AI開発の話。

AIの作り方は３ステップだ。

１．AIに学習させるためデータを集める（収集）

２．集めたデータを使ってAIを作る（学習）

３．完成したAIを使う（導入）

これを「業務システム」の臓器のセグメンテーション（画像上で臓器が移っている場所をマーカーで表示する）に当てはめてみる。

１．

セグメンテーションされた画像を沢山作る必要がある。

セグメンテーションされた画像をどうやって作るのか？

　→データは信頼できる、正しいものでなければならない。

　→DICOMという特別なフォーマットに対して、どうやってデータを作るのか？

**・データを集めるためのツールがいる。**

**・そのツールを専門医が扱い、データを作る必要がある。**

２．

プログラミングスキルが必要。

学習は、コンピューターを動かして行う。動かすためのプログラミングがいる。

マシン（ハードウェア）が必要。

　GPUを搭載した高価なマシンがいる。

**・マシン構築（インフラの構築）**１。のデータをマシンに送るインフラが必要だ。

**・学習アルゴリズムを構築するためのプログラミング**

**・学習を実行するジョブスケジューラ**

　効率良く学習を進めるなら、複数の学習計画を並列に進めていくべきだ。

　そのためには、マシンに対して複雑なジョブスケジュールをする必要がある。

　一方の学習計画にはどれくらいのGPUリソースを注ぐべきか？

　最短で学習を終えるには、どのような順番で学習を行うべきか？

３．

実際に「業務システム」へ作成したAIを導入するには、それを前提としたシステム設計が必要。

**・AI導入できるシステム設計**

------------------------

以上のステップをAI開発者は実行できるだろうか？

AI開発者とはどんな人が想定できる？

・医療従事者

・社内SE

　業務システムの開発に従事

・AI人材

　AIアルゴリズム開発が可能、知見を有する。

　AIを用いたシステム設計

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | データを集めるツール | データを作る | 学習アルゴリズムのプログラミング | マシン構築、 | **ジョブスケジュール** | 業務システムへ導入 |
| 医療従事者 | △ | 〇 | △ | △ | **×** | × |
| 社内SE | 〇 | × | × | △ | **×** | 〇 |
| AI人材 | 〇 | × | 〇 | △ | **×** | △ |

２．集めたデータを使ってAIを作る（学習）際の、ジョブスケジュールを実行できる人がいない。

AIを作るうえで一番時間がかかるのは「２．集めたデータを使ってAIを作る（学習）」。

ここを効率良くする「マシン構築」「ジョブスケジュール」が重要。

**（この辺の数字がほしい）**

でも、AI開発者ができる領域ではない（苦手、本業じゃない）。

だから、ここにサポートを入れたい。

（競合、市場）

KAMONOHASHI

マシンを用意すれば、OSSをインストールすることで、Webサイトから操作できる。

<https://kamonohashi.ai/>

AIRI（NVIDIA）

マシン、と、マシン専用のフレームワーク、学習ジョブスケジューラを搭載。

<https://it.impressbm.co.jp/articles/-/15936>

他にもあるはず。

2017、2018年ごろから、提供するところ増えている。らしい。

ジョブスケジュールは難しいのか？

「複数の学習を並列にやった方がいい。でも、マシンリソースをどう割り振るのが最適か分からない」

みんながやっていること。

・学習計画を一つずつ順番にやる。

　　CPUやGPUを一つの学習でフル稼働することはない。

　　マシンパワーを全部使っているとは言えない。

→複数の学習を並列にやった方が良い。

・リソースを均等に割り振る。

　　均等に割り振ることが、最適とは言えない。

　　均等に割り振ることで、リソース不足し、非効率になっている可能性がある。

→どのようにリソースを割り振ることが、最適であるか分からない。

「複数の学習を並列にやった方がいい。でも、マシンリソースをどう割り振るのが最適か分からない」　→　なぜ、分からない？

学習の時間は、CPUとGPUに強く依存する。

画像系の学習では、画像ピクセルごとに計算を行う。この計算は各画像ピクセル間で並列に行うことができる。

つまり、理論的には画像ピクセル数分だけ並列処理を書けることで高速化できる。

しかし、画像解像度だけの並列処理はCPUは不可能。GPUの得意分野。

だから、GPUが学習の時間に強く影響を与える。

CPUは、データの前処理に影響を与える。

前述のとおり、メイン処理はGPUで行うが、CPUが得意な処理もある。なので、GPUの計算が早くてもCPUの前処理が終わらないと、次の処理へ進まないことになる。

（備考）

メモリは、必要量がある。

基本的に学習の時間に影響は与えない。ただし、メモリ不足が発生すると、処理停止する恐れがある。

HDD、SSDは関係ない。優先度は極めて低い。

CPU計算、GPU計算、ともにメモリ上で演算を行うから。

学習データのI/Oはあるが、GPUなどの計算の方がはるかに重い。

「CPU」と「GPU」を引数にしたときの、「学習時間」が予測できない。

学習データ、学習アルゴリズムによって、全く変わる。

どのようになるかは、AI開発者にも分からない。

本当に何もわからないか？

・AI開発者は最低条件が分かっている（制約条件）

　　学習アルゴリズムで、CPU、GPUの最低条件は分かる。

　　学習データのサイズで、メモリの必要量は分かる。

・同じアルゴリズムを繰り返し行えば、傾向が見えてくる。

　　アルゴリズムを変えなければ、学習データのサイズに対して線形傾向がある。

　　収束の速度は変わるので全体の時間は読めないが、１世代の計算時間は見えてくる。

「制約条件をユーザーの入力させ、CPUとGPUを引数に実際の学習時間を観測し、学習時間が最短になるCPU、GPU値を探索する方法を考えられないだろうか」

　学習は、データを変えつつ、繰り返し実行する。

　だから、その間に少しずつCPU、GPUの使用効率が上がることは価値があるのではないか？

　自動制御できないにしても、その傾向をグラフなどでAI開発者に示すことに価値はないか？

外的要因による制約条件に対応する。

・優先度の高い研究（人）に対して、優先的にリソースを割り当てる。

・スケジュールの予約

（他に何かあるかな？）

（話が小さいかな）