Explain AI@Imperial Workshop

FUJITSU

shaping tomorrow with you

Explainable AI that Can be Used for Judgment with Responsibility

FUJITSU LABORATORIES LTD.

25th April 2018

About me

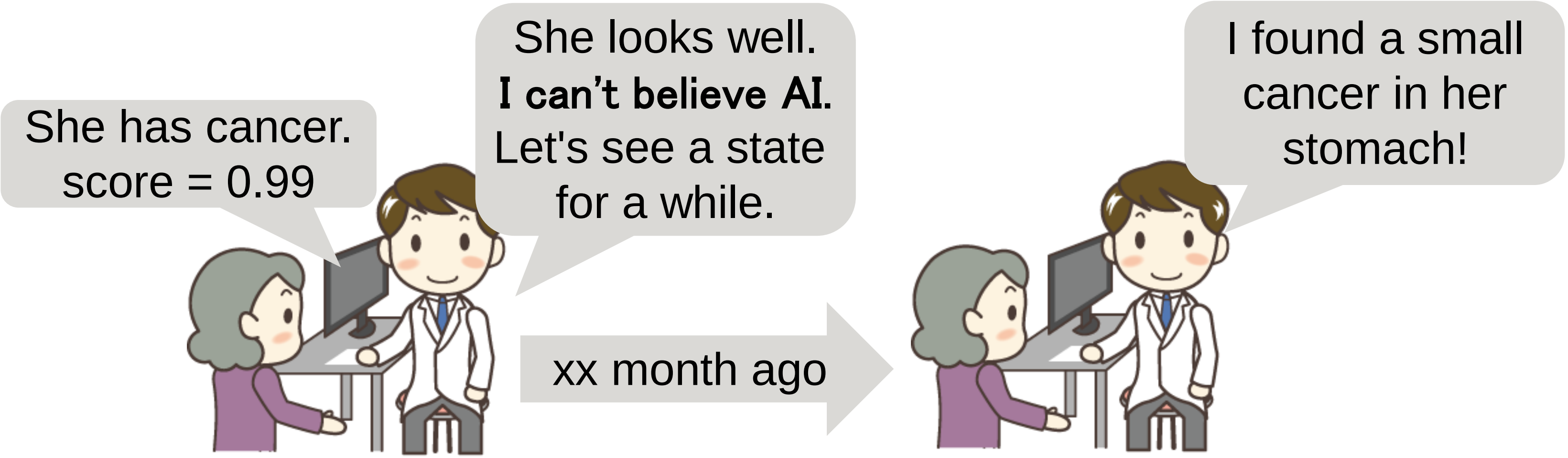
- Name: Hajime Morita
- Research interests:
  - Natural Language Processing
    - Summarization
    - Morphological Analysis
    - Information Extraction
- Recent jobs:
  - I received my Ph.D. from Tokyo Institute of Technology (2015)
  - Researcher @ Kyoto University (2015 - 2017)
  - Researcher @ Fujitu Laboratory (2017 -)

Outline

- Introduction of explainable AI
- Knowledge Graph (short)
- Deep Tensor
- Explainable AI

Why we need explainable AI

- If we applied **IN**-explainable AI to clinical decision.



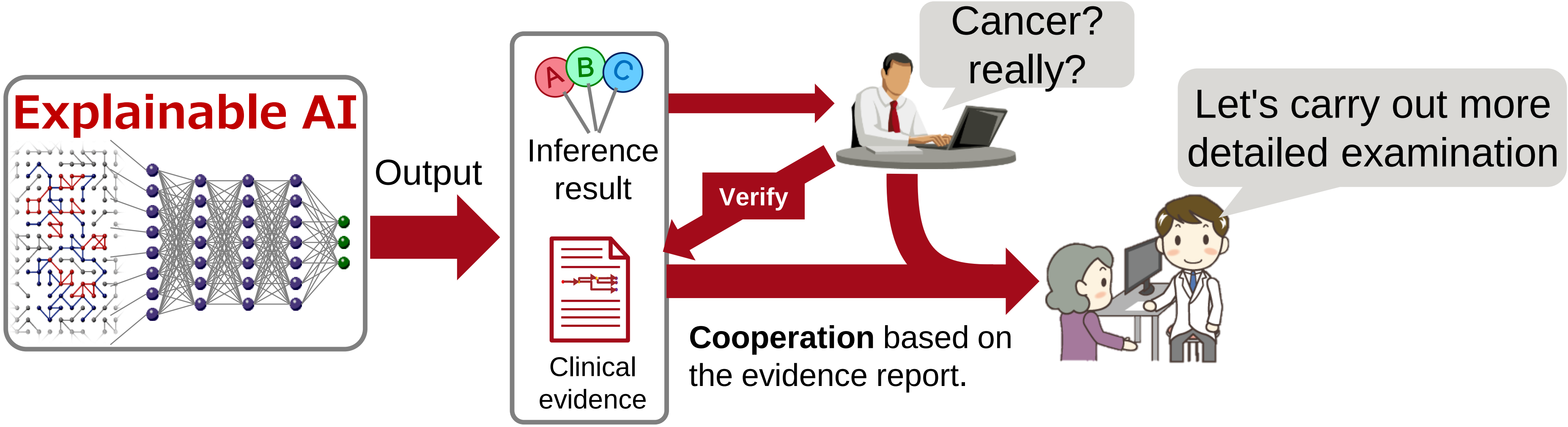
She has cancer.
score = 0.99

She looks well.
I can’t believe AI.
Let's see a state
for a while.

xx month ago

I found a small
cancer in her
stomach!

- If we applied **explainable AI** to clinical decision.



Explainable AI

Output

Inference
result

Clinical
evidence

Cancer?
really?

Verify

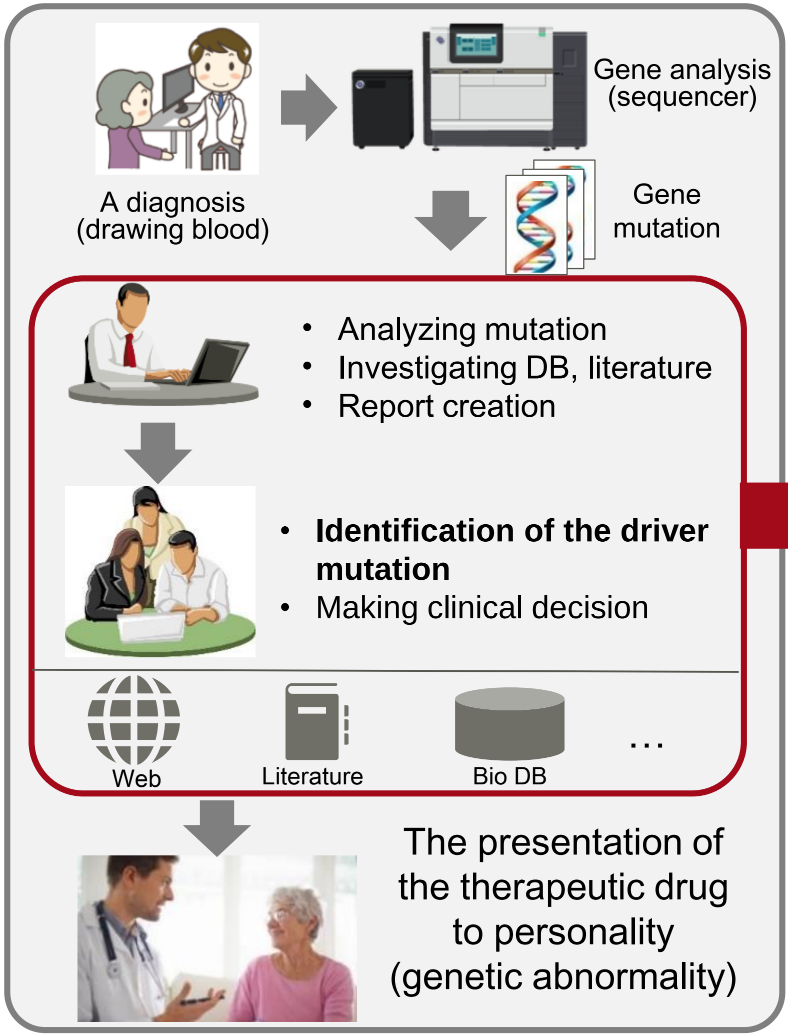
Let's carry out more
detailed examination

Cooperation based on
the evidence report.

Application to Genomic Medicine

FUJITSU

- We developed a prototype of AI that infers which gene mutation of the patient is associated with disease.



A diagnosis (drawing blood)

Gene analysis (sequencer)

Gene mutation

- Analyzing mutation
- Investigating DB, literature
- Report creation

- **Identification of the driver mutation**
- Making clinical decision

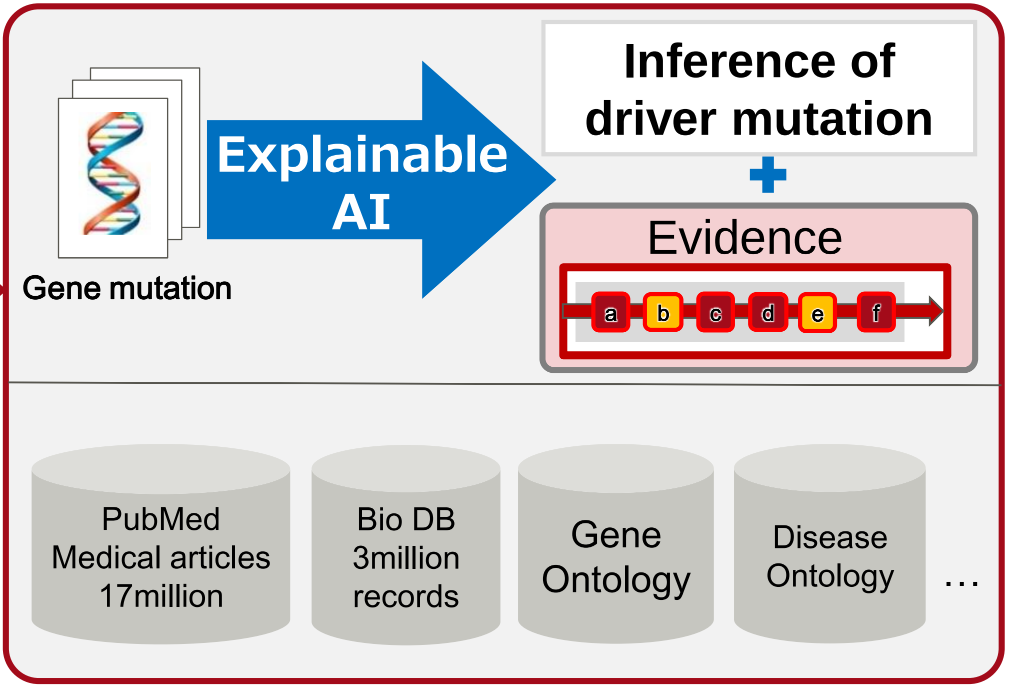
Web

Literature

Bio DB

…

The presentation of the therapeutic drug to personality (genetic abnormality)



Gene mutation

Explainable AI

Inference of driver mutation

+

Evidence

a b c d e f

PubMed Medical articles 17million

Bio DB 3million records

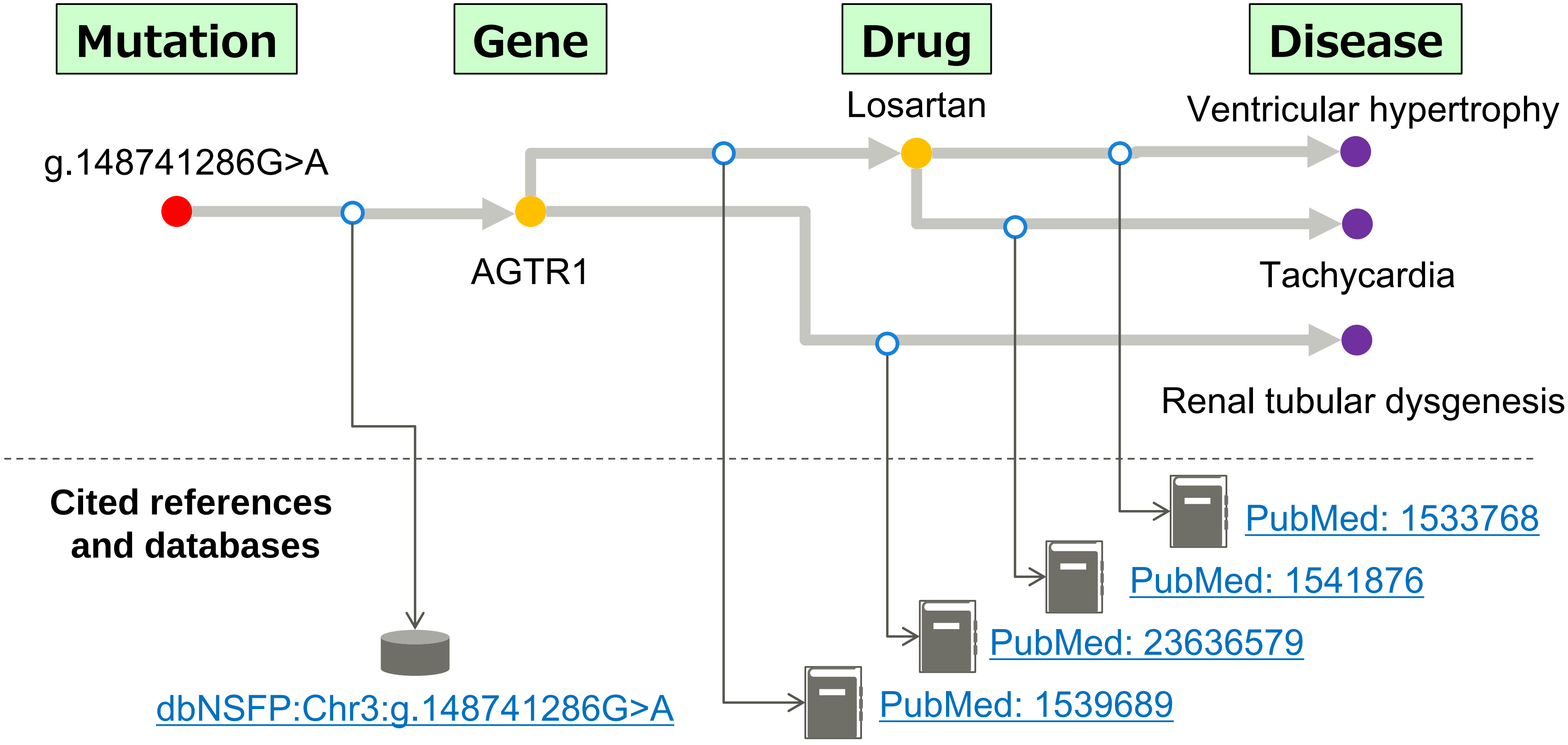
Gene Ontology

Disease Ontology

…

Ref. Hokkaido Univ. Hospital(http://www.huhp.hokudai.ac.jp/hotnews/detail/00001144.html)

Formation of Evidence Path by Knowledge Graph



The mutation cause an abnormality in the gene (AGTR1) and the abnormality can causes diseases such as tachycardia.

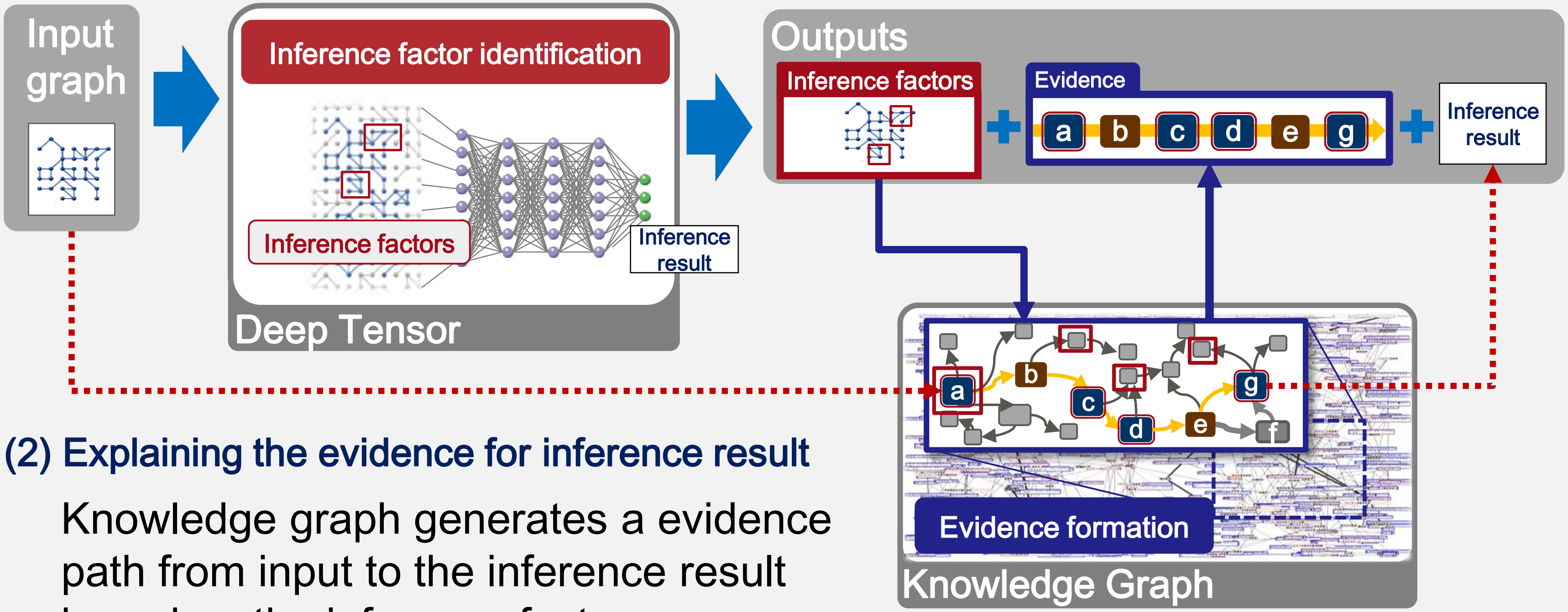
Overview

Two key components

- **Deep Tensor** is a neural classifier of **graph data**
- **Knowledge graph** is an extremely large graph knowledge base

(1) Explaining the important factor for the inference

Outputting the factors which strongly influenced the inference result through Deep Tensor



(2) Explaining the evidence for inference result

Knowledge graph generates a evidence path from input to the inference result based on the inference factors.

Knowledge Graph

What is knowledge graph?

Knowledge

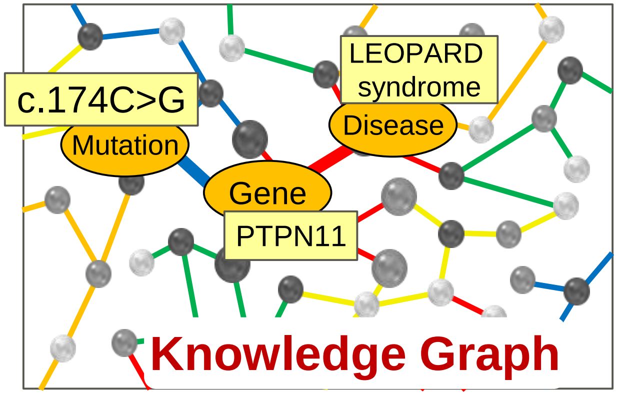
- Nodes and a edge that connects the nodes.

subject relation object

Francis Harry Compton Crick $\xrightarrow{\text{found}}$ DNA double helix

Knowledge graph

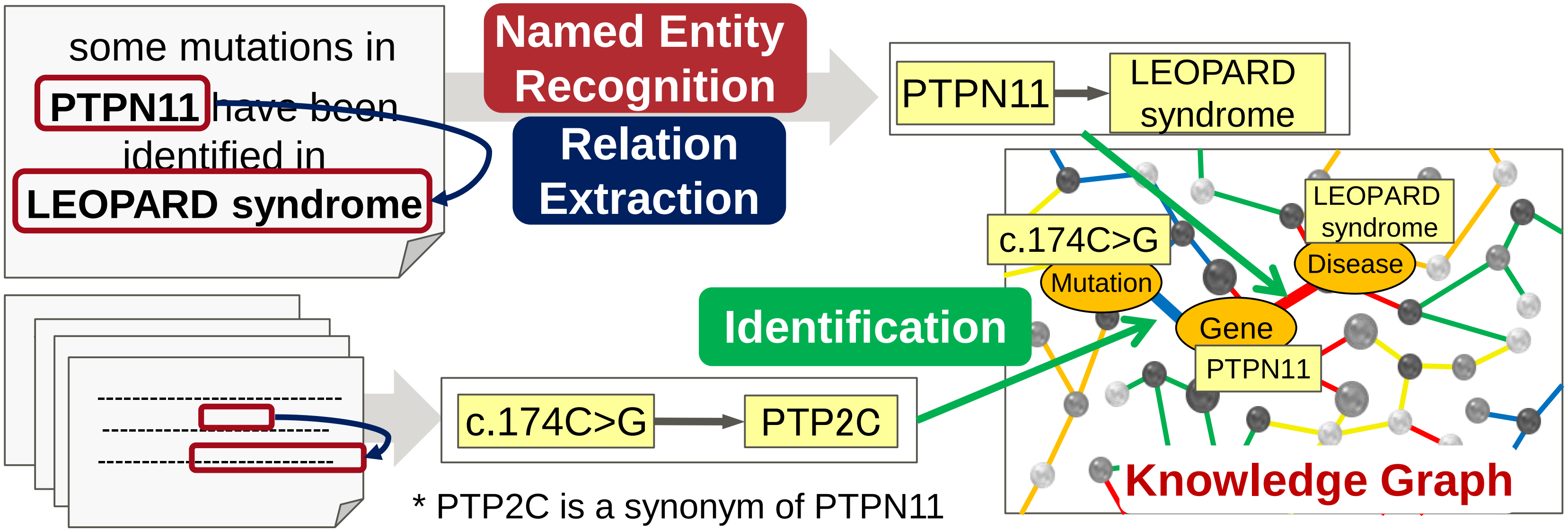
- Nodes
  - Gene, mutation, drug, disease, etc...
- Edges
  - Drug A is responsive to disease B
  - Gene C has an function D
  - Mutation E is located on Gene F
  - ...



How to make knowledge graph

- From existing database
  - Extraction of relational data from public databases
    - Relations between gene and its attributes (name, ID, function, etc.)
    - Relations between protains

- From literature
  - Knowledge extraction using Natural Language Processing technologies



some mutations in PTPN11 have been identified in LEOPARD syndrome

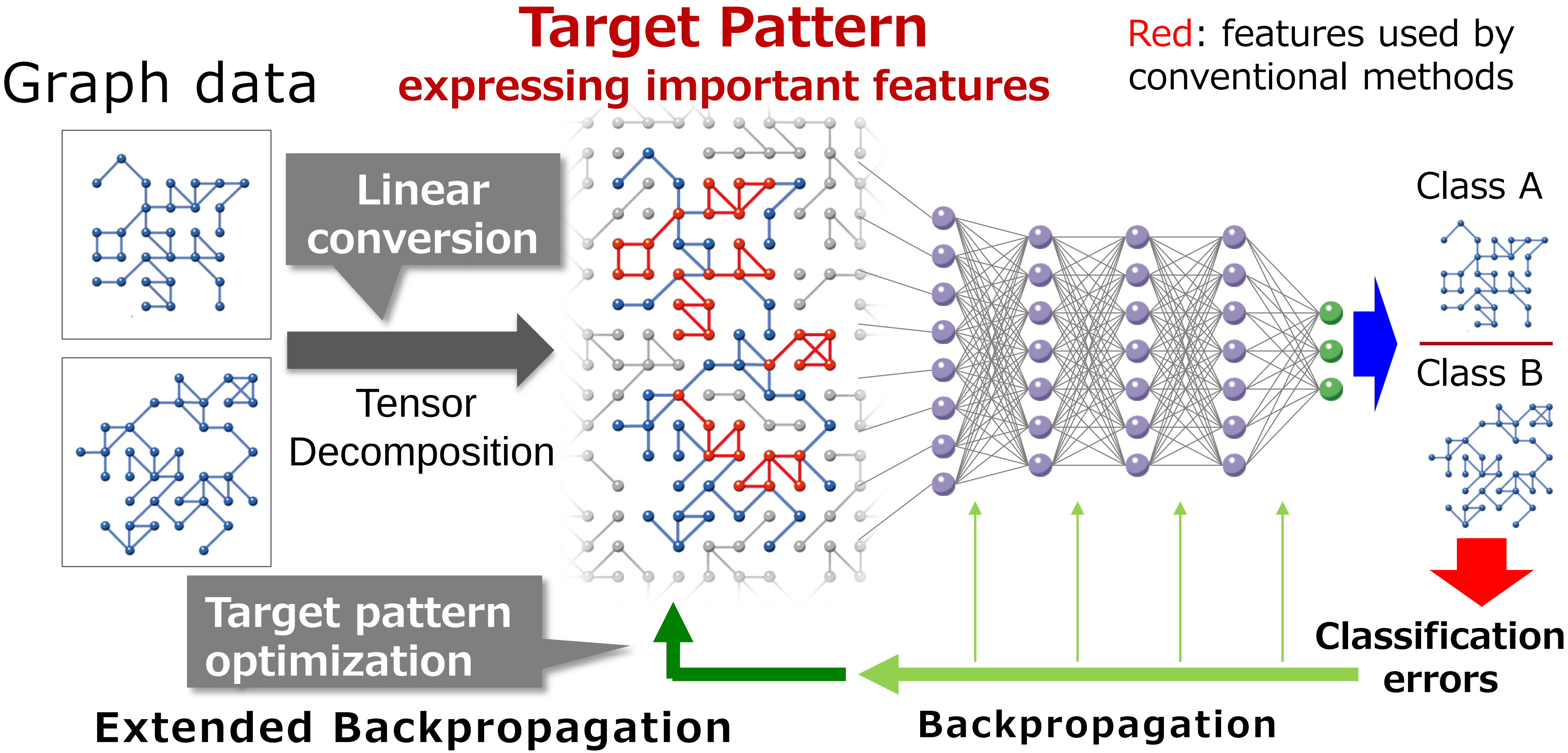
* PTP2C is a synonym of PTPN11

Deep Tensor

- Koji Maruhashi, Masaru Todoriki, Takuya Ohwa, Keisuke Goto, Yu Hasegawa, Hiroya Inakoshi, Hirokazu Anai, Learning Multi-way Relations via Tensor Decomposition with Neural Networks, AAAI 2018 (AAAI 2018), February 2018.

Naïve Idea

- Linearly convert graph data as close as a target pattern
- Optimize the target pattern along with Neural Networks



Classify graph data accurately with high interpretability

Difficulties Using Graph for Deep Leaning

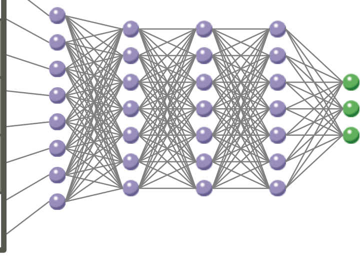
Image

The pixel coordinate number is fixed

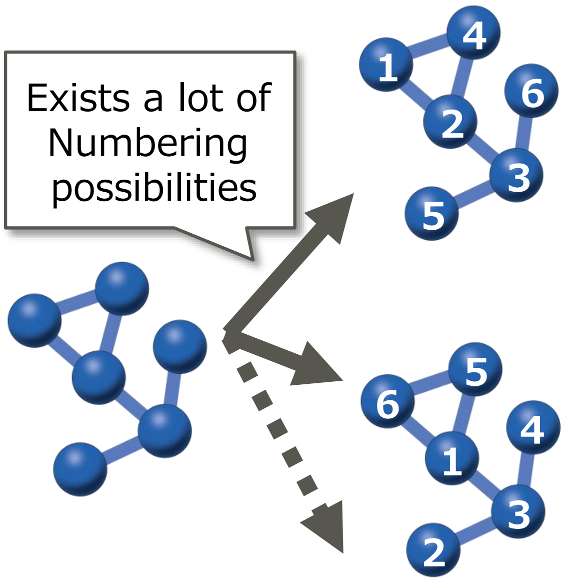
	1	2	3	4	5	6
1	0	1	1	1	1	0
2	1	0	0	0	0	1
3	1	1	1	1	1	1
4	1	0	1	1	0	1
5	1	1	1	1	1	1
6	0	1	0	0	1	0

Input

(1,1)	0
(1,2)	1
(1,3)	1
(1,4)	1
～	



Graph data

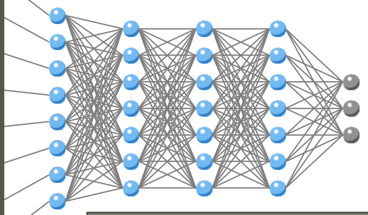


Exists a lot of Numbering possibilities

	1	2	3	4	5	6
1	0	1	0	1	0	0
2	1	0	1	1	0	0
3	0	1	0	0	1	1
4	1	1	0	0	0	0
5	0	0	1	0	0	0
6	0	0	1	0	0	0

Input

1-2	1
1-3	0
1-4	1
1-5	0
～	

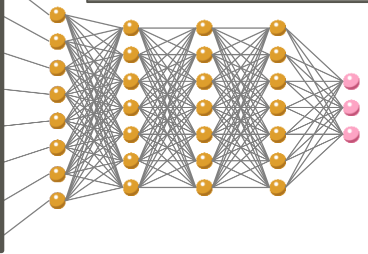


Coordinate label caused results unstabiliy

	1	2	3	4	5	6
1	0	0	1	0	1	1
2	0	0	1	0	0	0
3	1	1	0	1	0	0
4	0	0	1	0	0	0
5	1	0	0	0	0	1
6	1	0	0	0	1	0

Input

1-2	0
1-3	1
1-4	0
1-5	1
～	



⋮

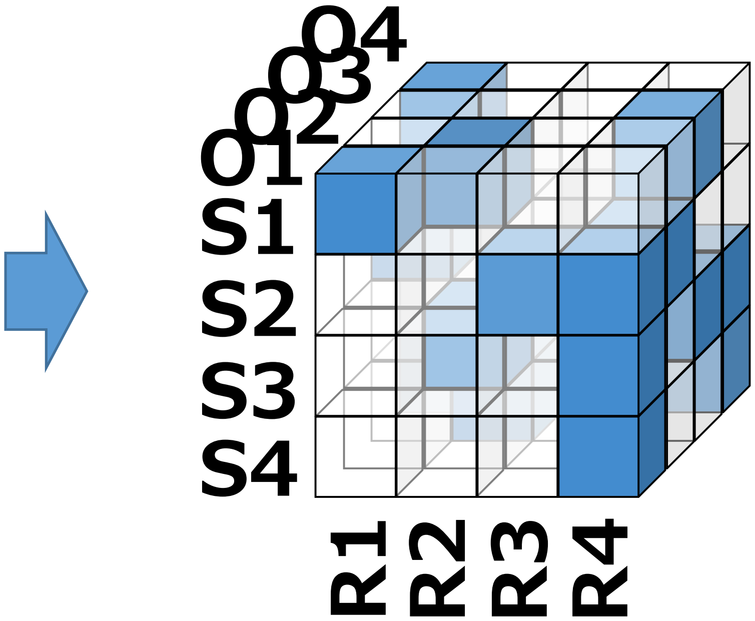
Tensor Representation

- Multi-way data can be represented as a tensor

Multi-way data

Sub	Relation	Obj
S1	R1	O1
S2	R3	O1
S3	R2	O2
…	…	…

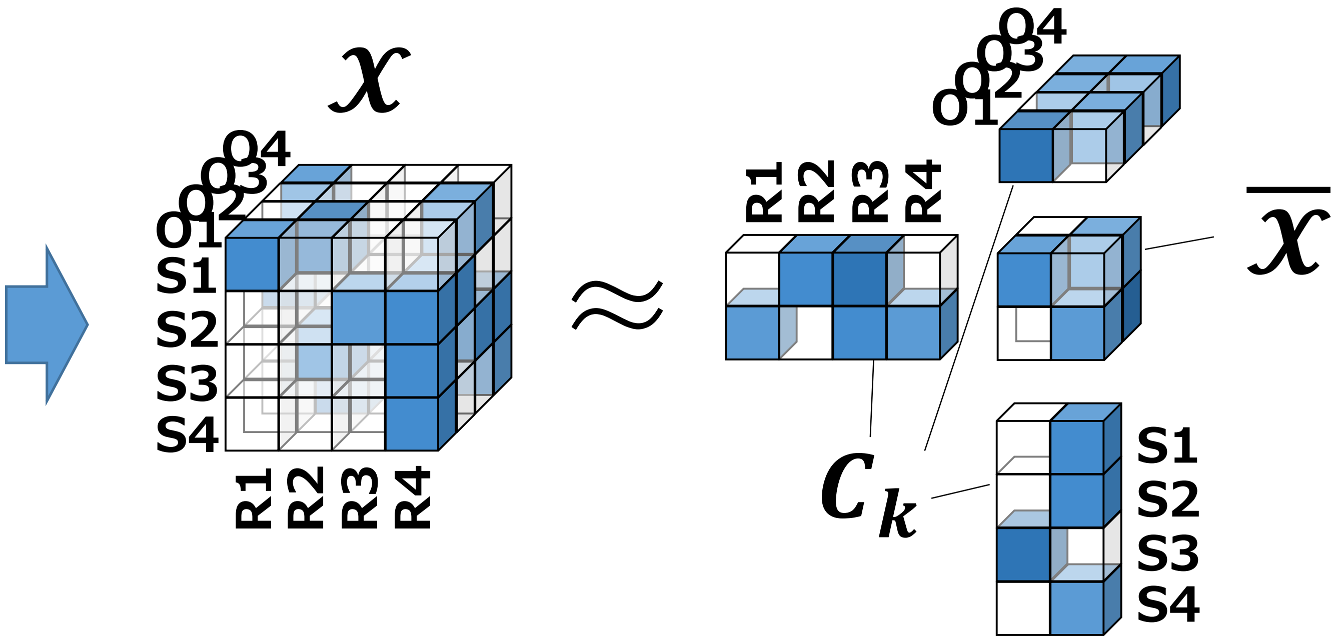
Tensor representation



Tensor Decomposition

- Approximate a tensor $\mathcal{X}$ by a core tensor $\overline{\mathcal{X}}$ multiplied by factor matrices $\{C_k\}$
- Results are <u>easy to interpret</u> in terms of nodes and edges.

sbj	rel	obj
S1	R1	O1
S2	R3	O1
S3	R2	O2
…	…	…

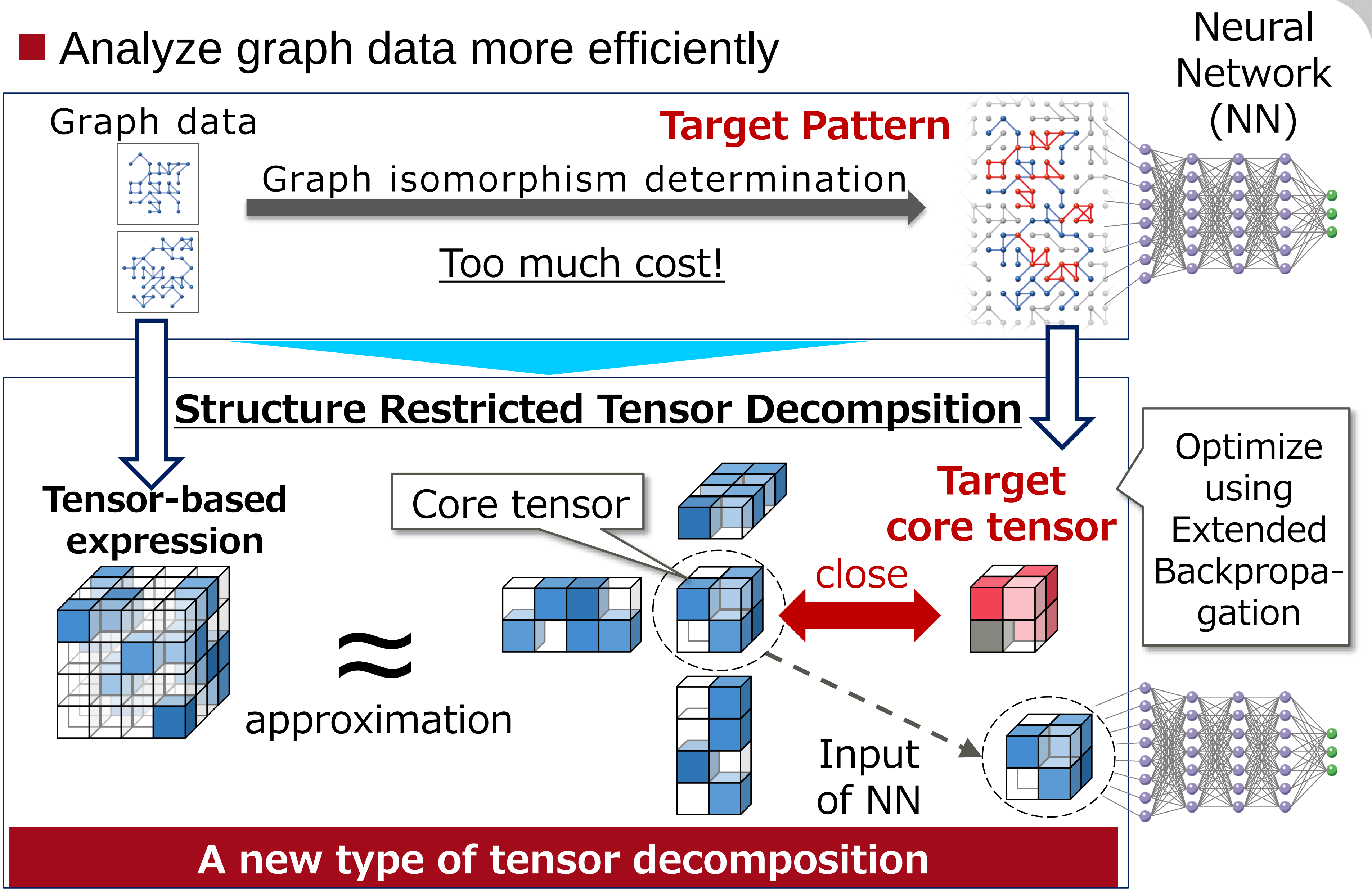


$$(\overline{\mathcal{X}}, \{C_k\}) = \operatorname{argmin}\left\|\mathcal{X} - \overline{\mathcal{X}} \textstyle\prod_k \times_k \ C_k^T\right\|_2^2$$

Can we leverage tensor decomposition?

Leveraging Tensor Decomposition

- Analyze graph data more efficiently



Graph data

Target Pattern

Graph isomorphism determination

Too much cost!

Neural Network (NN)

Structure Restricted Tensor Decompsition

Tensor-based expression

≈ approximation

Core tensor

Target core tensor

close

Optimize using Extended Backpropagation

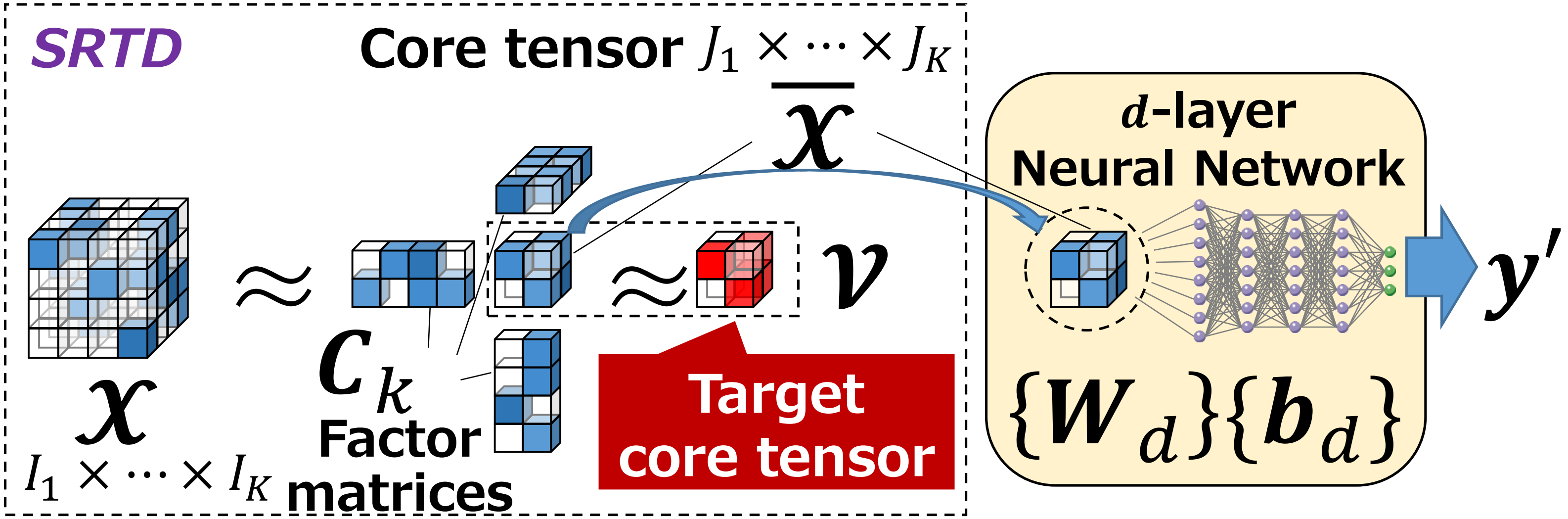
Input of NN

A new type of tensor decomposition

Structure Restricted Tensor Decomposition (SRTD)

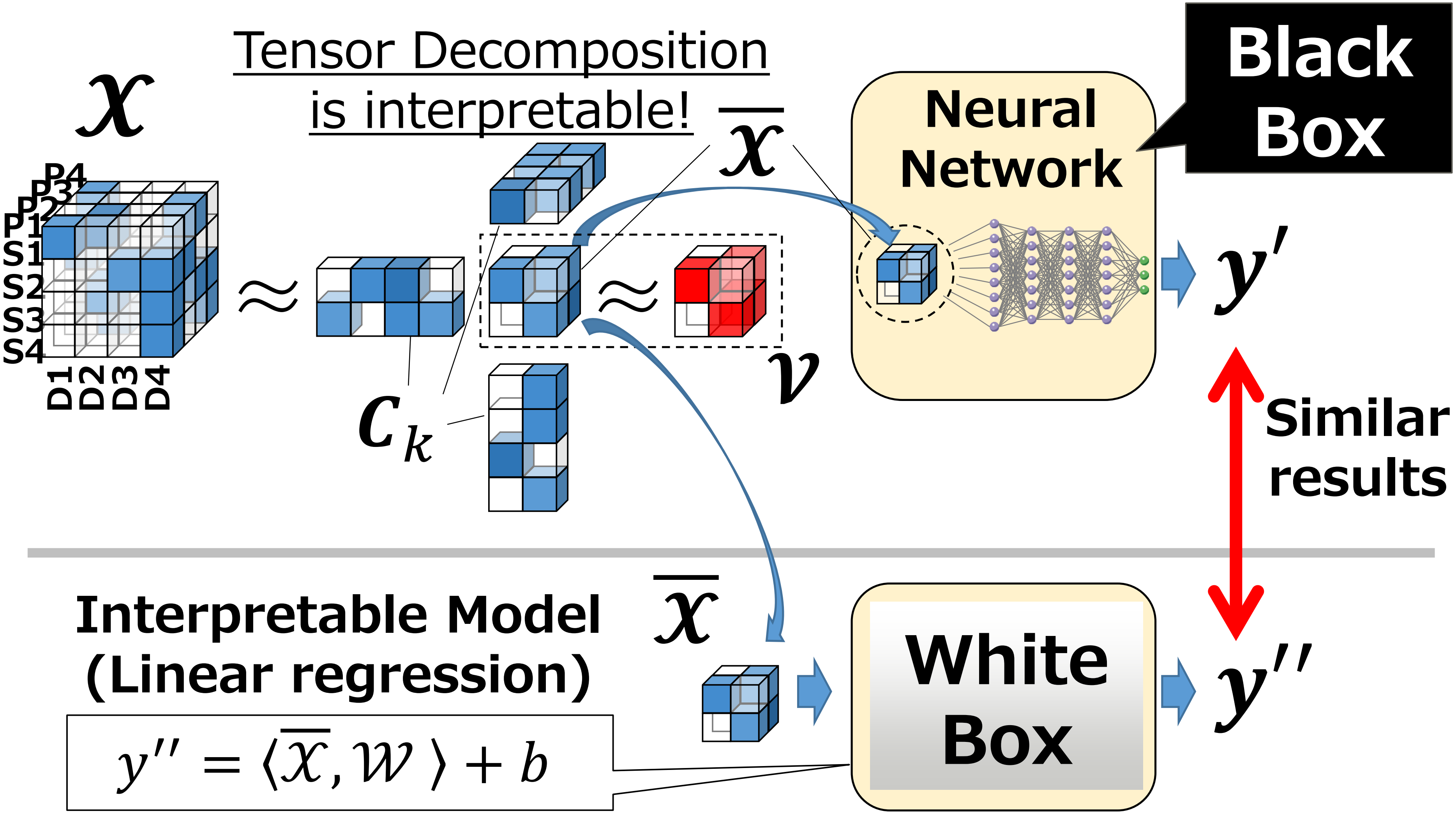
Details

- Calculate $\overline{\mathcal{X}}$ which minimize $\left\|\boldsymbol{\mathcal{X}} \prod_{k|I_k<J_k} \times_k \boldsymbol{C}_k - \overline{\boldsymbol{\mathcal{X}}} \prod_{k|I_k \geq J_k} \times_k \boldsymbol{C}_k^T \right\|_2^2$
- by using $\boldsymbol{C}_k$ which minimize $\left\|\boldsymbol{\mathcal{X}} \prod_{k|I_k<J_k} \times_k \boldsymbol{C}_k - \boldsymbol{\mathcal{V}} \prod_{k|I_k \geq J_k} \times_k \boldsymbol{C}_k^T \right\|_2^2$
- subject to $\boldsymbol{C}_k^T \boldsymbol{C}_k = \boldsymbol{I}(I_k \geq J_k), \boldsymbol{C}_k \boldsymbol{C}_k^T = \boldsymbol{I}(I_k < J_k)$


SRTD
Core tensor $J_1 \times \cdots \times J_K$
$\overline{\boldsymbol{\mathcal{X}}}$
$\boldsymbol{\mathcal{X}}$
$I_1 \times \cdots \times I_K$
$\approx$
$\boldsymbol{C}_k$
Factor matrices
$\approx$
$\boldsymbol{\mathcal{V}}$
Target core tensor
d-layer
Neural Network
$\{\boldsymbol{W}_d\}\{\boldsymbol{b}_d\}$
$\boldsymbol{y}'$

How to interpret prediction results?

- Solution: Learn *interpretable models* that output similar results as Neural Networks



Tensor Decomposition is interpretable!

Black Box

Similar results

Interpretable Model (Linear regression)

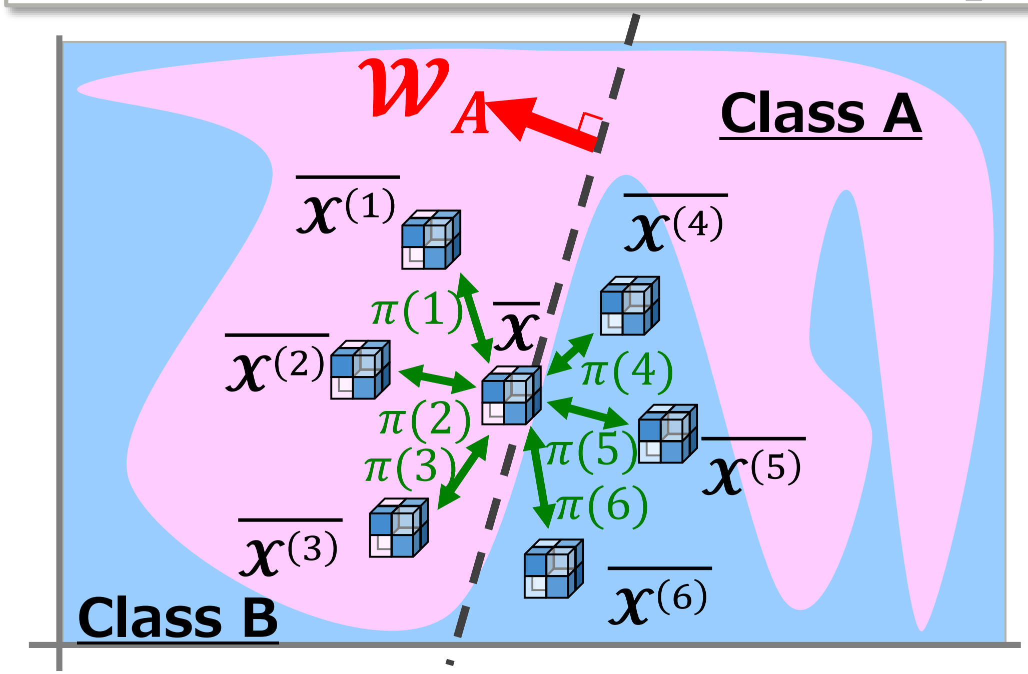
$$y'' = \langle \bar{\mathcal{X}}, \mathcal{W} \rangle + b$$

Local Interpretable Models

Details

- Use perturbed samples $\boldsymbol{x}^{(p)}$ based on LIME [Ribeiro+ KDD16]

  - Learn interpretable model $g_A\left(\overline{\boldsymbol{x}^{(p)}}\right) = \langle \overline{\boldsymbol{x}^{(p)}}, \boldsymbol{w}_A \rangle + \boldsymbol{b}_A$
  - which minimize $\sum_p \pi(p) \left\| y_A'^{(p)} - g_A\left(\overline{\boldsymbol{x}^{(p)}}\right) \right\|_2^2$
  - where $\pi(p) = \exp(-\left\| \overline{x} - \overline{x^{(p)}} \right\|_2^2 / \sigma^2)$



$\boldsymbol{w}_A$
Class A
$\overline{\boldsymbol{x}^{(1)}}$
$\overline{\boldsymbol{x}^{(4)}}$
$\pi(1)$
$\overline{\boldsymbol{x}}$
$\overline{\boldsymbol{x}^{(2)}}$
$\pi(4)$
$\pi(2)$
$\pi(3)$
$\pi(5)$
$\overline{\boldsymbol{x}^{(5)}}$
$\pi(6)$
$\overline{\boldsymbol{x}^{(3)}}$
$\overline{\boldsymbol{x}^{(6)}}$
Class B

Probability that $\boldsymbol{x}^{(p)}$ is classified into A by DeepTensor

$$y_A'^{(1)} = 0.9 \quad y_A'^{(4)} = 0.1$$
$$y_A'^{(2)} = 1.0 \quad y_A'^{(5)} = 0.0$$
$$y_A'^{(3)} = 0.9 \quad y_A'^{(6)} = 0.1$$

- *Contribution score* is calculated by $\boldsymbol{x} * \boldsymbol{w}_A \prod_k \times_k \boldsymbol{C}_k^T$

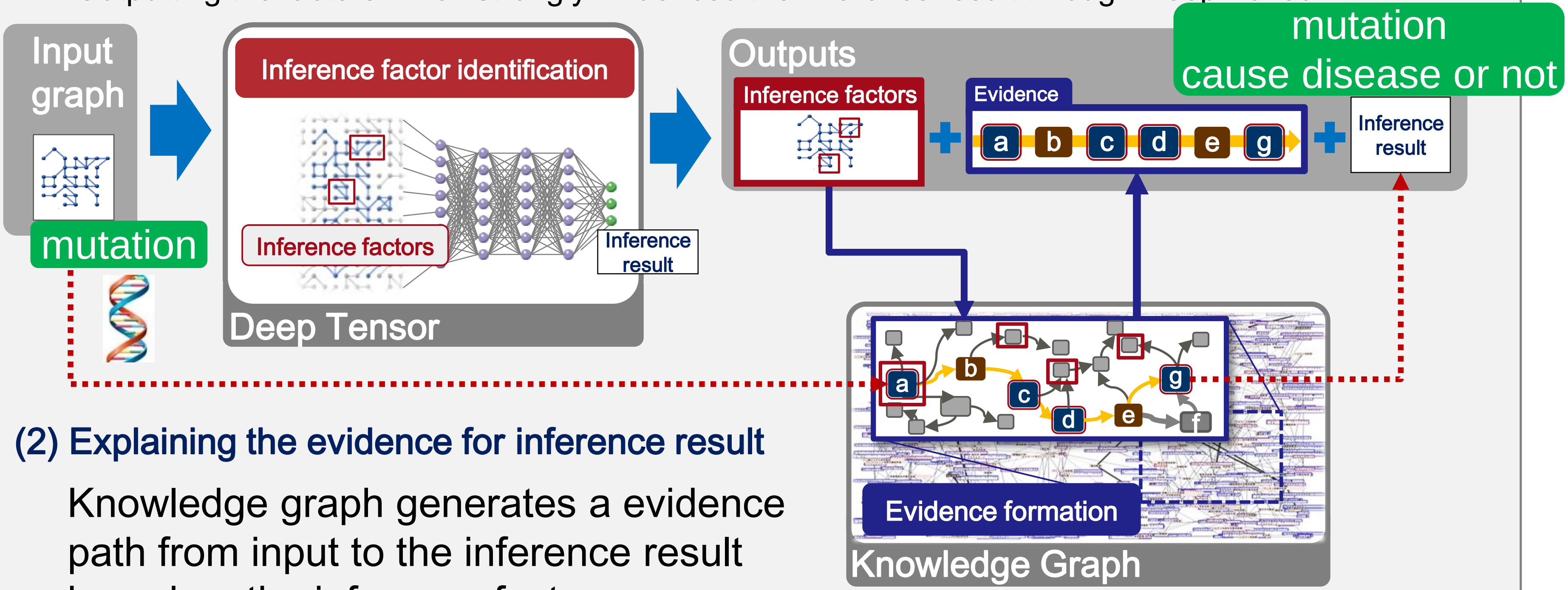
Recap: Overview

Deep Tensor + knowledge graph

- Knowledge graph generates input graph (extracting subgraph representing about the mutation)
- Deep Tensor infers which the mutation cause disease or not, and output inference factors.
- Knowledge graph makes evidence graph based on the inference factors.

(1) Explaining the important factor for the inference

Outputting the factors which strongly influenced the inference result through Deep Tensor



(2) Explaining the evidence for inference result

Knowledge graph generates a evidence path from input to the inference result based on the inference factors.

Conclusion

- Explainable AI is a key technology for cooperation of AI and humans

- We developed a prototype of explainable AI
  - The explainable AI explains important part of input data, and the evidence that explains the important part and inference result.

- We are now trying to proof of the concept of explainable AI, by cooperating with several medical groups.

FUJITSU

shaping tomorrow with you