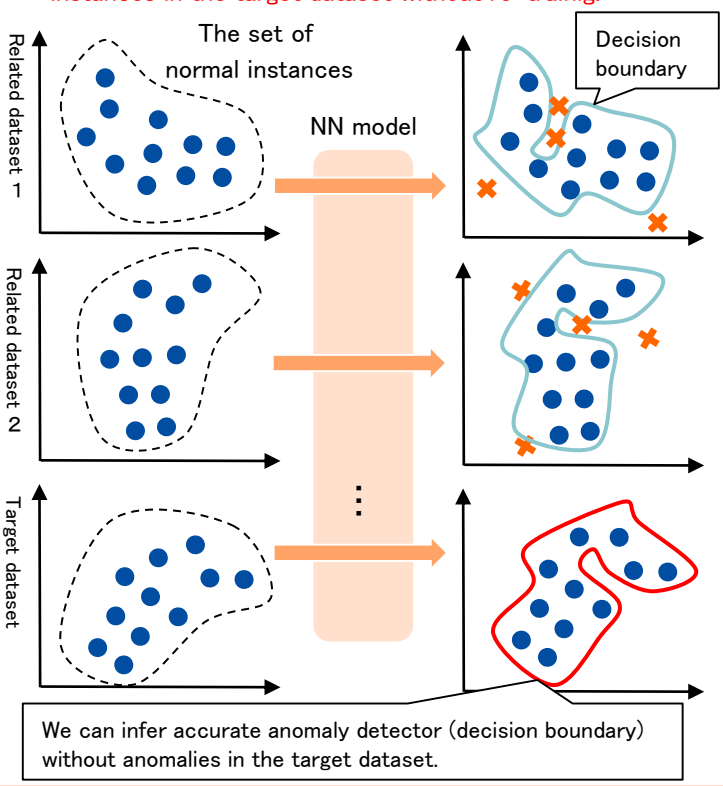
Approach

We use anomalous/normal instances in "related" datasets as well as normal ones in the target dataset.



Proposed Method

- We propose a neural network (NN) that infers an appropriate anomaly detector from the set of normal instances.
- We pre-trains this neural network with multiple related datasets.
  ⇒ It can infer the accurate anomaly detector from normal instances in the target dataset without re-trainig.



References

[1] A. Kumagai, T. Iwata, Y.Fujiwara, "Transfer anomaly detection by inferring latent domain representations," *Proc. 33rd Conference on Neural Information Processing Systems (NeurIPS)*, 2019.

Contact

Atsutoshi Kumagai Email: cs-openhouse-ml@hco.ntt.co.jp
Software Innovation Center

オープンハウス 2020

05 Anomaly detection with low false positive rate

Semi-supervised learning for maximizing partial AUC

Abstract

The partial area under a receiver operating characteristic curve (pAUC) is a performance measurement for binary classification problems that summarizes the true positive rate with the specific range of the false positive rate. Obtaining classifiers that achieve high pAUC is important in a wide variety of applications, such as anomaly detection and medical diagnosis. Although many methods have been proposed for maximizing the pAUC, existing methods require many labeled data for training. We propose a semi-supervised learning method for maximizing the pAUC, which trains a classifier with a small amount of labeled data and a large amount of unlabeled data. To exploit the unlabeled data, we derive two approximations of the pAUC: the first is calculated from positive and unlabeled data, and the second is calculated from negative and unlabeled data. A classifier is trained by maximizing the weighted sum of the two approximations of the pAUC and the pAUC that is calculated from positive and negative data.

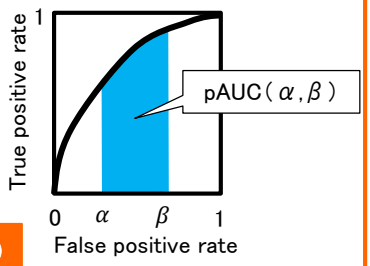
Partial AUC

Partial Area Under the Receiver Operating Characteristic Curve

PAUC : Area Under the ROC Curve when $\alpha \leq$ False positive rate $\leq \beta$

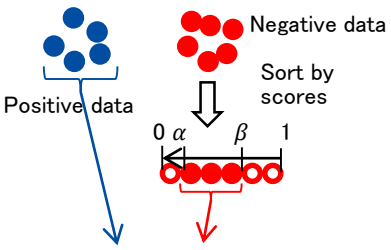
Evaluation measurement for binary classifiers when positive and negative sample size is unbalanced

Applications: Anomaly detection and medical diagnosis: reduce costs by reducing false alarm



Existing method (supervised)

Inputs labeled data (positive and negative data), and learns score function s by maximizing the pAUC



$$\mathbb{E}[I(s(\mathbf{x}^{\mathrm{P}}) > s(\mathbf{x}^{\mathrm{N}}))] = \mathrm{pAUC1}(\alpha, \beta)$$

pAUC calculated by positive and negative data

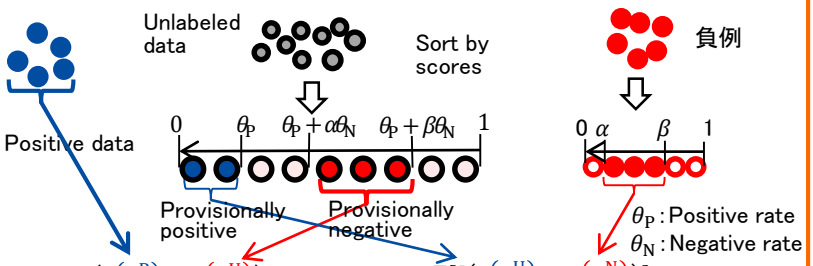
Objective function

$$L = \mathrm{pAUC1}(\alpha, \beta)$$

Proposed method (semi-supervised)

Derives two approximations of pAUC using unlabeled data, and uses them for learning the score function

Assumes data with high (low) scores as positive (negative) with the boundary of positive rate θ_{p}



$$\mathbb{E}[I(s(\mathbf{x}^{\mathrm{P}}) > s(\mathbf{x}^{\mathrm{U}}))] \approx \mathrm{pAUC2}(\alpha, \beta)$$

Approximated pAUC calculated by positive and unlabeled data

$$\mathbb{E}[I(s(\mathbf{x}^{\mathrm{U}}) > s(\mathbf{x}^{\mathrm{N}}))] \approx \mathrm{pAUC3}(\alpha, \beta)$$

Approximated pAUC calculated by negative and unlabeled data

Objective function: weighted sum of three pAUCs

$$L = \lambda_1 \mathrm{pAUC1}(\alpha, \beta) + \lambda_2 \mathrm{pAUC2}(\alpha, \beta) + \lambda_3 \mathrm{pAUC3}(\alpha, \beta)$$

References

[1] T. Iwata, A. Fujino, N. Ueda, "Semi-supervised learning for maximizing the partial AUC," *Proc AAAI Conference on Artificial Intelligence (AAAI)*, 2020.

Contact

Tomoharu Iwata Email: cs-openhouse-ml@hco.ntt.co.jp
Ueda Research Laboratory

オープンハウス 2020

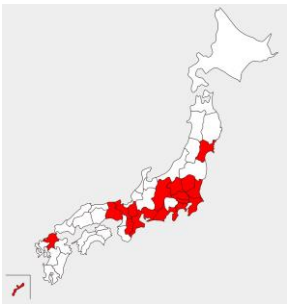
06 Is the data really biased?

Testing combinatorial correlation by decision diagrams

Abstract

We sometimes observe data with structures; population changes of cities on a map, traffic densities of roads on a traffic network, and reactions of sensors on a sensor network. Then, it is a natural question that the observations depend on the structure or not. Testing combinatorial correlation is a statistical method to answer the question: however, the test generally requires the exponential time because it considers all possible observations to evaluate the rarity of the current observation. In this research, we propose an efficient testing method using decision diagrams (DDs) that are a compact representation of a family of sets. We first compress the hypothesis patterns, which define the structure of the observations, by a DD and then construct another DD that compresses rare events to evaluate the rarity of the current observations. Our method reduces the testing time from 10^8 years to only 1 day in the case of testing binary observations on the Japanese prefecture map.

Combinatorial Correlation — Observations depend on Structure?



Ex1: Hotspot of Diseases

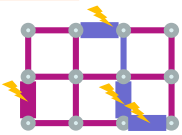
Red areas have so many patients.

White areas have a few patients.

This disease has

1. Hotspot (locality)? **Null Hypothesis**
2. No Hotspot (Bias)? **Alternative Hypothesis**

※ This observation is just an example.



Ex2: Sensor Network

Sensor Reaction

1. Intruder?
2. Noise?

Hist.	#
A B C	3
A B	5
B C	1
A	1
B	1

Ex3: Shopping History

Combination of A&B is

1. Popular?
2. Coincident?

Difficulty — Consider all possible observations to compute the P-value.

Combinatorial Correlation Testing

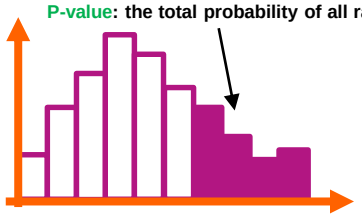
1. Get observations x and **hypothesis patterns** $\mathcal{F}$
2. Compute **Scan statistics** $K(x)$
3. Compute the **P-value** of $K(x)$ by **rare events** $\mathcal{W}$
4. Reject Alt. Hypo. if **P-value** ≦ Significance level (0.05)

$\mathcal{F}$ and $\mathcal{W}$ are exponentially huge. ⇒ Naïve computation takes 10^8 years!! Our method takes only 1 day!!

$p(K(x) = k)$

Rare Event: w **s.t.** $K(w) \geq K(x)$

P-value: the total probability of all rare events.



Histogram of K(x) considering all possible observations

Proposed method — Compute $K(x)$ and P-value on decision diagrams (DDs) of $\mathcal{F}$ and $\mathcal{W}$.

Scan Statistics

$$K(x) = \max_{S \in \mathcal{F}} \sum_{v \in S} x_v$$

Compute without Decompression

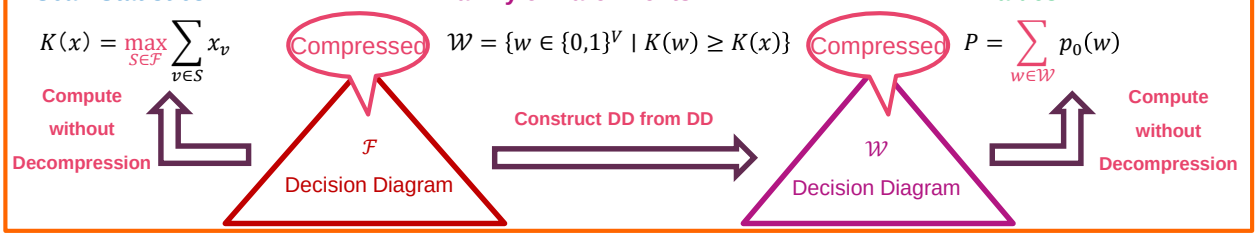
Family of Rare Events

$$\mathcal{W} = \{w \in \{0,1\}^V \mid K(w) \geq K(x)\}$$

P-values

$$P = \sum_{w \in \mathcal{W}} p_0(w)$$

Compute without Decompression



References

[1] M. Ishihata, T. Maehara: "Exact Bernoulli scan statistics by binary decision diagrams," *The 28th International Joint Conference on Artificial Intelligence (IJCAI-2019)*, 2019.

Contact

Masakazu Ishihata Email: cs-openhouse-ml@hco.ntt.co.jp

Learning and Intelligent Systems Research Group, Innovative Communication Laboratory

オープンハウス 2020

07 What happens if every player rushes selfishly?

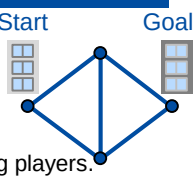
Equilibrium computation of congestion games

Abstract

In a road or telecommunication network, an edge (link) may be congested and may incur more cost if many people use it. We can compute how much each edge will be congested if each user of such network chooses a path (route) selfishly, i.e. chooses a path with minimum cost without any guidance or control. The computation of congestion requires calculation of probability for each of all available paths, which is generally prohibitive because there are a great many number of paths. To overcome this problem, we use a data structure called binary decision diagram to represent all available paths compactly. With this data structure, we can speed up the computation dramatically, and can compute such state of a network with realistic size. This study enables us to predict the state of congestion for a road and telecommunication network in a simple way, which is useful for designing such networks.

Congestion Games and Equilibrium State

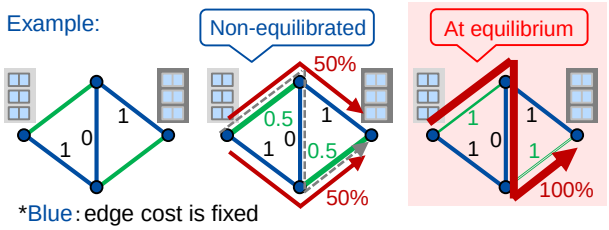
- There are infinite num. of players.
- Each player chooses a path.
  - Chooses a path with lower cost.
- Available paths are fixed.
- Edge cost increases w.r.t. mass of using players.



-> What path is chosen by each player?

Equilibrium state : each player chooses a path with (currently) minimum cost from available paths

Example:



*Blue : edge cost is fixed

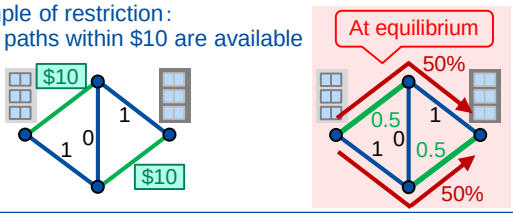
*Green : edge cost is equal to proportion of players using it

Increasing cost w.r.t. mass of using players

Computing Equilibrium

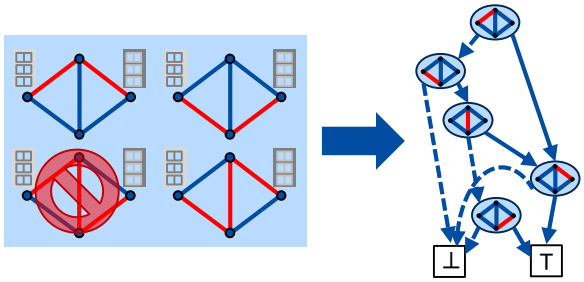
- All paths are available -> easy to compute
- Available paths are restricted -> difficult to compute since all available (enormous) paths should be enumerated

Example of restriction:
Only paths within $10 are available

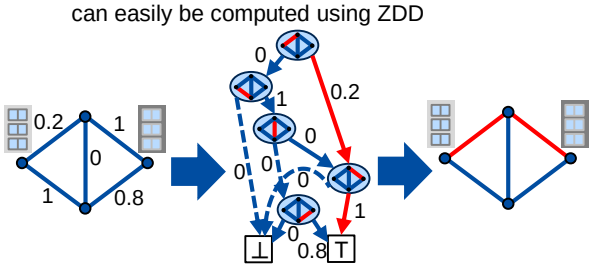


Solution with Decision Diagrams

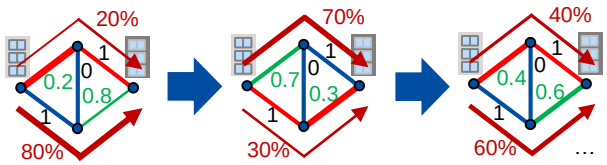
Using zero-suppressed binary decision diagram (ZDD), all available paths can be represented in compact form
- e.g. 8 quadrillion paths may be represented in 1MB



Key 1: the path with minimum cost among available ones can easily be computed using ZDD



Key 2: equilibrium can be computed by performing the following steps repetitively:
- computing path with currently minimum cost
- augmenting players using this path



References

[1] K. Nakamura, S. Sakaue, N. Yasuda, “Practical Frank—Wolfe method with decision diagrams for computing Wardrop equilibrium of combinatorial congestion games,” *Proc. 34th AAAI Conference on Artificial Intelligence (AAAI)*, 2020.

Contact

Kengo Nakamura Email: cs-openhouse-ml@hco.ntt.co.jp
Linguistic Intelligence Research Group, Innovative Communication Laboratory

オープンハウス 2020

08 Handle a huge quantum world through a tiny window

Investigation of the ability of indirect quantum controls

Abstract

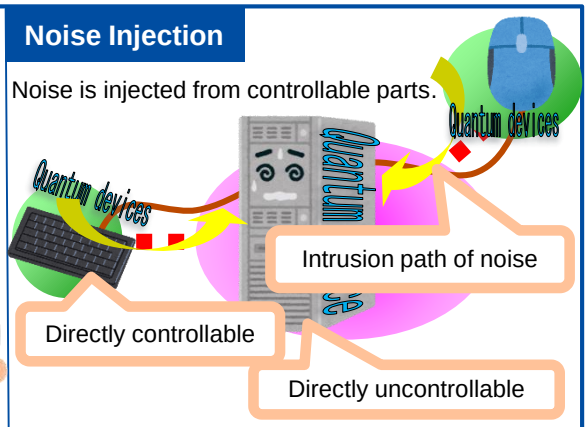
It is difficult to manipulate an entire large quantum system directly. When we try to do so, huge noise will be injected into the system. However, if we can indirectly control the system via a restricted part of it, we will be able to suppress the injected noise. In this research, we investigated the effect of the restriction mathematically, and succeeded in completely categorizing the set of operations in the case of indirect control. This result indicates that, if the degree of the freedom of the controllable part is more than two, we can universally control the whole quantum system in effect even when the degree of freedom of the uncontrollable part is very large. This knowledge provides a new strategy for constructing a noise-less quantum computer or any other noise-less device for quantum information processing. If we can construct such a device, we can realize quantum information processing, e.g. factorization of huge numbers with a quantum computer.

Current Quantum Computer

- Company I and Company G construct 53-qubit quantum computers with the noise.
- Noise increases as the number of qubits increases.
- To factorize a thousand-bit integer, millions of qubits are needed even when the noise is small.

Noise Injection

Noise is injected from controllable parts.

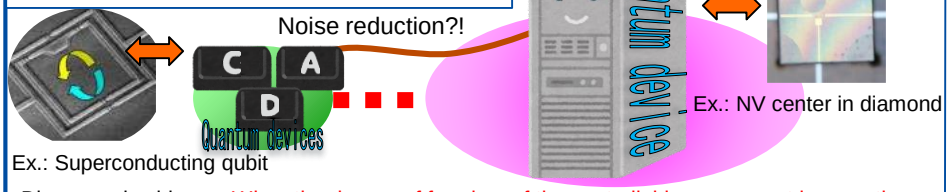


New strategy: Prevent noise injection at the sacrifice of controllability

Theoretical approach for noise reduction

A quantum computer can be made as a composite system consisting of a directly controllable component and a directly uncontrollable component with a spontaneous interaction between them.

—Question—
Can the whole composite system be controlled even if the directly controllable component is small?



Ex.: Superconducting qubit

Ex.: NV center in diamond

Discovered evidence: When the degree of freedom of the controllable component is more than two, we can completely control the whole system in effect for any spontaneous interaction, though we can not when it is two.

With this strategy, the noise could be suppressed even if the total number of qubits is drastically increased!

Mathematical basis

In the case of indirect control of a quantum system, we prove the fact that a set of all the executable operations (matrices) for a fixed interaction can be written as $L \approx \mathcal{L}(su(d_S) \otimes J \cup i\, I \otimes [J, J])$ where the set J must satisfy a certain condition.

The condition depends on the degree of freedom of the controllable component.

References

[1] G. Kato, M. Owari, K. Maruyama, “Algebra and Hilbert space structures induced by quantum probes,” *Annals of Physics*, 412（2020）168046

[2] G. Kato, M. Owari, K. Maruyama, “Hilbert space structure induced by quantum probes,” *Proceedings of the 11th Italian Quantum Information Science conference（IQIS2018）*, Catania, Italy, 12（2019）4

Contact

Go Kato Email: cs-openhouse-ml@hco.ntt.co.jp
Computing Theory Research Group, Media Information Laboratory

オープンハウス 2020

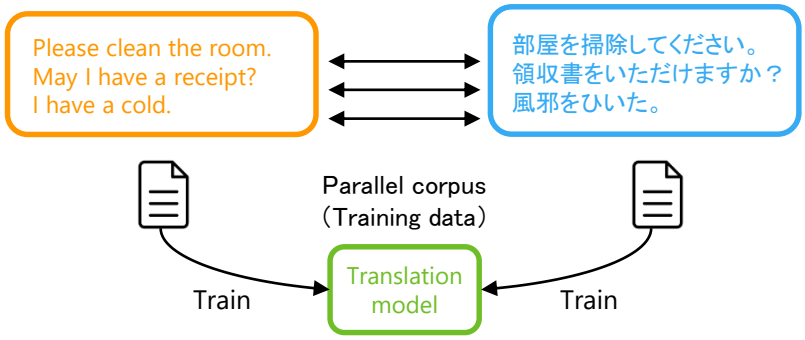
09 Tuning machine translation with small tuning data

Domain adaptation with JParaCrawl, a large parallel corpus

Abstract

Recent machine translation algorithms mainly rely on parallel corpora. However, since the availability of parallel corpora remains limited, only some resource-rich language pairs can benefit from them.
We constructed a parallel corpus for English-Japanese, for which the amount of publicly available parallel corpora is still limited. We constructed the parallel corpus by broadly crawling the web and automatically aligning parallel sentences.
Our collected corpus, called JParaCrawl, amassed over 10 million sentence pairs. JParaCrawl is now freely available online for research purposes.
We show how a neural machine translation model trained with it works as a good pre-trained model for fine-tuning specific domains and achieves good performance even if the target domain data is limited.

Training Machine Translation Model



- Machine Translation (MT) model learns automatically from the parallel sentences (parallel corpus).
  → The amount of parallel sentences is the key to its accuracy.
- We need a large parallel corpus for each domain to make a practical MT system, but such domains are limited.
- Our purpose is to accurately translate low-resource domains, which only have a small number of parallel sentences.

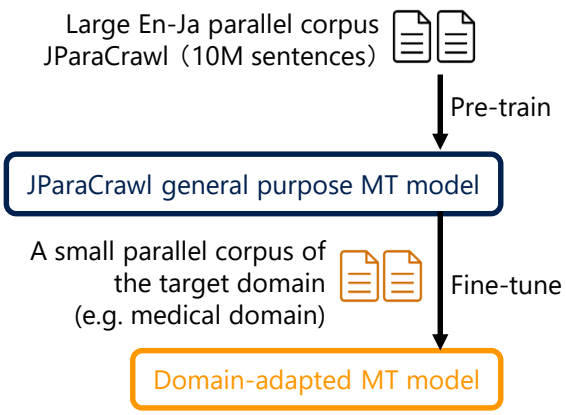
How to create a large En-Ja parallel corpus

Crawl the web (150k websites, 14TB) → Find parallel sentences

- We created a large-scale English-Japanese parallel corpus "JParaCrawl" that contains more than 10M sentences by largely crawling the web and automatically aligning the parallel sentences.
- Until now, freely available parallel corpora are up to 3.0M, so this is more than three times larger than the previous largest one.
- This corpus covers broad domains since it is based on the web.

Our corpus is freely available online for research purposes.
http://www.kecl.ntt.co.jp/icl/lirg/jparacrawl/

Domain adaptation with small tuning data



- Even if the amount of the target domain parallel corpus is small, we can achieve good performance with the combination of JParaCrawl.
- Fine-tuning only needs small computational cost.

References

[1] M. Morishita, J. Suzuki, M. Nagata, "JParaCrawl: A large scale web-based Japanese-English parallel corpus," *Proc. 12th International Conference on Language Resources and Evaluation (LREC)*, May 2020.

Contact

Makoto Morishita Email: cs-openhouse-ml@hco.ntt.co.jp
Linguistic Intelligence Research Group, Innovative Communication Laboratory

オープンハウス 2020

10 Assessing children's emotional development

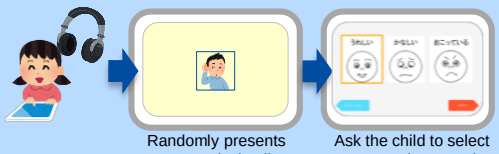
Investigating developmental changes via multiple cues

Abstract

Understanding one's own and others' emotions is an essetial skill in interpersonal communication. To investigate the development of emotion understanding in children, we computerized the Affect Knowledge Test (Japanese version) and developed an emotion voice test. These assessment tools allow us to examine how children percieve and process multiple emotion cues and to objectively quantify their development. Compared to the traditional method in which only trained experimenters can administer assessments to children, the computerized tests made it possible to assess children's emotion understanding in any setting without experts. Using the tests, researchers as well as teachers would be able to detect children with developmental delays and those who struggle with interpersonal communication, and to find which emotional cue is difficult for them to percieve. Obtaining this information would allow us to develop an appropriate intervention to facilitate their development.

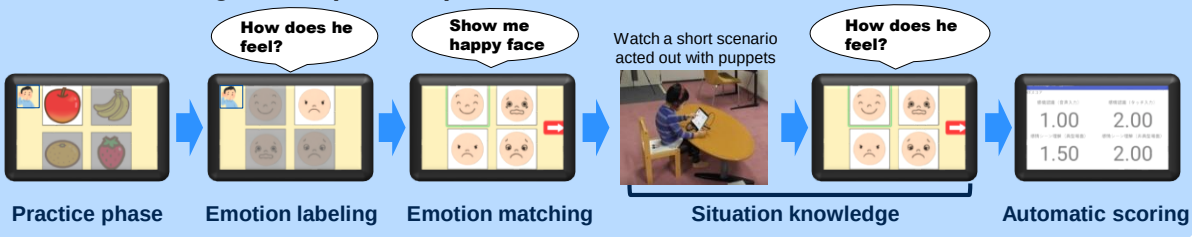
Developing assessment tools for children's emotion understanding

Emotion Voice Test For age 3-12



Affect Knowledge Test* (AKT: Japanese version) For preschoolers (age 3-5)

*Originally developed in US



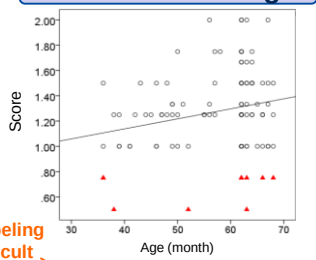
Emotion development in 3-5 years old

Participants:
116 Japanese 3 to 5-year-olds (65 boys)

○ Typically developing children
▲ Developmentally delayed children

Can detect children with developmental delays

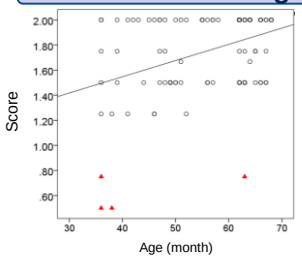
Emotion labeling



Emotion labeling is more difficult then other tasks

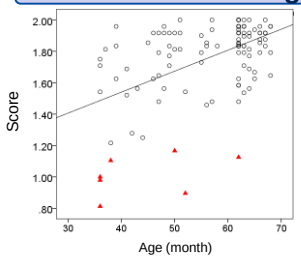
M = 1.27, *SD* = .33

Emotion matching



M = 1.76, *SD* = .33

Situation knowledge



M = 1.76, *SD* = .26

References

[1] N. Watanabe, S. A. Denham, N. M. Jones, T. Kobayashi, H. H. Bassett, D. E. Ferrier, "Working toward cross-cultural adaptation: Preliminary psychometric evaluation of the Affect Knowledge Test in Japanese preschoolers," *SAGE Open*, 2019.

[2] N. Watanabe, T. Kobayashi, "Computerization of an emotion knowledge assessment for preschoolers: Supporting their school readiness," *Proc. International School Psychology Association 41st Annual Conference*, 2019.

Contact

Naomi Watanabe Email: cs-openhouse-ml@hco.ntt.co.jp

Learning and Intelligent Systems Research Group, Innovative Communication Laboratory

オープンハウス 2020

11 Creating a personalized picture book

Support for parent-child picture book interaction

Abstract

Previous research on developmental psychology have extensively shown that picture book reading promotes child language development. To support a parent-child interaction for language development, we propose a method to create a personalized picture book that suits child's interest and vocabulary level. On the application site, parents answers the words that their child can produce and what they are interested in. Then, we try to estimate the words that the child is likely to produce based on statistical models in child vocabulary development database, to arrange each object picture for the estimated words, and to customize a story to match the child's interest. We are currently conducting a field trial in Onna village, Okinawa, to distribute a personalized picture book at a public library for parent and child who participate in health check-up. Through the field trial, we are investigating whether the personalized picture book contribute to the increase of parent-child interaction and book reading activity.

■ Personalized Picture Book

Akio Kashiwara, NTT Print Corp.

Target Age: 1-2 years of age

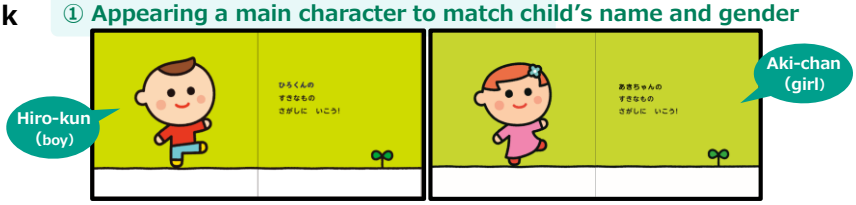
Enter personal information from site

https://ehon.nttprint.com

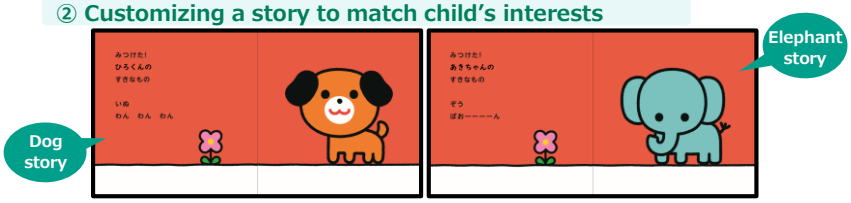
- child's name、boy/girl
- favorite objects
- vocabulary checklist (productive words)

Child Vocabulary Development Database

① Appearing a main character to match child's name and gender



② Customizing a story to match child's interests



③ Arranging pictures for words that the child is about to produce

■ Field Trial in Onna Village

with the corporation of NTT Print Corp

Feb 2020 – March 2021

- Lead parent and child to a public library
- Give opportunity to encounter picture books

Familiarize a child with books by personalized picture book

Inform parents at health check-up
(Onna Health & Welfare Center)

Apply for personalized picture book
(Onna Culture & Information Center)

Creating and sending the book for 10 days

Support book reading activity at home

References

[1] NTT, NTT Print Corp., Onnna Village, "Joint field trial about parent-child activity using personalized picture books in Onna Village, Okinawa," *Topics in NTT Group Web Site* (https://www.ntt.co.jp/topics/oyakoehon/index.html), 2020.

[2] NTT Print Corp. & Ehon Navi, "Parent-child interaction using personalized picture books created from psychological research in child language development," *Ehon Navi* (https://style.ehonnavi.net/ehon/2020/03/17_358.html), 2020.

Contact

Tessei Kobayashi Email: cs-openhouse-ml@hco.ntt.co.jp

Interaction Research Group, Innovative Communication Laboratory

オープンハウス 2020

12 How many words do you know?

– Vocabulary size test, *Reiwa* edition –

Abstract

More than 20 years ago, NTT conducted psychological experiments to investigate word familiarity and thus construct a Japanese lexicon of about 77,000 words. This lexicon is still used in many fields.

Now, we reinvestigated and reconstructed an unparalleled-scale Lexicon of 163,000 words through crowdsourcing. By applying careful screening, we succeeded in obtaining highly reliable results. This makes it possible to make comparisons with results of 20 years ago. Based on this, we created the *Reiwa* edition of a vocabulary-size estimation test. Furthermore, this lexicon allows us to estimate the vocabulary size appropriate for the present day. We are working on vocabulary-size investigations for a wide range of ages, including elementary school to high school students. In the future, we will investigate and analyze the relationship between vocabulary size, reading comprehension, and academic ability, aiming to achieve effective educational support.

Word Familiarity Examination via Cloud

Ask good workers to evaluate other sets

Pre-screening
Cloud workers

Holder of official Kanji test, level 2 or above, or top performers in original Kanji reading test

Rating word familiarity

Evaluate 1000–2000 words/set at a time (5% of words are duplicated)

Post-screening
Cloud workers

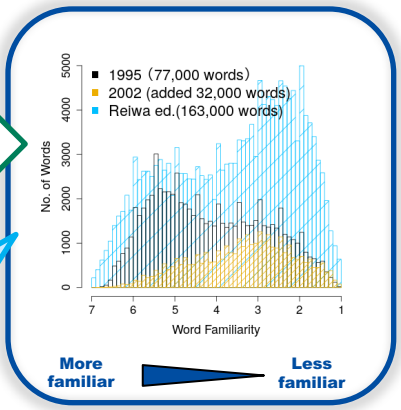
Correlation coefficient of duplicated evaluation is more than 0.5

Point ① Earnest and eager evaluators who have knowledge of Kanji

Point ② About 163,000 target words, including new words, e.g., smartphone, internet, heterodimer, laba…

Word Familiarity Database, *Reiwa* edition

Comparison of Histograms with previous data

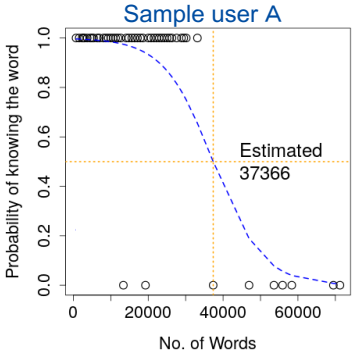


Vocabulary-size estimation

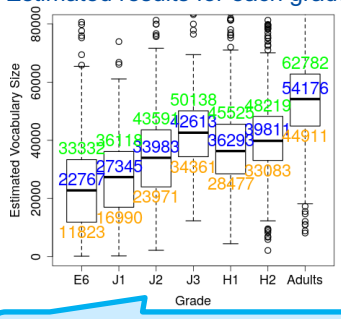
Please select words that you know.
- ☐ bank
- ☐ charge
- …
- ☐ intimacy
- ☐ recast

Point ③ Checklist of only about 50 words

Estimated vocabulary size, Sample user A



Estimated results for each grade



Point ④ About 4,600 persons, including elementary/junior high/high school students

Vocabulary-size test:
http://www.kecl.ntt.co.jp/icl/lirg/resources/goitokusei/

References

[1] S. Fujita, T. Kobayashi, "Reexamination of word familiarity and comparison with past examination," *The 26th Annual Meeting of the Association for Natural Language Processing*, pp. 1037-1040, 2020.

[2] S. Fujita, T. Kobayashi, T. Yamada, S. Sugawara, T. Arai, N. Arai, "Vocabulary size of elementary, junior high and high school students and analysis of relationship with word familiarity," *The 26th Annual Meeting of the Association for Natural Language Processing*, pp. 355-358, 2020.

Contact

Sanae Fujita, email: cs-openhouse-ml@hco.ntt.co.jp
Linguistic Intelligence Research Group, Innovative Communication Laboratory

オープンハウス 2020

13 Kyomachi Seika will guide you!

Training the role-play AI with community cooperation

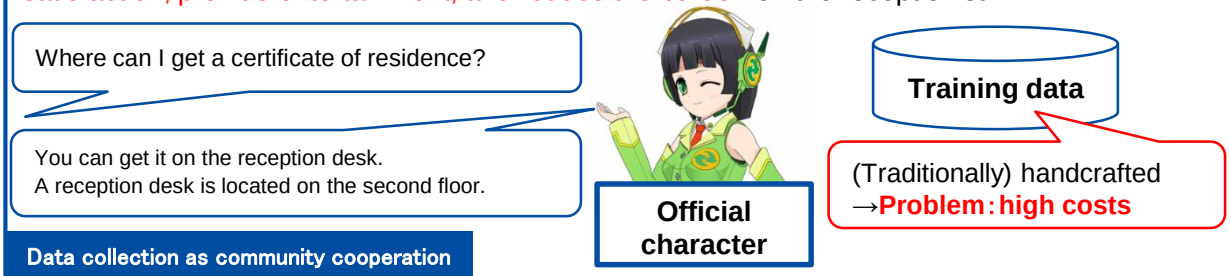
Abstract

We propose a **“role-play AI” that can respond to inquiries, receptions, and guides at city offices.** Conventional AI training requires a lot of accurate training data, and data collection has been an extremely costly and difficult task. In this study, we solve this problem by utilizing community cooperation. By making the data collection work a community cooperation activity, we collect accurate training data at a low cost. By **connecting “people who live in the area” and “people who interested in the area,” we have collected very high-quality training data**, and we have made possible the training of role-play AI that is closely linked to the community. By using this technology, we provide “role-play AI” to learn according to local demand. In the future, we aim to realize AI technology that can be used in a wide range of situations, not only by local governments.

The efficiency of business operations is being improved through AI. In particular, the realization of AI, which provides information to users through dialogues such as receptionists and guide, is attracting attention as a means of reducing the burden on human and promoting collaboration with people and AI. In this research, we realize the useful AI in the real world through joint experiments with Seika-town.

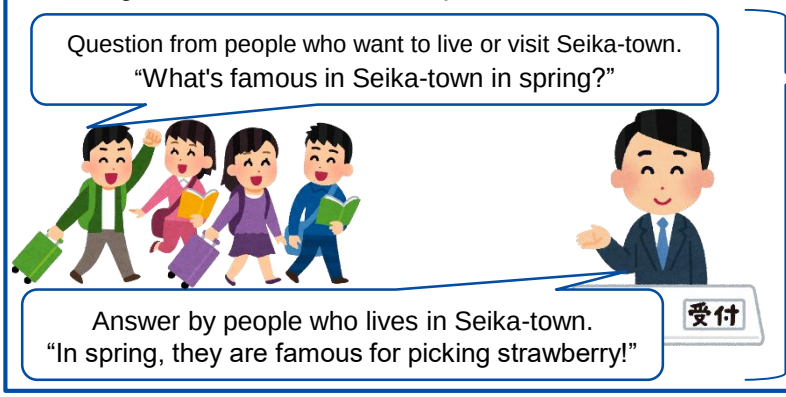
Task-oriented dialogue with the role-play AI

"Role-play AI" is a dialogue system that imitates the character. Having an official corporate character or a celebrity serve as a receptionist or guide is expected to improve customer satisfaction, provide entertainment, and reduce the burden on the receptionist.



Data collection as community cooperation

We solved the problem of the data collection by matching "people with interest" and "people with knowledge“. →The data collection by residents of Seika-town is carried out as community cooperation.



■POINTS:
Training data collection
・Question from people who interested
＝Demand information
・Answer by people who lives in Seika-town.
＝Accurate information
Data collection with community cooperation
＝Low cost and high quality

References

[1] R. Higashinaka, et al. "Role play-based question-answering by real users for building chatbots with consistent personalities." *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*. 2018.

Contact

Masahiro Mizukami Email: cs-openhouse-ml@hco.ntt.co.jp
Interaction Research Group, Innovative Communication Laboratory

オープンハウス 2020

14 What does he/she think in this situation?

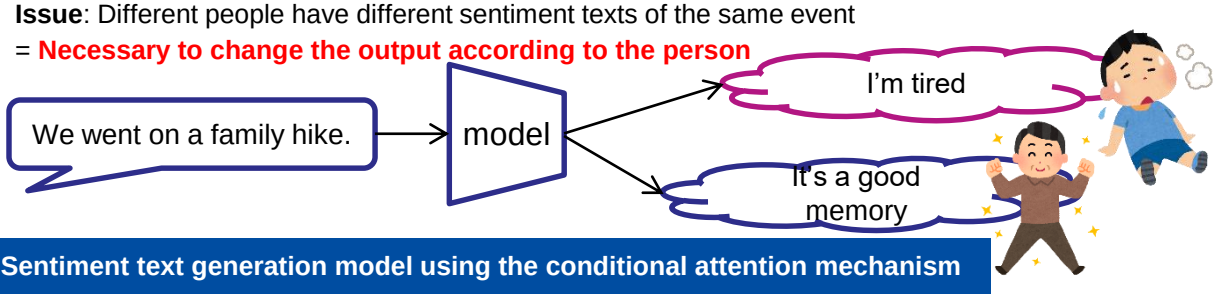
Sentiment text generation based on personality

Abstract

In order for a dialogue system to have a good relationship with humans, it is essential to understand and express human emotions. In previous studies, the expression of emotions has been done in typologies such as joy and anger. In contrast, this research will realize a dialogue system that can better understand and express human emotions by predicting the feelings associated with people like "what does he/she think this situation." This model take "who" and "what they did" as input and estimate the output of "how they feel" according to a particular "person". This allows the dialogue system to use a more flexible way of expressing impressions, depending on the person. In the future, we will realize a method that learns "characteristics of the expression of impressions" from users in real time during the dialogue, and estimates the kind of person from the user's characteristics of the expression of impressions.

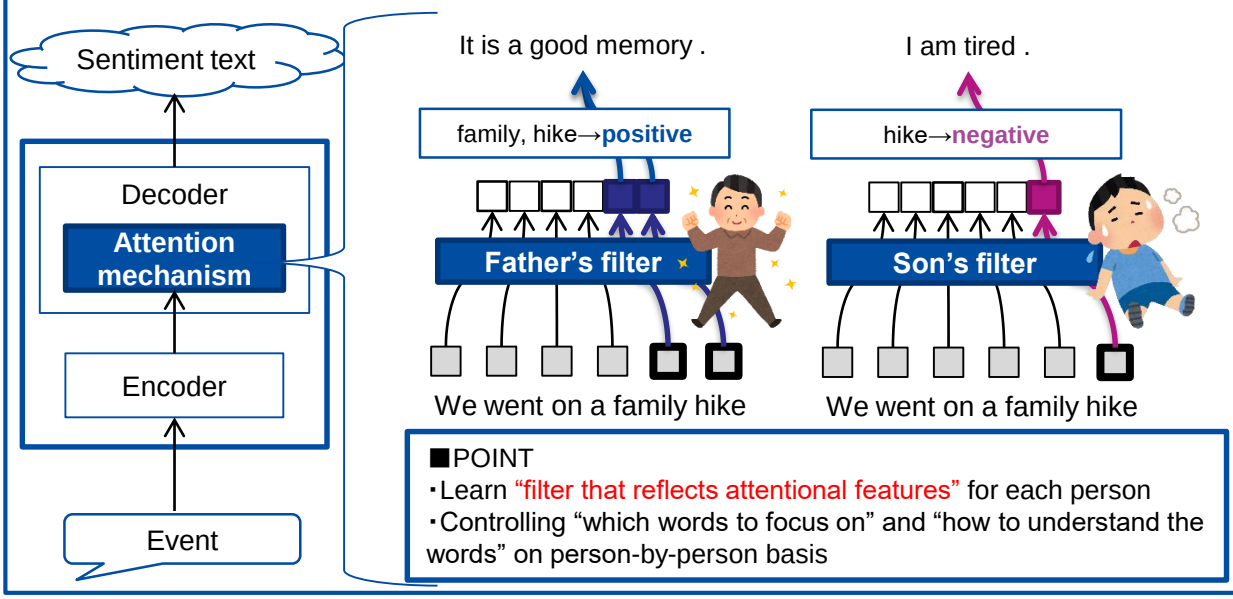
Sentiment text generation is a technique for estimating how he/she feel when an event occurs.

Previous works: Estimating whether the majority believes it positive or negative
Issue: Different people have different sentiment texts of the same event
= **Necessary to change the output according to the person**



Sentiment text generation model using the conditional attention mechanism

We focus on **the attention mechanism** → Introduce a filter that reflects personality



■POINT
・Learn "filter that reflects attentional features" for each person
・Controlling "which words to focus on" and "how to understand the words" on person-by-person basis

References

[1] M.Mizukami, H. Sugiyama, H. Narimatsu, "Event data collection for Recent Personal Questions," in *Proc. LACATODA*, 2018.
[2] M.Mizukami, H. Sugiyama, H. Narimatsu, T. Arimoto, R. Higashinaka, "話者情報を考慮する注意機構を用いた応答生成手法の検討," *人工知能学会*, 2020.

Contact

Masahiro Mizukami Email: cs-openhouse-ml@hco.ntt.co.jp
Interaction Research Group, Innovative Communication Laboratory

オープンハウス 2020

15 Can you guess the age from this voice?

Deep speaker attribute estimation with speaker clustering

Abstract

Estimating speaker-attributes such as age and gender is an important task with a wide range of applications. While the recent proposed deep neural network models have been achieving high performance, the estimated results tend to be less reliable because of the overfitting problem. In order to solve this problem, we propose a general framework for correcting the unreliable results of the arbitrary speaker-attribute estimation models. The proposed algorithm first applies speaker clustering to the target utterances to detect similar speakers of target utterances. Then, the speaker-attribute class of each cluster is determined by voting on the utterances assigned to the cluster. Finally, we can correct the result of unreliable utterances by replacing their result with the clusters' speaker-attribute class. Our approach is evaluated on age-gender classification and gender regression tasks, yielding significant improvements in classification accuracy and mean absolute error.

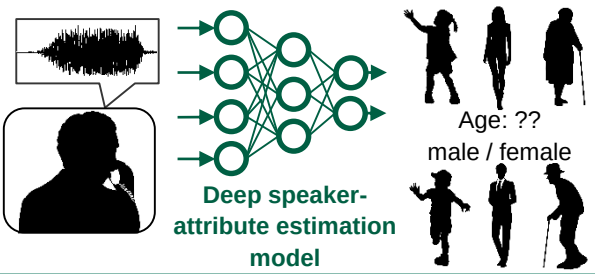
Deep speaker attribute estimation

Problem definition

Estimate speakers' age and gender from their voices with deep learning model

Applications

Call center response decision, marketing support, realization of a voice dialogue system that changes behavior according to user attributes, etc.



Difficulties of task

- There is **large deviation in amount of training data for each age-group** (Fig.1)
- **Models tend to be overfitting to specific speaker / age** (Fig. 2)

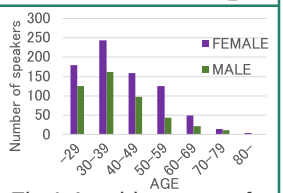


Fig.1 Age histogram of NIST-SRE08

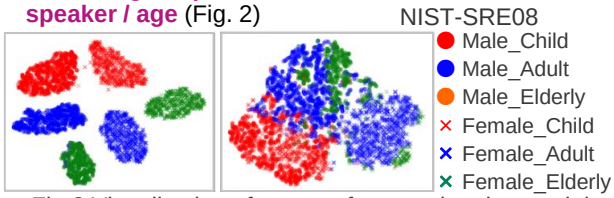
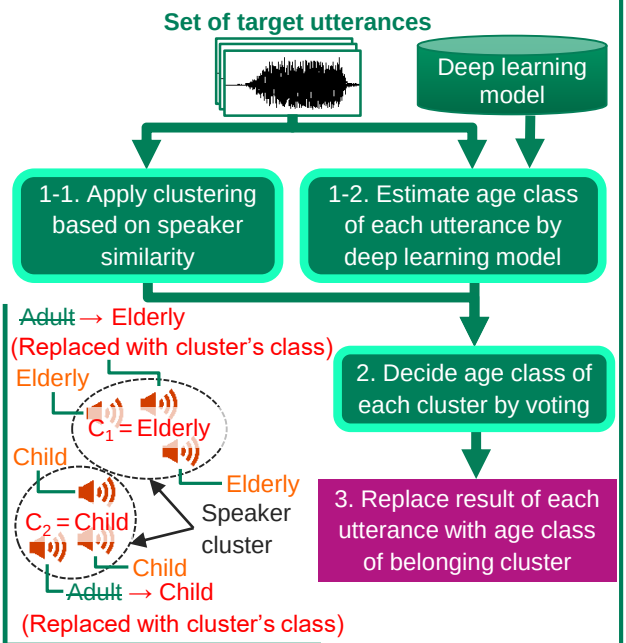


Fig.2 Visualization of output of age-estimation model (cited from [1]) (left: training data, right: evaluation data)

Error correction based on speaker clustering

Correct estimation results of deep learning model by majority voting of results of similar speakers



: Target speech and estimated class

C.: Class decided by voting for each cluster

Evaluation criteria	Conventional	Proposed
Classification accuracy (Child, Adult, Elderly)	59 %	**72 %**
Mean absolute error	± 10.9 years	**± 8.7 years**

References

[1] N. Tawara, H. Kamiyama, S. Kobashikawa, A. Ogawa, "Improving speaker-attribute estimation by voting based on speaker cluster information," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 6594-598, May 2020.

[2] N. Tawara, H. Kamiyama, S. Kobashikawa, A. Ogawa, "Frame-level phoneme-invariant speaker feature extraction for text-independent speaker recognition on extremely short utterances," *Reports of the autumn meeting the Acoustical Society of Japan.*, pp. 815-816, Sept. 2019.

Contact

Naohiro Tawara Email: cs-openhouse-ml@hco.ntt.co.jp

Signal Processing Research Group, Media Information Laboratory

オープンハウス 2020

16 More wireless microphones are available in a room

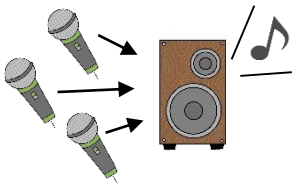
BRAVE: Bit-error-robust low-delay audio and voice encoding

Abstract

We have developed a bit-error-robust speech-and-audio codec working in low-delay conditions. Inter-device audio transmission, such as in microphones, requires strict real-time processing. It is a challenge to enhance the compression efficiency in such conditions, which enables us to use more microphones at once in a room. Sometimes, this kind of inter-device transmission encounter errors occurring in the encoded data, and codecs have to deal with them to avoid severe decoding errors. Especially in low-delay conditions, it is hard to protect codes with additional information keeping the bitrates. Therefore, we proposed a bit rearrangement technique, which makes lower the impact of the errors compressing data efficiently. Using this technique, the developed codec BRAVE can compress speech and audio data in a very short time and is robust for bit errors. It is thus expected to be useful also for other use cases such as the Internet of things (IoT).

Speech and audio codec BRAVE

We developed a speech-and-audio coding scheme for real-time inter-device audio transmission like wireless microphones



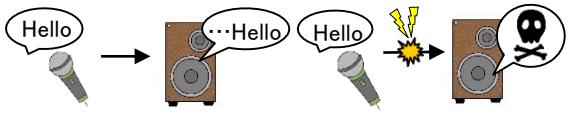
Input: 48-kHz 16-bit monaural
Bitrate: 96 kbps (About half of the conventional one)
Algorithmic delay: 1.5–3.0 ms

Lower bitrates allow us to use more microphones at once

Technical difficulty

Errors may happen in the codes during transmission
- Many codecs protect codes by using packets
- Packets need frame-wise headers
- Headers weigh too much in low-delay conditions

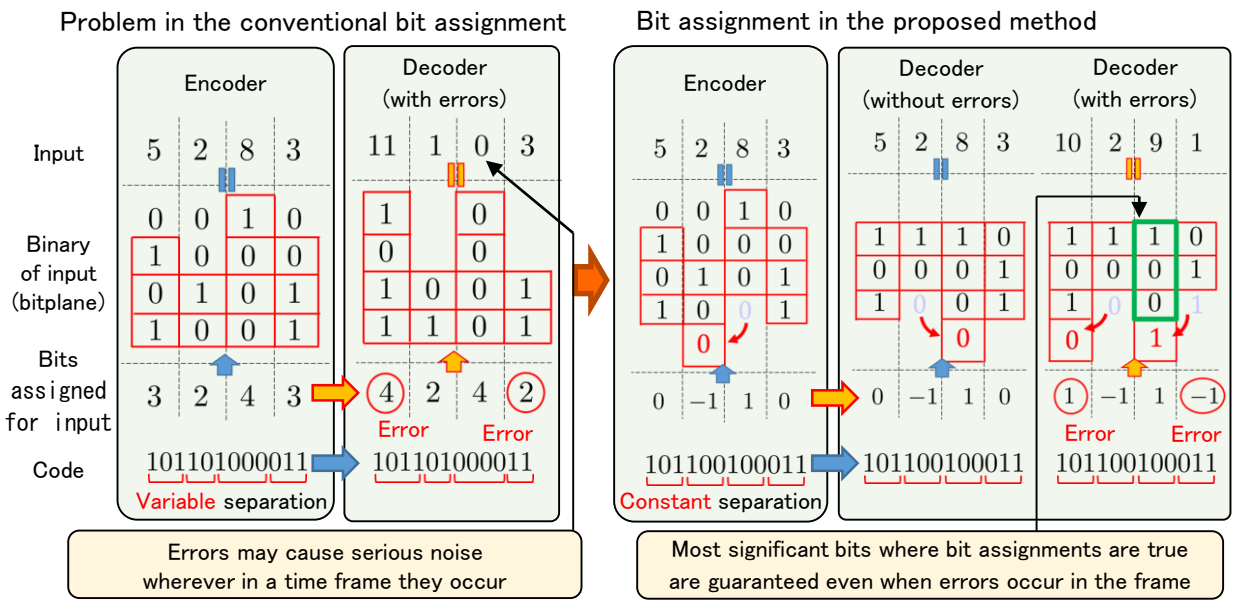
→ We want to deal with errors without using packets



Short-time processing　　Suppressing error effects

Proposal: Bitplane rearrangement

Avoiding serious impacts on the sound quality



References

[1] R. Sugiura, Y. Kamamoto, T. Moriya, "Spectral-envelope-based least significant bit management for low-delay bit-error-robust speech coding," *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018.

Contact

Ryosuke Sugiura Email: cs-openhouse-ml@hco.ntt.co.jp
Moriya Research Laboratory

オープンハウス 2020

17 Pay attention to the speaker you want to listen to (II)

Neural selective hearing with audio-visual speaker clues

Abstract

Human beings have the ability to concentrate on listening to a desired speaker (= selective hearing) even when multiple people are speaking at the same time. The purpose of this research is to realize the selective listening mechanism of human beings on a computer. In this research, we propose multimodal selective hearing technology that uses video information as the target speaker's clues in addition to audio information. By utilizing multiple information sources like humans, the technology become advanced that can operate stably even in situations, where audio clues are useless, such as conversations between speakers with similar voice characteristics. This technology will become fundamentals of various devices that take human voice as input. For example, it will contributes to the realization of robots and smart speakers that recognize people and change their response.

Selective Hearing with Audio Speaker Clue

Selective Hearing

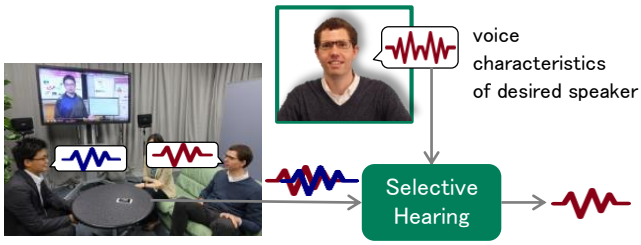
- Ability to focus on listening to desired speaker from mixture signals
- In daily conversations, multiple speakers often speak at same time

⇨ Humans easily perform such selective hearing, but it is difficult for conventional computers

⇨ First proposal of neural selective hearing with audio speaker clue (OPEN HOUSE 2018)

Problem

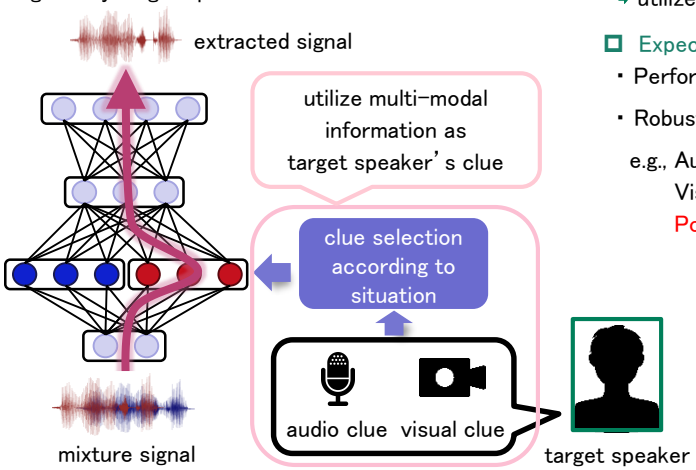
With audio clues, extraction performance degrades for mixture signals with similar voice characteristics



Utilization of Audio and Visual Speaker Clues

SpeakerBeam (= Selective Hearing based on Deep Learning)

Deep learning–based model, which extracts desired speaker's voice from mixture signal given by target speaker's clue



Solution: Proposal of Multimodal SpeakerBeam

In addition to voice characteristics (audio info.), use mouth motion (visual info.) as speaker clues

⇨ utilize multi–modal information like humans

Expected effect

- Performance improvement by utilizing multiple modality
- Robustness improvement against lack of speaker clues

e.g., Audio clue is useless (similar voice characteristics)
Visual clue is missing (face not detected)
Possible to extract even in above situations

＊ About target speaker's clue

Pre–recorded audio data of target speaker

Video data (around mouth) recorded same time as mixture signal

References

[1] T. Ochiai, M. Delcroix, K. Kinoshita, A. Ogawa, T. Nakatani, "Multimodal SpeakerBeam: Single channel target speech extraction with audio-visual speaker clues," *Proc. Interspeech*, 2019.

[2] K. Zmolikova, M. Delcroix, K. Kinoshita, T. Ochiai, T. Nakatani, L. Burget, J. Cernocky, "SpeakerBeam: Speaker aware neural network for target speaker extraction in speech mixtures," *IEEE Journal of Selected Topics in Signal Processing,* 2019.

Contact

Tsubasa Ochiai Email: cs-openhouse-ml@hco.ntt.co.jp
Signal Processing Research Group, Media Information Laboratory

オープンハウス 2020

18 Controlling voice expression using face expression

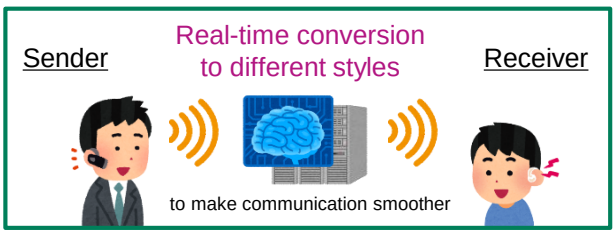
Crossmodal voice expression control

Abstract

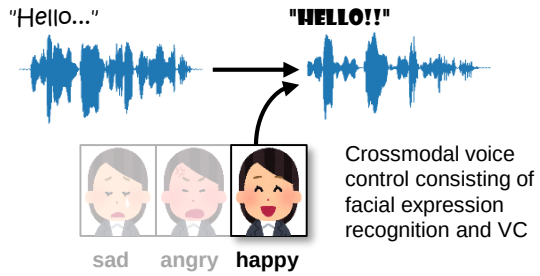
There are many kinds of physical or mental barriers that prevent individuals from smooth verbal communication. One key technique to overcome some of these barriers is voice conversion (VC), a technique to convert para/non-linguistic information contained in a given utterance without changing the linguistic information. Here, we propose a crossmodal voice control system, which offers a way to control the vocal expression of emotion in speech through the facial expression in a face image. The proposed system consists of performing facial expression recognition (FER) followed by VC. For VC, we have developed a method based on sequence-to-sequence (S2S) learning, which is designed to convert the prosodic features as well as the voice characteristics in speech conditioned on the output of the FER system. We believe that this work can provide some insight on what it is like to be able to control our voice through different modalities.

Communication augmentation system

Using voice conversion (VC) technique to help overcome barriers that prevent us from smooth communication

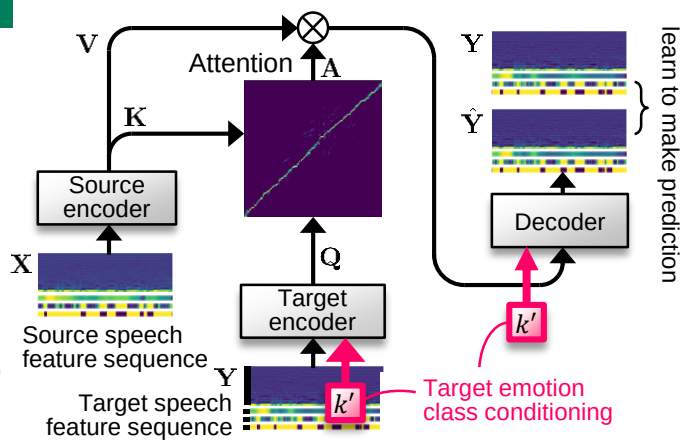


Voice expression control through face



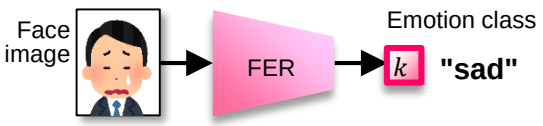
Conditional sequence-to-sequence VC

- Sequence-to-sequence (S2S) learning
  - Offers a general framework for transforming one sequence into another variable length sequence
  - Encoder/decoder structure and attention mechanism make it possible to learn conversion rules that reflect long-term dependencies in input/output sequences
  - Usually requires large-scale parallel corpora
- VC based on S2S learning (S2S-VC) [1,2]
  - Voice expressions are characterized by prosodic features (e.g., intonation and rhythm)
  - S2S-VC is able to convert prosodic features as well as voice characteristics in input speech with limited amount of training data



Facial expression recognition (FER)

- FER using attentional convolutional network
  - After prediction, output is passed to VC system



References

[1] H. Kameoka, K. Tanaka, T. Kaneko, N. Hojo, "ConvS2S-VC: Fully convolutional sequence-to-sequence voice conversion," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, in Review.

[2] K. Tanaka, H. Kameoka, T. Kaneko, N. Hojo, "AttS2S-VC: Sequence-to-sequence voice conversion with attention and context preservation mechanisms," *Proc. 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2019)*, pp. 6805-6809, May 2019.

Contact

Hirokazu Kameoka Email: cs-openhouse-ml@hco.ntt.co.jp
Media Recognition Research Group, Media Information Laboratory

オープンハウス 2020

19 Learning to search like human

Adaptive spotting for efficient object search

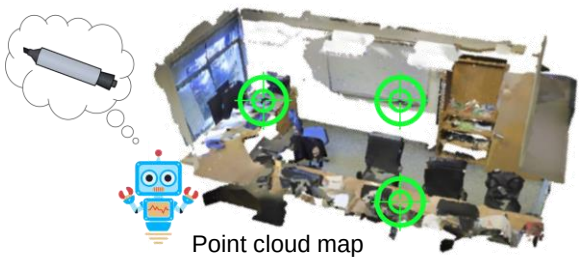
Abstract

We propose Adaptive Spotting, a deep reinforcement learning approach to object search from a scene represented by a 3D point cloud map. A straightforward approach using exhaustive search is often not promising due to poor computational efficiency. To solve this problem, our approach simultaneously learns the features of a given object and its efficient search path. Our network is designed to have a pose estimation module to estimate promising locations to be explored. The network is trained in an end-to-end manner to learn efficient search paths by using a reinforcement learning strategy that gives a higher reward when it finds the target in fewer search steps. Evaluation results demonstrate that our approach outperforms several state-of-the-art methods in both search accuracy and the number of search steps required. It is expected to be used in areas such as logistics, manufacturing, and transportation, which require the ability to search for objects in 3D space fast and accurately.

Point Cloud Search

Search for object with certain shape in point cloud map of space captured by 3D sensor (LiDAR, etc.)

Space often huge and non-uniform
⇒ Exhaustive search is often undesirable.

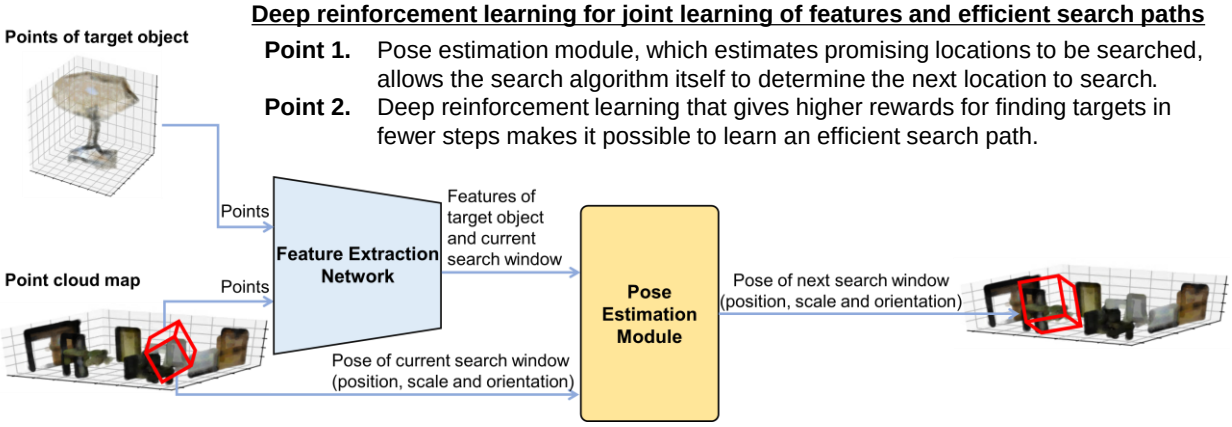


Adaptive Spotting

Deep reinforcement learning for joint learning of features and efficient search paths

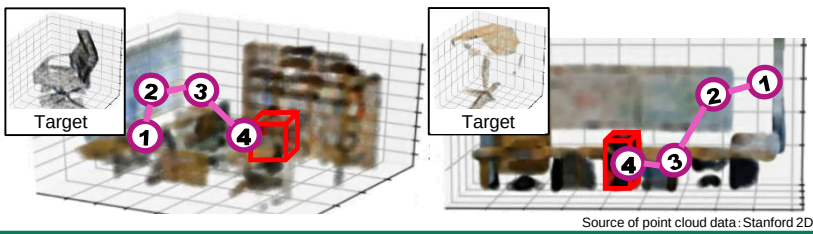
Point 1. Pose estimation module, which estimates promising locations to be searched, allows the search algorithm itself to determine the next location to search.

Point 2. Deep reinforcement learning that gives higher rewards for finding targets in fewer steps makes it possible to learn an efficient search path.



Examples of Search Paths (Searching chair in office room)

Possible to find target within a few steps



Performance improved significantly

Source of point cloud data : Stanford 2D-3D-Semantics Dataset http://buildingparser.stanford.edu/dataset.html

References

[1] O. Krishna, G. Irie, X. Wu, T. Kawanishi, K. Kashino, "Learning search path for region-level image matching," *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019.

[2] O. Krishna, G. Irie, X. Wu, T. Kawanishi, K. Kashino, "Deep reinforcement template matching," *Meeting on Image Recognition and Understanding (MIRU)*, 2019.

Contact

Onkar Krishna Email: cs-openhouse-ml@hco.ntt.co.jp
Recognition Research Group, Media Information Laboratory

オープンハウス 2020

20 Deep learning without data aggregation from nodes

Asynchronous consensus algorithm on distributed networks

Abstract

While use of massive data benefits on training DNN models, aggregating all data into one physical location (e.g. a cloud data center) may not be possible due to data privacy concerns from consumers. For example, according to EU GDPR, it is preferable to minimize data transmission between processing nodes.

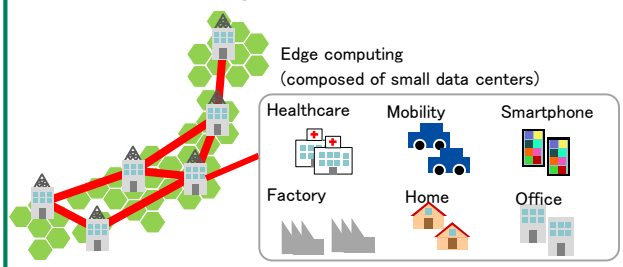
Our goal is to construct training algorithms to obtain a global DNN model that can be adapted to all data, even when individual nodes only have access to different subsets of the data. We assume that this algorithm is allowed to communicate autonomously between nodes, exchanging information such as model variables or their update difference, but data are prohibited from being moved from node they reside on.

Now, several platformers provide advanced services by aggregating/monopolizing data. However, we aim to create a society where data ownership belongs to individual and can be used for a variety of services while protecting data privacy.

Background, goal

Background: We are entering an era of distributed aggregation of due to data volume, privacy protection and legal regulations (e.g. GDPR).

Goal: To obtain a global DNN model without data aggregation (where asynchronous communication among nodes, such as model variable exchange, is allowed).



Problem

Problem: When the data at each node is statistically heterogeneous, a global DNN model cannot be obtained by just minimizing each nod cost function.

Approach: We solve problem by minimizing sum of cost functions under a consensus constraint that all node models are identical with each other.

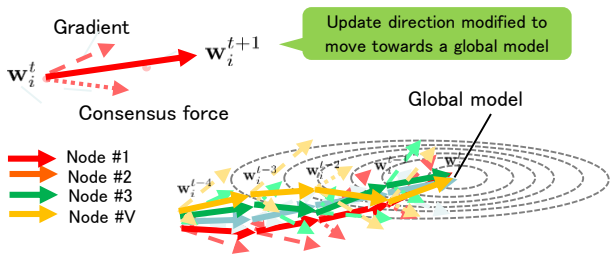
$$\inf_{\{\mathbf{w}_i | i\in\mathcal{V}\}} \sum_{i\in\mathcal{V}} F_i(\mathbf{w}_i;\mathbf{x}_i)$$

$$\text{s.t. } \mathbf{A}_{i|j}\mathbf{w}_i + \mathbf{A}_{j|i}\mathbf{w}_j = \mathbf{0} \qquad \mathbf{A}_{i|j} = \begin{cases} \mathbf{I} & (i>j,\ j\in\mathcal{N}(i)) \\ -\mathbf{I} & (j>i,\ j\in\mathcal{N}(i)) \end{cases}.$$

Data sets are placed across V nodes ($x_1,\cdots,x_V$). Model variables are updated (i) such that minimizes sum of cost functions (Σ Fi) (ii) under a consensus constraint that models are identical among V nodes (s.t.....)

Asynchronous consensus algorithm

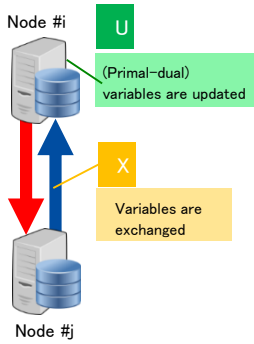
Proposed algorithm: A training algorithm is constructed to obtain a global model by asynchronously exchanging primal model variables and Lagrangian dual variables among nodes. (this enables to work on arbitrary network structure).

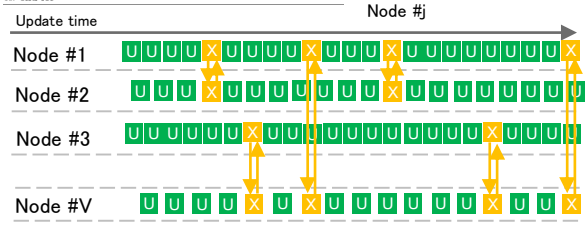


Proposed algorithm

Algorithm 1 PDMM SGD/ADMM SGD

1: Initialization of $\tilde{\mathbf{z}}_{i|j}^{(0)}, \mathbf{w}_i^{(0)}$
2: **for** $k \in \{0,\dots,K-1\}$ **do**
3: ▷ Step 1: Update model for each node
4: **for** $i \in \mathcal{V}$ **do**
5: $\mathbf{w}_i^{k+1} \leftarrow \Big(\mu\mathbf{w}_i^k - \nabla F_i(\mathbf{w}_i^k;\mathbf{x}_{i,\mathrm{MB}}^k) + \sum_{j\in\mathcal{N}(i)}\big(\alpha\mathbf{A}_{i|j}^{\mathrm{T}}\tilde{\mathbf{z}}_{i|j}^k + \gamma\mathbf{w}_j^k\big)\Big)\Big/\big(\mu+\alpha|\mathcal{N}(i)|+\gamma|\mathcal{N}(i)|\big)$
6: **for** $j \in \mathcal{N}(i)$ **do**
7: $\tilde{\mathbf{y}}_{i|j}^{k+1} \leftarrow \tilde{\mathbf{z}}_{i|j}^k - 2\mathbf{A}_{i|j}\mathbf{w}_i^{k+1}$
8: **end for**
9: **end for**
10: ▷ Step 2: Exchange and update variables at random time k
11: **for** $i \in \mathcal{V}$ **do**
12: Select $j \in \mathcal{N}(i)$ at random
13: **Transmit** $j \mapsto (\mathbf{w}_j^{k+1}, \tilde{\mathbf{y}}_{j|i}^{k+1})$
14: $\tilde{\mathbf{z}}_{i|j}^{k+1} \leftarrow \tilde{\mathbf{y}}_{j|i}^{k+1}$ (PDMM SGD) / $\tilde{\mathbf{z}}_{i|j}^{k+1} \leftarrow \theta\tilde{\mathbf{y}}_{j|i}^{k+1} + (1-\theta)\tilde{\mathbf{z}}_{i|j}^k$ (ADMM SGD)
15: **end for**
16: **end for**





References

[1] T. Sherson, R. Heusdens, and B. Kleijn, Derivation and analysis of the primal-dual method of multipliers based on monotone operator theory, *IEEE transactions on signal and information processing over networks 5*, 2 (2018), 334–347.

[2] K. Niwa, N. Harada, G. Zhang, B. Kleijn, Edge-consensus learning: deep learning on P2P networks with nonhomogeneous data, submitted to KDD 2020.

Contact

Kenta Niwa Email: cs-openhouse-ml@hco.ntt.co.jp
Communication Science Laboratory, Media Intelligence Laboratory

オープンハウス 2020

21 Cardiac model that makes it heart

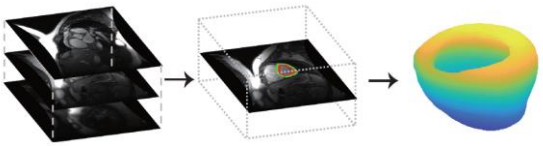
Gaussian process with physical laws for 3D cardiac modeling

Abstract

Cardiovascular disease is one of the leading causes of both morbidity and mortality all over the world. Early diagnosis and treatment planning are demanded for the wide variety of etiologies and pathophysiologies. In the last decades, intensive research in the field of computational biology has demonstrated the potential ability of three-dimensional (3D) cardiac computational models to give us a clue to perform early diagnosis or to have high affinity with machine learning for treatment planning. We introduces some physical laws into a Gaussian process for a statistical 3D cardiac computational model. The heart shape must be ruled by some physical laws, which should be an important clue for the statistical shape estimation. For demonstration, we apply our model into the pipeline that estimates the heart shape from cardiovascular magnetic resonance (CMR) imaging, by combining it with the deep neural networks-based anatomical segmentation of CMR imaging.

Heart shape estimation

- Cardiac modeling from magnetic resonance imaging (MRI)



MRI → Anatomical classification → Shape estimation

Deep learning for classification

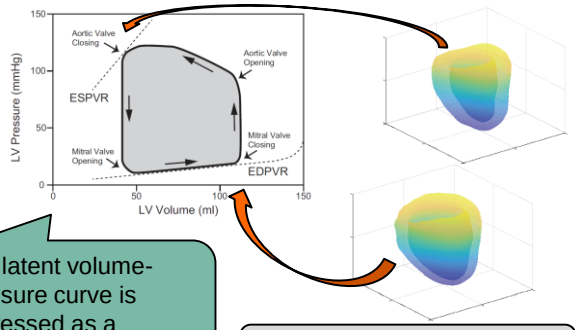
Statistical shape modeling for regression

State-of-the-art deep nets [2]

Our contribution: Gaussian process (GP) with physical laws for the heart shape [1]

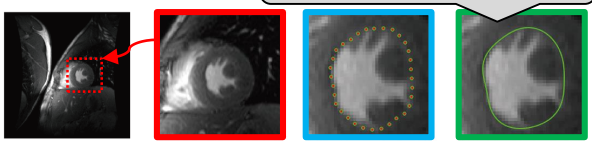
Regression with physical laws

- The Frank-Starling law is introduced into a Gaussian process-based statistical shape model.



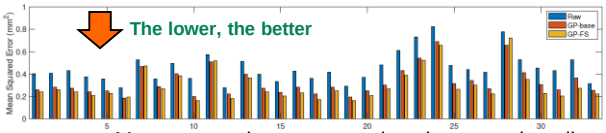
The latent volume-pressure curve is expressed as a hidden Markov model.

Our method can obtain the similar shape to that handles by experts.



From left to right. **First**: MRI. **Second**: Zoom-in. **Third**: The endocardium manually handled by experts. **Fourth**: Predictive endocardium.

Goal: shape prediction comparable to experts



Mean squared error comparison between baseline methods and our model.

For 29 out of 33 subjects, the mean squared error is improved by around 9%.

- Raw: Regression without GP fitting.
- GP-base: GP without physical laws.
- GP-FS: GP with the Frank-Starling law.

Difficulty

- It is not easy to obtain training datasets for the heart shape, which is typically handled by experts.
  - **We introduce physical law into unsupervised learning.**
- It is difficult to estimate details of the heart shape, since MRI provides only statistical information locally averaging over time and space.
  - **We extend the static model into s time-varying statistical shape model.**

References

[1] M. Nakano, R. Shibue, K. Kashino, S. Tsukada, H. Tomoike, "Gaussian process with physical laws for 3D cardiac modeling," under review.

[2] T. Ngo, Z. Lu, G. Carneiro, "Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance," *Medical Image Analysis*, Vol. 35, pp. 159–171, 2019.

Contact

Masahiro Nakano Email: cs-openhouse-ml@hco.ntt.co.jp

Recognition Research Group, Media Information Laboratory

オープンハウス 2020

22 Listening carefully to your heart beat

Cardiohemodynamical analysis based on stethoscopic sounds

Abstract

A variety of sounds are constantly emitted from the human body as a result of life activities. By listening to and analyzing those sounds, we can obtain useful information about the function and condition of the body, which is called auscultation. In this research, we are focusing on heart sounds to estimate the function and condition of the heart and blood vessels based on the observation of acoustic signals. In our system, multiple microphones are attached to several places, such as the chest, to detect heart activity. Based on the captured sound, it estimates the degree of normality as a score and generates an explanatory statement as a sentence. We have confirmed that the normality estimation and description generation with a specified degree of detail work effectively for test data. We aim to realize an "AI stethoscope" that contributes to the prevention and early detection of diseases in many people, as skilled doctors can accurately understand and explain the condition through auscultation.

Concept of AI Stethoscope

- Multiple small microphones are attached to your body to collect useful information and visualize it in various ways.
- By our machine learning techniques, advanced media conversion such as text generation from audio [1] is possible, in addition to abnormality detection or pattern classification.
- The system will be extended to a visualizing and analyzing tool for heart activity and hemodynamics, which is a part of the "digital twin computing" concept that we are pursuing.

Figure 1: Prototype of the Heart Sound Collector

Generation of Explanatory Text and Score

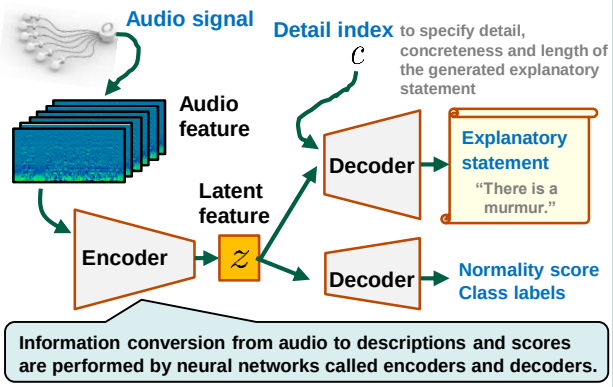


Figure 2: Sequence Conversion Model in This Study [1]

Working Example of Prototype

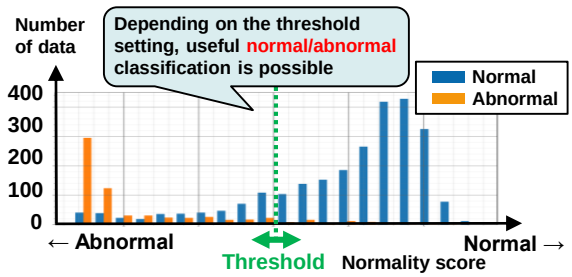


Figure 3: Distribution of Normality Scores for Test Data [3]

Table 1: Generated Description Examples

Depending on the specified level of detail, useful description is generated.

C index	Example of generated text
20	Your heart sounds are abnormal.
40	There is a systolic murmur in the heart sound.
60	There is a systolic murmur in the heart sound, and it may be a sign of cardiomyopathy.

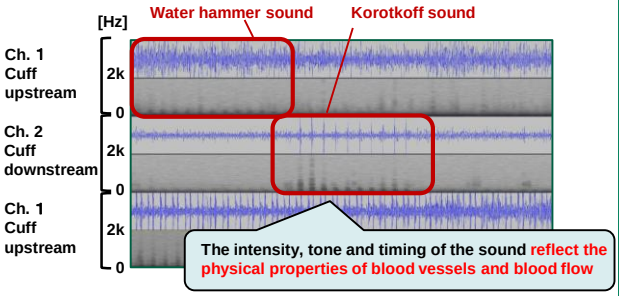


Figure 4: Measurement Example of Heart Sound and Blood Flow Sound during Blood Pressure Measurement in the Left Upper Arm
(Each Ch. Top: Sound signal waveform, Bottom: Sound spectrogram)

Part of this research is commissioned to NTT Research, Inc. (Medical and Health Informatics Laboratories; MEI Lab.) by NTT.

References

[1] S. Ikawa, K. Kashino, "Neural audio captioning based on conditional sequence-to-sequence model," In *Proc. DCASE 2019 Workshop*, 2019.
[2] M. Nakano, R. Shibue, K. Kashino, S. Tsukada, H. Tomoike, "Gaussian process with physical laws for 3D cardiac modeling," under review.
[3] The PhysioNet Computing in Cardiology Challenge, *http://physionet.org/content/challenge-2016/1.0.0/*, 2016.

Contact

Kunio Kashino Email: cs-openhouse-ml@hco.ntt.co.jp
Media Information Laboratory and Biomedical Informatics Research Center

オープンハウス 2020

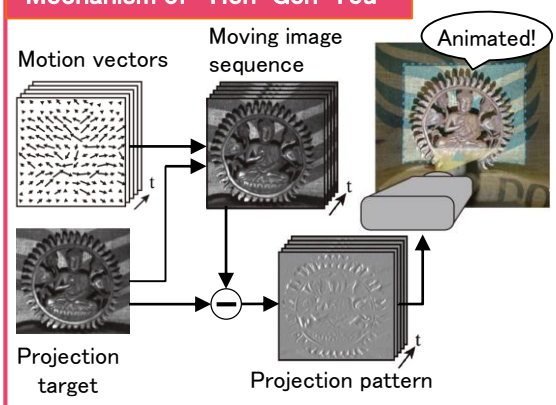
23 Make natural-looking illusions by perceptual model

Adaptive motion retargeting for illusion-based projection AR

Abstract

Hen-Gen-Tou, invented by CS Labs, is an illusion-based projection mapping that adds motion impressions to real static objects. It produces illusory motion impressions in the projection target by projecting luminance motion signals that selectively drive the motion detectors in the human visual system. However, in order to successfully "fool" human vision, the amount of movements must be properly adjusted because there is a limit in shift size that can create the illusion. Here, to automate this laborious adjustment task, we propose an optimization framework that adaptively retargets the motion information in real time based on a perceptual model. The perceptual model predicts the perceived deviation of a projected pattern from an original surface pattern using a computational model of human visual information processing. This technique will broaden the range of applications of Hen-Gen-Tou, including interactive applications.

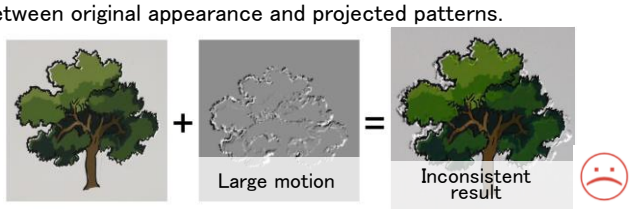
Mechanism of "Hen-Gen-Tou"



The human visual system separately analyzes color, pattern, and motion information and later integrates them. Taking advantage of this characteristic, HenGenTou adds illusory motion impressions on static objects by projecting only luminance motion signals.

Problem: shift limit

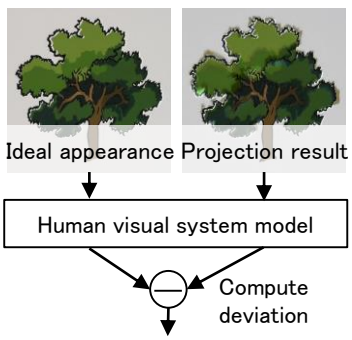
Trying to add too large motion leads to **subjective inconsistency** between original appearance and projected patterns.



Manual adjustments? ➡ Impossible for interactive applications

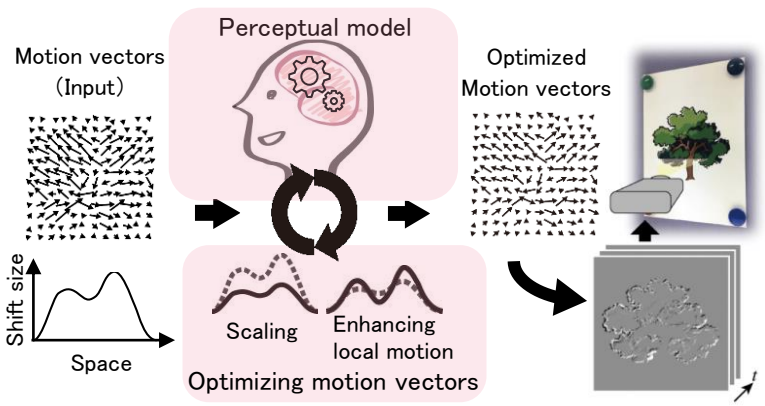
Perceptual model to predict subjective inconsistency level

Predict subjective inconsistency level of the projection result as a deviation from the ideal appearance in representation space of the human visual system.



Optimizing motion vectors based on the perceptual model

Maximize motion impressions while keeping the subjective inconsistency level predicted by the perceptual model within the tolerable range.



Automatically generates results in real time, comparable to conventional methods that require manual fine-tuning

References

[1] T. Fukiage, T. Kawabe, S. Nishida, "Perceptually based adaptive motion retargeting to animate real objects by light projection," *IEEE Transaction on Visualization and Computer Graphics*, Vol. 25, No. 5, pp. 2061-2071, 2019.

[2] T. Kawabe, T. Fukiage, M. Sawayama, S. Nishida, "Deformation Lamps: A projection technique to make static objects perceptually dynamic," *ACM Transactions on Applied Perception,* Vol, 13, No. 2, pp. 1-17, 2016.

Contact

Taiki Fukiage Email: cs-openhouse-ml@hco.ntt.co.jp

Sensory Representation Group, Human Information Science Laboratory

オープンハウス 2020

24 Tiny eye movements reflect cognitive states

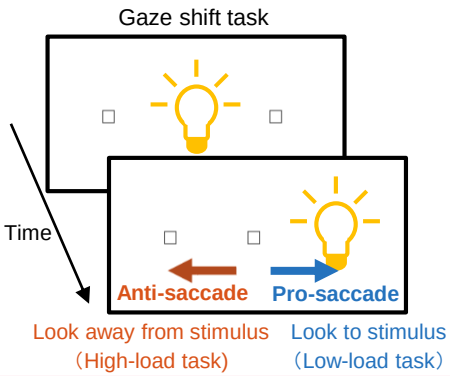
Relation of eye-movement dynamics with cognition and pupil

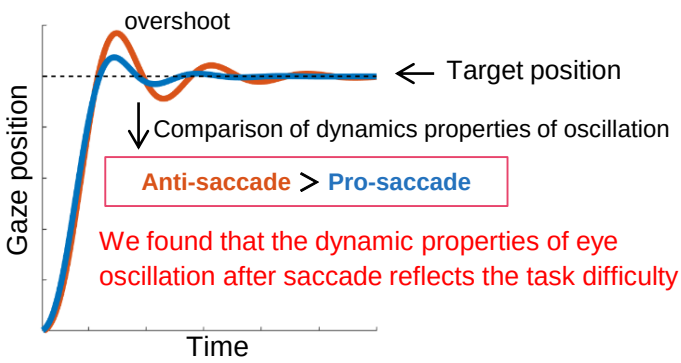
Abstract

Recent studies have shown that the characteristics of the eye movements reflect the cognitive state varying moment by moment. In this study, we investigated the relationship between cognitive states and the detailed dynamics of eye movements which have been regarded as mere mechanical oscillation. We found that the dynamic properties of eye oscillation after saccade reflect the task difficulty in gaze shift task. In addition, we showed that the oscillation dynamics was greater for pupil-centric motion than motion of entire eyeball. The correlation between the pupil-centric oscillation and pupil size indicates that it reflects the instantaneous states of eye tissue inside iris (e.g., stiffness of muscles controlling pupil size). There is a potential that the measurement of the tiny eye movements can be applied to a tool for monitoring the time-varying cognitive state (for example, monitoring the worker who engages in task requiring attention or cognitive load).

Relation between dynamics of saccadic eye movement and cognitive task

- The gaze position after a saccade does not stop exactly at the target position but oscillate around it (overshoot)
- We tested the relationship between overshoot, which has been regarded as mere mechanical vibration, and cognition

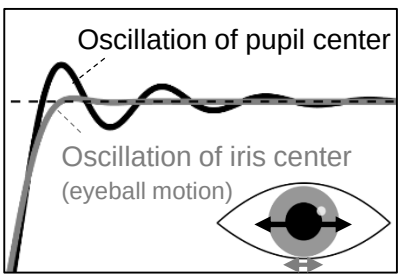




We found that the dynamic properties of eye oscillation after saccade reflects the task difficulty

Eye-movement dynamics reflects the physical properties of eye tissue

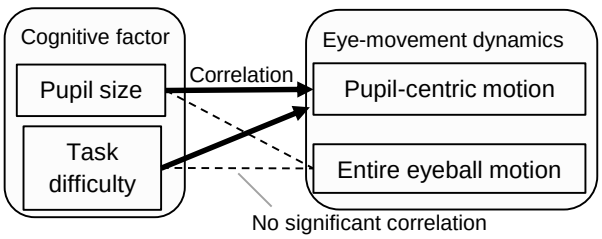
- Extracts the pupil-centric and iris-centric motions



Viscoelasticity of the pupil-centric oscillation is greater than iris-centric one (entire eyeball)

Correlation between pupil size and dynamics of pupil-centric motion

- Calculates the correlation coefficient between eye-movement dynamics and pupil size or task difficulty



Dynamics correlate with pupil size as well as task difficulty

➡ Reflects state of muscles controlling pupil size

References

[1] S. Yamagishi, M. Yoneya, S. Furukawa, "Relationship of postsaccadic oscillation with the state of the pupil inside the iris and with cognitive processing," *J Neurophysiol*, 2020.

Contact

Shimpei Yamagishi Email: cs-openhouse-ml@hco.ntt.co.jp

Sensory Resonance Research Group, Human Information Science Laboratory

オープンハウス 2020

25 Haptic metameric textures

Direct control of perceived texture of 3D printed stimuli

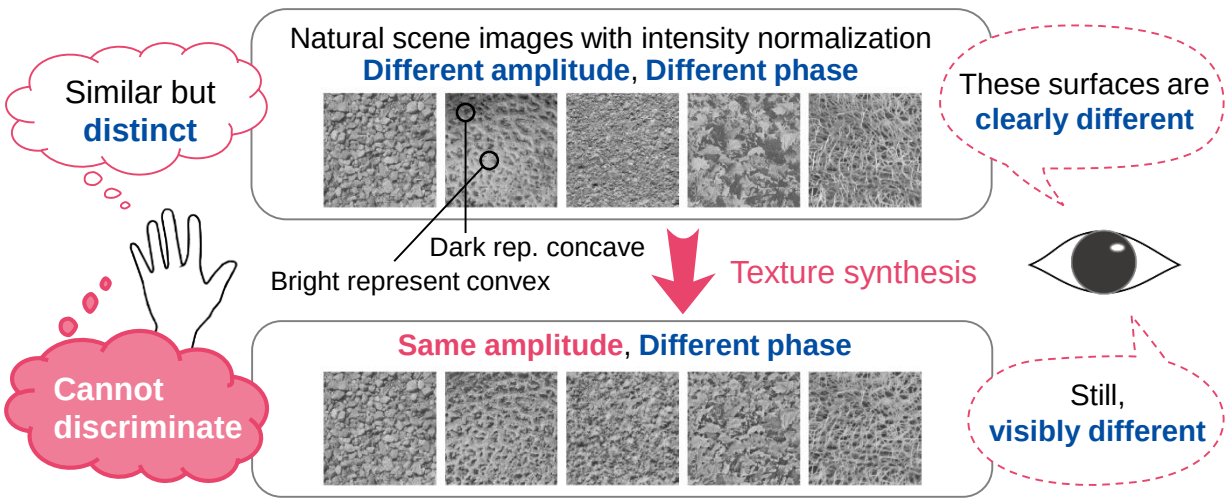
Abstract

Humans sense spatial patterns through their eyes and hands. Past studies have revealed differences (as well as similarities) between vison and touch in texture processing (e.g., eye is good at detecting texture boundaries, while hand can discriminate subtle texture differences), but the underlying computational differences remain poorly understood. Here we transcribed various textures as surface relief patterns by 3D-printing, and analyzed the tactile discrimination performance regarding the sensitivity to image statistics. We found that visually very different patterns cannot be distinguished by touch if they differ only in higher-order statistics. Human tactile texture processing differs from visual one not only in spatio-temporal resolution but also in (in)sensitivity to higher-order image statistics.

Manipulation of surface texture by image processing

Treat a height map as a grayscale image

- Converting the depth pattern of surface carving into a luminance value of a monochrome image
- Conduct image processing on the image then 3D print
- Investigated that tactile texture perception is sensitive to the Fourier amplitude spectrum while not to the Fourier phase spectrum



Touch is more sensitive to small differences in the amplitude spectrum
Vision is more sensitive to higher-order statistics, including the phase spectrum

➤ Realize the design of different looks with identical tactile feeling

References

[1] S. Kuroki, S. Sawayama, S. Nishida, “Haptic metameric textures,” bioRxiv, 2019. doi: https://doi.org/10.1101/653550

Contact

Scinob Kuroki Email: cs-openhouse-ml@hco.ntt.co.jp
Sensory representation research Group, Human and Information Science Laboratory

オープンハウス 2020

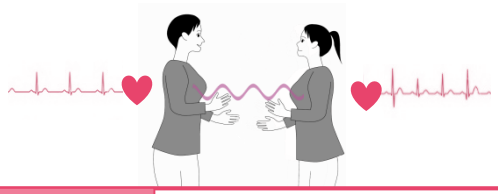
26 What causes emotional change?

Monitoring emotion in experimental settings and daily life

Abstract

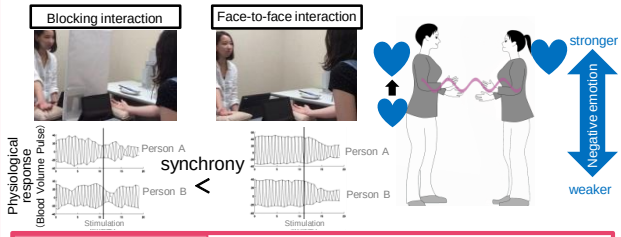
Our emotions are influenced by changes in both our internal states and external environment including interactions with others. In this study, we aimed to investigate how emotional responses change through social interaction in the experimental setting, and to develop a new framework for monitoring the internal change of emotional states in daily life. The findings of two experiments measuring autonomic responses during interaction suggested that negative emotions and positive emotions are transmitted differently through interaction. The interpersonal dynamics of emotional change found here will help us to understand larger group phenomena such as crowd joy or panic. Furthermore, for the purpose of logging internal states which dynamically change through daily life, we developed a new self-tracking method using exclamations or onomatopoeias (e.g., “NIKONIKO”, “SHOBON”). This kind of framework will contribute to creating the system that support our wellbeing.

Emotional change through social interaction in the experimental setting



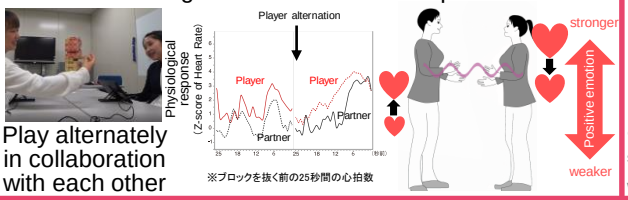
Experiment 1 Negative emotion — Sharing pain by simultaneously stimulation of thermal stimulus

Participants with weaker reactions elevated their physiological reactivity in response to their partner’s reactions during face-to-face interaction, but not vice versa.



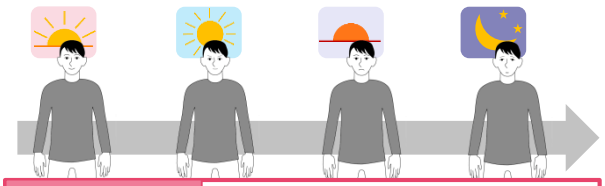
Experiment 2 Positive emotion — Sharing excitement during collaborative block game

Participants changed their physiological reactivity in response to their partner’s reactions, whether their reaction is stronger or weaker than their partner.



Play alternately in collaboration with each other

A new self-tracking system for monitoring emotion in daily life



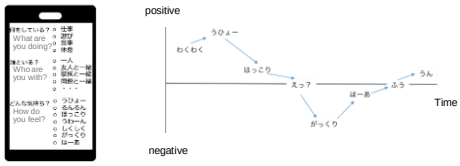
Commonly used method — Numerical rating for each adjective expressing emotion

Happy 0 : not at all ～ 10 : very much
Sad 0 : not at all ～ 10 : very much

- High cognitive load (difficult to answer intuitively)
- Unsuited to capture physical emotional experience
- Hesitate to answer honestly (esp. negative emotions)

The method we propose — Reporting emotional states using exclamations and onomatopoeias

We propose a new self-tracking method using emotional expression words (exclamations, onomatopoeias) which are suitable to express physical emotional experience.



Words were mapped to emotion type and strength using a large-scale survey (N=14,000) to enable such word-based emotion assessments.

ex.)	Happiness	Sadness	Anticipation	Surprise	Anger	Fear	Disgust	Accept
strong	うひょー	がっくり	びりびり	えー	いらいら	びくびく	げっ	うっとり
weak	にこにこ	しゅん	そわそわ	どきっ	かちん	ぞくっ	うわぁ	うん

References

[1] A. Murata, H. Nishida, K. Watanabe, and T. Kameda. “Convergence of physiological responses to pain during face-to-face interaction,” *Scientific Reports,* 10(1), 1-10. 2020.

[2] 村田 藍子・熊野 史朗・渡邊 淳司. “協力場面における対人インタラクションの当事者評価と客観評価,” *電子情報通信学会技術研究報告*, vol. 118, No. 487, pp.111-114, 2019.

Contact

Aiko Murata Email: cs-openhouse-ml@hco.ntt.co.jp
Sensory Resonance Research Group, Human Information Science Laboratory

オープンハウス 2020

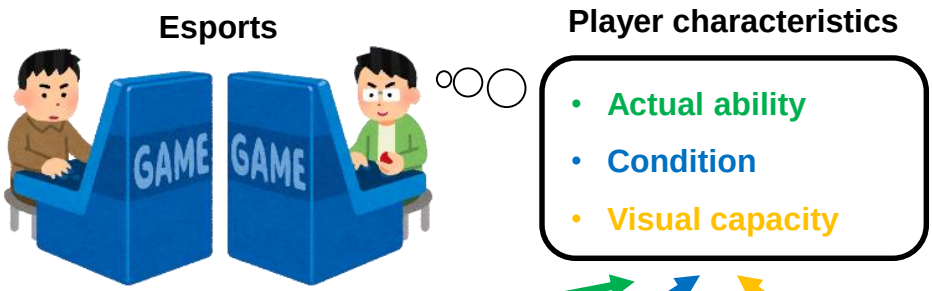
27 Special cognitive abilities of esports experts

Performance, physiological state, and brain activity

Abstract

Since esports is not easily influenced by physical factors, it is said that their actual ability, condition, and visual capacity are difficult to analyze. To objectively evaluate these characteristics of esports players, we have investigated the relationship between game performance and the physiological/brain states of esports players by taking a neuroscience approach. The results of our experiments show that synchrony of heart rate between players reflects their actual abilities and that parietal brain activities are associated with the outcome of the next round. Further investigations, including a vision-science-based test of the capacities of visual information processing of esports players, will reveal the physiological/brain states and cognitive capacities related to performance and will enable us to establish a neuroscience methodology for esports players to improve their performance in competitive environments.

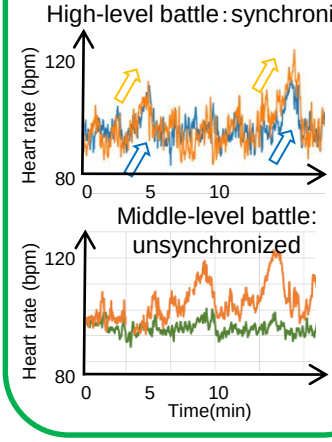
Esports and player characteristics



Approach

Heart rate (HR)

Synchrony of HR reflects actual abilities of players [1]



Brain activity

Brain activities are associated with the outcome of next round

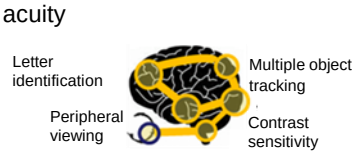
Parietal brain activity **increases**

Parietal brain activity **decreases**

Visual test

Examination on superior visual capacities of players

Focusing on the various types of visual capacity as well as visual acuity



Measuring capacities using our unique vision test battery [2]

References

[1] K. Watanabe, N. Saijo, M. Kashino（2019）"The across-player correlation of the physiological change reflecting the fight-or-flight response in esports." *Neuroscience Meeting Planner. Chicago, IL: Society for Neuroscience, 2019*. Online. Program No. 769.20.

[2] K. Hosokawa, K. Maruya, S. Nishida, M. Takahashi and S. Nakadomari（2019）"Gamified vision test system for daily self-check," *2019 IEEE Games, Entertainment, Media Conference（GEM）*, New Haven, CT, USA, 2019, pp. 1-8.

Contact

Sorato Minami Email: cs-openhouse-ml@hco.ntt.co.jp
Kashino Diverse Brain Research Laboratory

オープンハウス 2020

28 Realizing a harmony in rugby scrum

Easy assessment of player coordination with wearable sensors

Abstract

It is crucial for team plays in sports that the players synchronize their actions, but objectively assessing player coordination is not easy. We propose a convenient measurement method to immediately evaluate and feed back some aspects of player coordination by attaching compact inertial measurement units (IMUs) to each player; we use the example of scrumming in rugby. In a scrum, a pack of eight forwards (players) are required to coordinate their forward drives, timing and direction, to maximize forward pressure. The IMU data allows us to calculate the acceleration vectors and its peak time structures for the group of players involved. Constant storage of these values can yield a useful database for understanding each player's characteristics and developing suitable combinations of players to improve scrumming performance. This measurement system also makes it easy to visualize, as well as sonify, player coordination during various joint activities other than rugby.

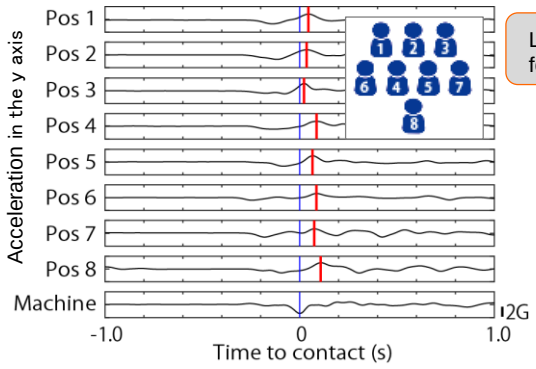
Convenient wearable sensor

Analysis, Database

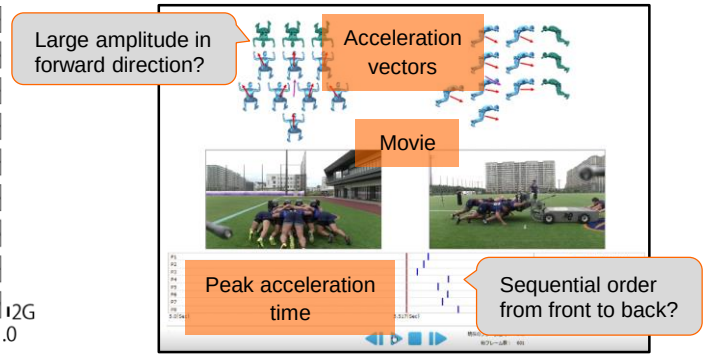
Daily training

Feedback of player coordination

Example of acceleration profiles

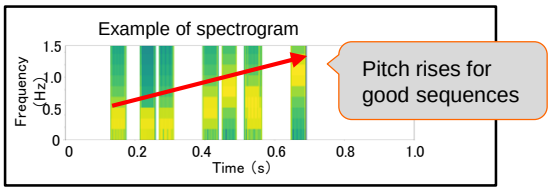


Example of feedback screen



- Easy acquisition via wearable sensors
- Rapid visualization of player coordination
- Developing suitable combinations of players
- Potential of application to various activities

Sonifying the peak time sequence of players



＊This research is in collaboration with NTT Communications Shinning Arks.

References

[1] T. Kimura, N. Saito, H. Okamoto, K. Ohta, "Evaluation of cooperation between players in the rugby scrum using IMU," *Proc. Sports Engineering and Human Dynamics 2019*, 2019.

.

Contact

Toshitaka Kimura Email: cs-openhouse-ml@hco.ntt.co.jp
Kashino Diverse Brain Research Laboratory

オープンハウス 2020

29 What is a "straight" ball?

Physical and perceptual attributes of a pitched ball

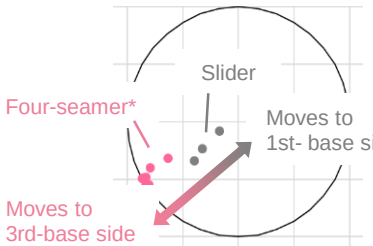
Abstract

Athletes' perceptions are extraordinarily sensitive, but they do not necessarily capture the physical world as it is. We accurately quantified the physical characteristics of pitched baseballs, and then investigated how batters perceived them. In the physical measurement, we devised a technique to easily measure the 3D rotation characteristics of the ball using a single camera. In the perceptual measurement, we found that the batters' perception of the ball's horizontal movement in the trajectory was systematically biased, even though they discriminated small differences in the movement. Moreover, we found that the direction of their bias was reversed, depending on the pitcher's handedness. In sports science, physical measurement tends to be emphasized, and the gap between the physical characteristics and players' perception is at issue. Combining physical and perceptual measurement to identify the cause of such a gap could lead to a dramatic revolution in training and coaching.

Physics

3D measurement of ball rotation

Directions of ball rotation axis for an exemplary right-handed pitcher. 2D ball diagram from the batter's view. The rotation axes (dots) and ball's horizontal movements were analyzed in detail.



*Four-seamer is "straight" in Japanese.

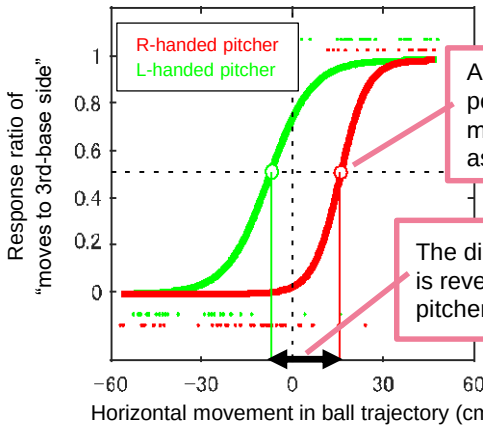
Physical trajectory of four-seamer

Most four-seamers for right-handed pitcher physically move to 3rd-base side

This measurement technique is being tested for practical use with professional and amateur baseball teams, and the Japan Women's National Softball Team.

Perception

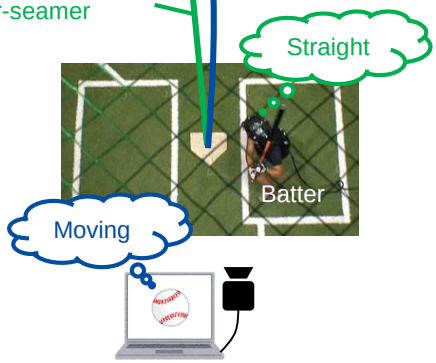
An example of batter's perception



A batter illusorily perceive a physically moving ball's trajectory as "straight".

The direction of the illusion is reversed, depending on pitcher handedness.

Perceptual trajectory of four-seamer



Perception is biased in the direction the four-seamer is perceived as "straight".

References

[1] D. Nasu, T. Kimura, M. Kashino, "Do baseball batters perceive straight ball trajectory as straight?" *Proc. 2020 Conference on North American Society for Psychology of Sport and Physical Activity*, 2020.

Contact

Daiki Nasu Email: cs-openhouse-ml@hco.ntt.co.jp
Kashino Diverse Brain Research Laboratory

オープンハウス 2020

30 Body representation for quick and skillful action

Uncertainty of hand-state estimate regulates stretch reflex

Abstract

Reflexive motor control mechanisms are embedded in the brain to unconsciously correct ongoing body movements. They do this by detecting changes in the external world and ones own posture, via sensory signals from eyes and limbs. We investigated the information processing underlying the functional and context-dependent regulation of reflexive responses. We found an interesting attenuation of muscle response to resisting sudden changes in limb movement, which occurred if visual feedback of the limb movement was not given or was distorted. The result suggests that the brain regulates reflexive responses depending on body states estimated by combining multimodal information such as vision and bodily sensations, rather than single modality information as previously thought. We will further explore the computational mechanisms of reflexive sensorimotor control, which may be beneficial to analyzing the performance of athletes or to developing effective sports training methods.

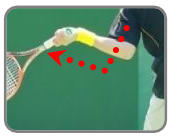
Movement correction without consciousness

* Reflexive muscle contraction occurring in response to sudden extension of a muscle to recover the posture

Consciousness

Motor correction through consciousness (≥ 200 ms)

Stretch reflex: Somatosensory-motor response (~50 ms)



Motor command

Somatosensory input

- Postural change
- Interaction with environment

Reflex responses enable quick compensation against disturbances to movement and posture

Informational processing for reflex regulation

Conventional hypothesis: unimodal processing

Somatosensory information → Stretch reflex modulation

Proposed hypothesis: multimodal processing

Visual information

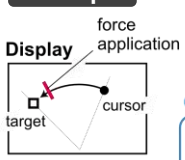
Somatosensory information

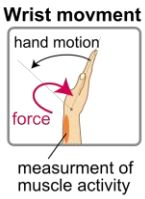
Multisensory integration → Stretch reflex modulation

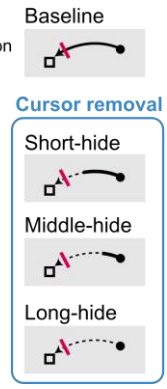
Visual information contributes to regulation of stretch reflex

- Stretch reflex was evoked by applying force during a wrist movement
- Testing if removing visual cue of hand position (cursor) affects the reflexive response
- The stretch reflex was smaller with longer cursor removal
- Suggesting uncertainty of hand states estimated by multisensory integration underlies the reflexive motor control

Setup

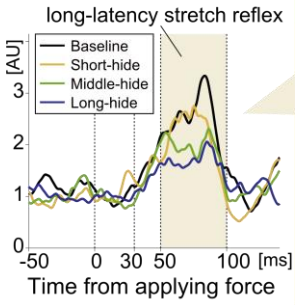




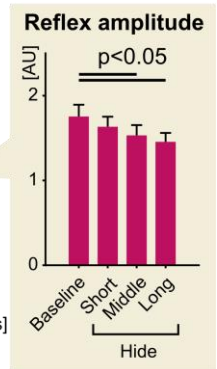


Results

Muscle activity around force application



Reflex amplitude



References

[1] S. Ito, H. Gomi, "Visually-updated hand state estimates modulate the proprioceptive reflex independently of motor task requirements," *eLife*, 9:e52380, 2020.

[2] S. Ito, H. Gomi, "Online modulation of proprioceptive reflex gain depending on uncertainty in multisensory state estimation," *Proc. The Society for Neuroscience 49th Annual Meeting*, 2019.

Contact

Sho Ito Email: cs-openhouse-ml@hco.ntt.co.jp

Sensory and Motor Research Group, Human Information Science Laboratory

オープンハウス 2020

31 Unconscious is smarter than conscious

Environmental dependency in visuomotor responses

Abstract

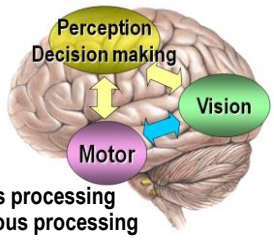
We may tend to think that human understand current states of self-body and external environments, and then consciously control our limbs according to those states. However, dominant part of actual skilled movements are controlled unconsciously. We are trying to reveal those implicit sensorimotor control mechanisms to understand the brain processing for skillful motor control. By inflicting different postural stability and/or noisy visual motion conditions, we investigated the adaptability of voluntary and reflexive responses to visual motion stimuli, and found that only reflex responses can be adjusted suitably to the different situations. This suggests that unconscious processing would be smarter than conscious processing for a particular condition. By understand the mechanisms of brain processing for sensorimotor control, we will be able to designe more sophisticated communication and man-machine interface, and novel training method for athletes.

Body and arm movement control

- In daily life, human has various physical interactions with external environments while his/her body is moving, which cannot be easily realized by current industrial robots.
- How can we realize these dexterous interactions by using the brain information processing whose transmission speed is much slower than that in current computers?

Implicit and explicit sensorimotor processes

It is generally considered that implicit process for sensorimotor control is faster but is less flexible than the explicit and voluntary processing.



The current study revealed that an implicit process can regulate sensorimotor response according to a particular environmental situation whereas an explicit process cannot [1].

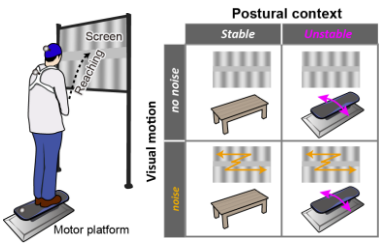
Experiment: Context dependency of implicit and explicit processes for visual motion

Experiment

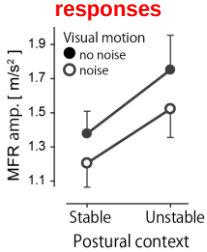
- Supply environmental context of postural fluctuation and/or random visual motion.
- Measure reflexive and voluntary visuomotor response.
  1. Reflexive response (Manual Following Response: MFR)
  2. Voluntary response (Motion Direction Discrimination)

Results

1. Reflexive: Context dependent and rational modulation
2. Voluntary: Context independent modulation



Context-dependent responses

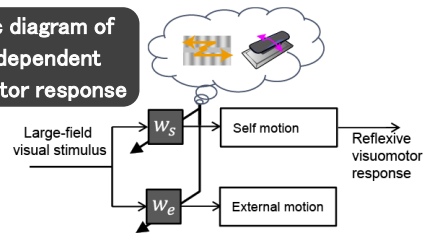


These results suggest that reflex mechanisms properly connect the relationship between postural stability context and visual motion while voluntary motor system cannot.

Unstable posture context ⇒ large MFR

Movable environment context ⇒ small MFR

Schematic diagram of context dependent implicit motor response



References

[1] N. Abekawa, H. Gomi, "Modulation difference in visuomotor responses in implicit and explicit motor tasks depending on postural stability," *Proc. The Society for Neuroscience 47th Annual Meeting*, 2017.

[2] H. Gomi, K. Kadota, N. Abekawa, "Dynamic reaching adjustment during continuous body perturbation is markedly improved by visual motion," *Proc. The Society for Neuroscience 40th Annual Meeting*, 2010.

Contact

Hiroaki Gomi Email: cs-openhouse-ml@hco.ntt.co.jp

Sensory and Motor Research Group, Human and Information Science Laboratory

オープンハウス 2020