大規模言語モデルを用いた神経科学論文からの解剖学的神経投射の記述を抽出するためのプロンプトの考案

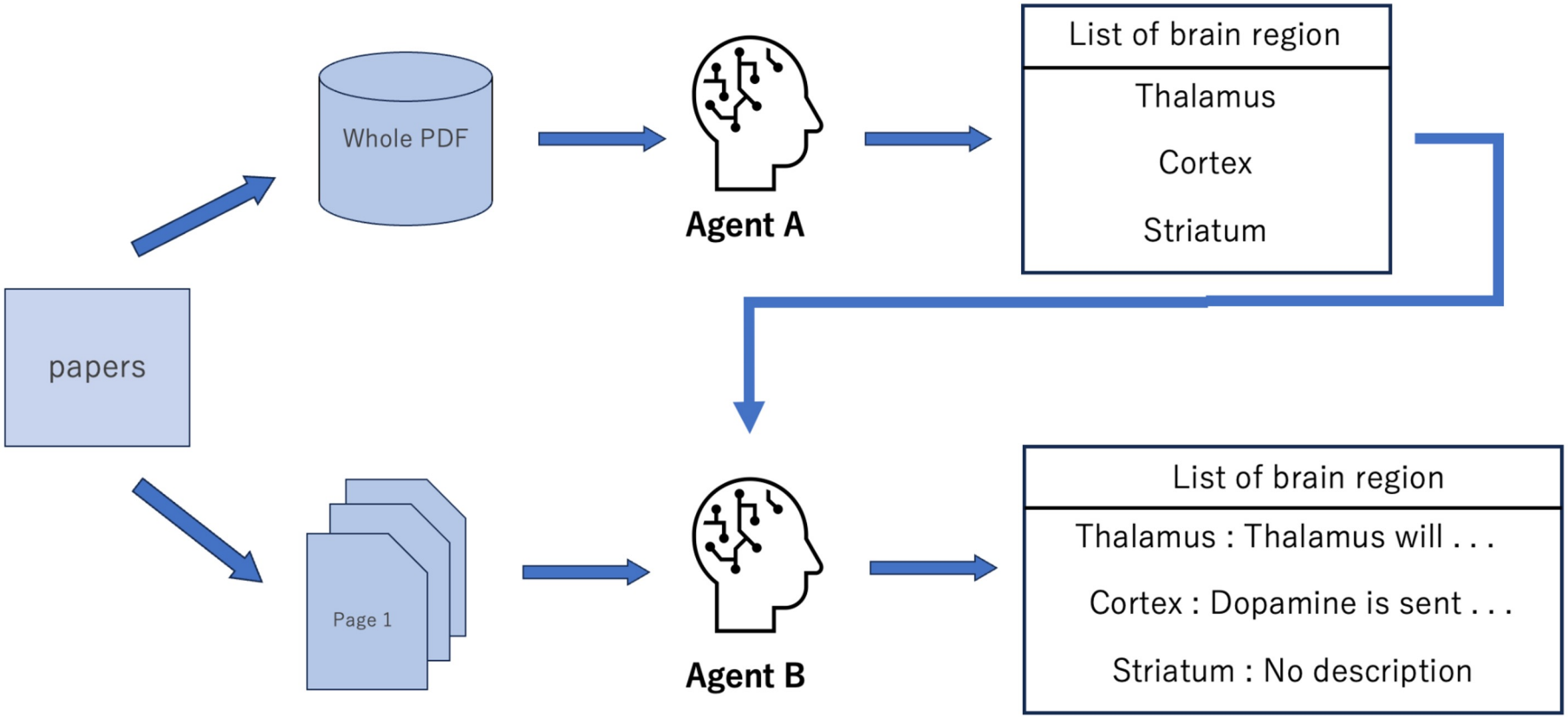
堀口　維里優
東京大学

BIFの自動抽出

Sender Circuit ID (sCID)	Receiver Circuit ID (rCID)	Reference ID	Taxon	Pointers on literature	Pointers on figure	In-depth literature
LEC_II	DG	Moser, 2010		All layer II principal cells issue an axon coursing straight toward the angular bundle, where the axon continues to its main targets in the DG and CA3/ CA2 (Fig. 17.1).	Fig.17-1	Tamamaki, 1993
LEC_II	DG	Witter, 2000		McNaughton and Barnes, 1977	Fig. 3	
LEC_II	CA3	Moser, 2010		All layer II principal cells issue an axon coursing straight toward the angular bundle, where the axon continues to its main targets in the DG and CA3/ CA2 (Fig. 17.1).	Fig.17-1	Tamamaki, 1993

現在BIF情報の抽出は人力で行われているが，時間がかかる

大規模言語モデルの導入



開発の現状

エージェントA

```
You are an agent that is specialized for extracting information from article.
In the article, there are some brain regions appeard.
You have you excerpt EACH and EVERY one of the regions.
Give the answer in .json form like above.


==answer==
[
  {"name" : "Basal ganglia"},
  {"name" : "Thalamus"},
   ,,,

]
```

対象論文：Parkinson’sdiseaseasasystem-leveldisorder
の実行結果

```
[
  {"name" : "Cerebellum"},
  {"name" : "Thalamus"},
  {"name" : "Basal ganglia"},
  {"name" : "Motor cortex"},
  {"name" : "Cortex"},
  {"name" : "Pedunculopontine nucleus"},
  {"name" : "Prefrontal cortex"},
  {"name" : "Subthalamic nucleus"},
  {"name" : "Striatum"},
  {"name" : "External globus pallidus"},
  {"name" : "Internal globus pallidus"},
  {"name" : "Substantia nigra pars compacta"},
  {"name" : "Substantia nigra pars reticulata"},
  {"name" : "Primary motor cortex"},
  {"name" : "Supplementary motor area"},
  {"name" : "Pre-supplementary motor area"},
  {"name" : "Ventral intermediate nucleus of thalamus"},
  {"name" : "Pontine nuclei"},
  {"name" : "Substantia nigra"}
]
```

取りこぼしを複数のLLMや
画像データを使って補完する

開発の現状

エージェントB

```
You are an agent that is specialized for extracting information from article.
In the article, there are some discription about connection between brain regions.
I will specify the region name.You have to exverpt information below for the regions.

sender circuit : name of the brain region
connection discription : if there is a discription about connection ( YES/NO and if NO put "ND" to
reciever circuit : name of the brain region connected to sender
Pointers on literature : Texts referring to projection in the literature.exerpt EXACTLLY from the
Pointers on figure : Drawings showing projection within the literature. (Fig.1. etc)
Size : Number of axons included in this projection

If you can't find the information simply put ND(No Discription)
Extract all the information above and give the answer in .json like below.

Now do this for { region name }

==answer==
[
  {"sender circuit" : "Basal ganglia",
   "connection discription" : "YES",
   "reciever circuit" : "Striatum",
   ,,,
  }
]
```

```
[
  {
    "sender circuit": "Cerebellum",
    "connection discription": "YES",
    "reciever circuit": "Thalamus, Motor Cortex",
    "Pointers on literature": [
      "The neuronal code (s) of the cerebellum. J. Neurosci. 33, 17603–17609 (2013)
      "Activity of neurons in cerebellar-receiving and pallidal-receiving areas of
      "The cerebellum in Parkinson's disease. Brain 136, 696–709 (2013).",
      "Pallidal dysfunction drives a cerebellothalamic circuit into Parkinson tremo
    ],
    "Pointers on figure": "ND",
    "Size": "ND"
  },
  {
    "sender circuit": "Thalamus",
    "connection discription": "YES",
    "reciever circuit": "Cerebellum",
    "Pointers on literature": "Cer and BG receive input from, and send output to, C
    "Pointers on figure": "ND",
    "Size": "ND"
  },
  {
    "sender circuit": "Basal Ganglia",
    "connection description": "YES",
    "receiver circuit": "Cortex and Cerebellum",
    "Pointers on literature": "Traditionally, the basal ganglia have been considere
    "Pointers on figure": "ND",
    "Size": "ND"
  }
]
```

- エージェントAで取得した各脳部位の投射情報を抽出
- 定量評価を進行中

開発の現状

エージェントC

```
You are an agent that is specialized for extracting information from article.
In the article, there are information appeard.
explained below.

author : author of the article
year : year of the publication of the article
Taxon : animal species used to investigate in the article
Measurement method : A measurement technique used to determine anatomical structure
(If there is no appropriate option, please ignore the error and write in)
Doc. Link : Link of literature i.e. URL
Journal names : Name of the journal, book, etc. in which the literature is published
Literature type : Type of literature content (review, text, etc.)

Extract all the information above and give the answer in .json like below.

==answer==
[
  {"author" : ". . ."},
  {"year" : "05/12/2012"},
  {"Taxon" : "Callithrix jacchus"},
   ,,,

]
```

対象論文：Parkinson’sdiseaseasasystem-leveldisorder
の実行結果

🤗 Taxonが取得できている

😢 Link を捏造する

```
[
  {"author" : "Daniele Caligiore, Rick C Helmich, Mark Hallett, Ahmed A Moustafa, Lars Timmermann,
  {"year" : "2016"},
  {"Taxon" : "Human"},
  {"Measurement method" : "Not specified in the document"},
  {"Doc. Link" : "https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5516580/pdf/npjparkd201625.pdf"},
  {"Journal names" : "npj Parkinson's Disease"},
  {"Literature type" : "Review Article"}
]
```

まとめ

現在は人力で行っているHCDの記述を**自動化**

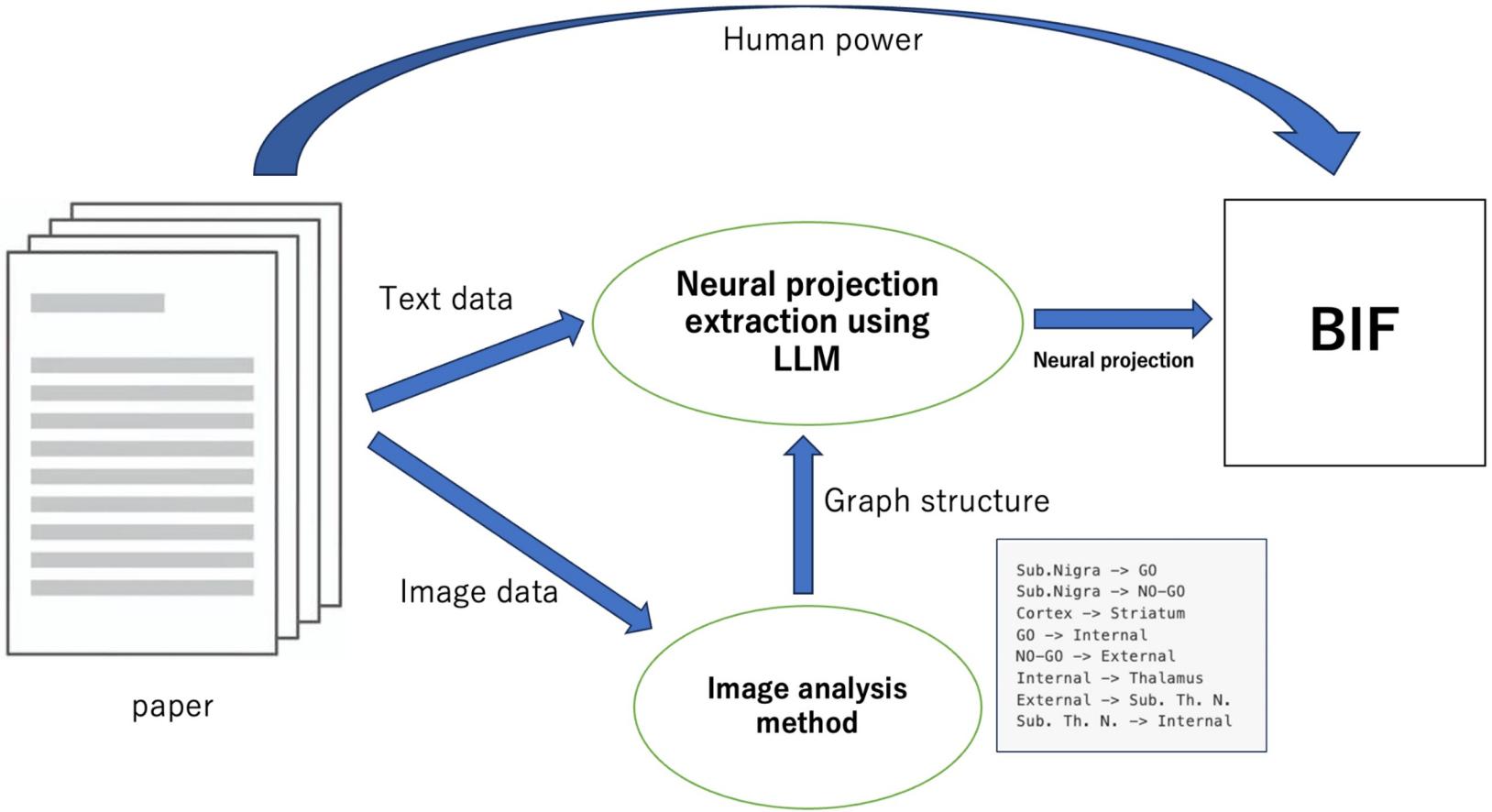
エージェントA：文献の本文から脳の領域のリストを抽出

エージェントB：脳領域間の神経投射に関連する説明を抽出

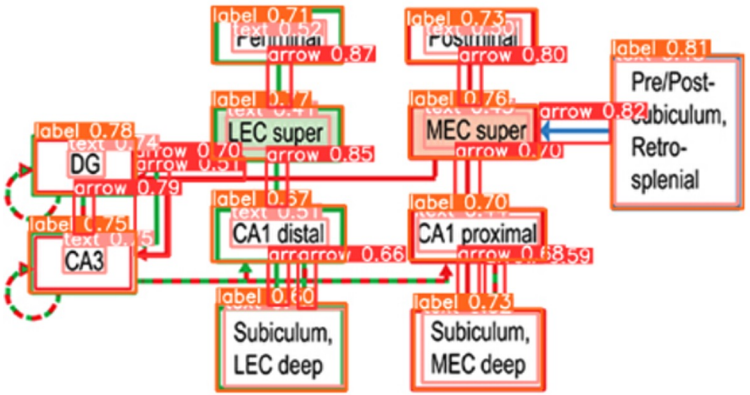
エージェントC：論文からHCDに記述すべきメタ情報を抽出

今後は定量的評価を実施

付録：全体のフレームワーク



付録：グラフ画像の利用



Pre1Post-Subiculum.Retro-splenial Go To **MEc super**
Postrhinal Go To **MEc super**
MEc super Go To **CA1 proximal**
CA1 proximal Go To **Subiculum. MEC deep**
CA1 proximal Go To **Subiculum. MEC deep**
Perirhinal Go To **LEC super**
LEC super Go To **CAl distal**
CAl disal Go To **Subiculum. LEC deep**
CAl disal Go To Subiculum. LEC deep
DG Go To **CA3**
DG Go To **CA3**
CA3 Go To **DG**