

😊TensorFlow や BERT などを使って…

😊AI エンジニア気分になっている墮落なエンジニアへ

☑CNN や RNN (LSTM、GRU)、RL (DQN), Attention などの理論/技術の理解や開発は、面倒だ

😊特に数学的背景の理解は難しく判らない🙄

☑米国製フレームワークを使って検索や分類、質問応答、チャットボットなどのアプリ開発をしている…AI エンジニア気分の君へ

♣機械学習フレームワーク

{NumPy, Pandas, SciPy, Matplotlib, scikit-learn, dlib, …}

♥深層学習フレームワーク

{TensorFlow (Agent), BERT, Keras, PyTorch, Caffe, DL4j, Paddle²,
MXNet, GPT-3, MSCognitiveToolkit, Chiner, …}

◇日本語形態素辞書

{IPA, Unidic, NEologd, …}

♣日本語解析器

{MeCab, KNP, JUMAN(++), GiNZA, COTOHA, Sudachi (Py), JANOME, …}

□意味とは？ 意味理解とは？ 概念とは？ 深層学習とは？ 活性化関数とは？

□係り受けとは？ 体言止めとは？ ロールとは？ 述語論理とは？ 双線形とは？

□欧米語と日本語の表現の違いは？ Onomatopes とは？ 双曲線関数 tanh とは？

□DeepLearning と Quasi-NewtonMethod との違いは？

(判らない人は、このサイトの「技術情報」の「基礎知識確認試験」を参照せよ)

☑開発しているアプリは、思うような成果/精度が出ていない

☑製品化に必要な精度基準 95%以上…は、難しい現状だ

□精度が思うように出ていない原因は…何だと思いませんか？

□良質なビッグデータが足りないから…ですか？ 違うでしょう

□いろいろ原因はあるが、欧米語と日本語の意味表現の多様性と深さの相違は大きい

■まず Embedding で素性関数から生成される「ベクトル」に原因のひとつがある

□Word2Vec で生成された周辺語だけでは意味解析をする上で情報が足りない

□周辺語に文字種や品詞、固有表現を追加しても…まだまだ情報が足りない

□何が足りないのか…？

✓照応解析と階層的な概念タグそして文脈で変化する意味タグの情報などが足りない

□これらの情報が Embedding 層で補充されて初めて米製 Framework が使えるものになる

◇照応解析とは、照応詞(代名詞, ゼロ代名詞)から先行詞を推定すること

パソコン

(例文「それを使って。」)

◇もちろん、照応解析も Attention 付 LSTM で解析ができるが、面倒な事前処理が必要だ

◇階層的な概念タグとは、名詞だけの簡単な固有表現(場所、組織、…)でなく、助詞や副詞など、すべての品詞に関する概念を階層的に構築/構造化されたタグのこと

組織 方向 移動

(例文「会社へ行く。」)

◇意味タグとは、複数の単語が連なることで固有表現や意味が変化する意図を表すタグのこと

勤務先 出社

予算 オーバー

(例文「会社へ行く。」) (例文「足が出る。」)

・要約文生成なら、概念/意味タグを頼りに意味理解された抜粋文ではない生成された要約文ができる (seq2seq や GPT-3 のような方法論では精度が低く使いものにならない)

- ・ Q&A なら、質問に対して「問い返し/聞き返し/切り返し」などのナビやガイドで回答に導くことができる（「問い返し」ができるシステムは世界に現存していない）
- ・ Chatbot や雑談なら、トピックスやパーソナリティに合った話題で会話が楽しめる
- ・ 相談や診断なら、詳細な質疑応答で解決策や処方箋までを示唆することができる

△▽ニューラル言語モデルへの変遷△▽

□ふた昔前は、「統計的確率言語モデル」 ⇨ 頻度論による統計的確率モデル

- ・ 条件付き確率場 (CRF、マルコフ確率場)
- ・ Bayes の定理、Bayesian
- ・ Entropy モデル（結合、条件付き、相対、最大）
- ・ ジェンセンの不等式、ラグランジュの未定係数法
- ・ EM アルゴリズム、反復スケールリング法
- ・ 推定法（ヘルドアウト、グッド・チューリング）
- ・ N グラムモデル (N-gram)
- ・ 隠れマルコフモデル (HMM)
- ・ 確率文法（確率文脈自由文法、確率依存文法）

□ひと昔前は、「機械学習」 ⇨ 学習による分類系言語モデル

- ・ 教師付き回帰
 - ・ 最小二乗学習
 - ・ スパース学習、ロバスト学習
- ・ 教師付き分類
 - ・ サポートベクトル分類
 - ・ アンサンブル分類
 - ・ 確率的分類
 - ・ 系列データ分類
- ・ 教師なし学習
 - ・ 異常検出
 - ・ スパース性と次元削除
 - ・ クラスタリング
- ・ オンライン学習
- ・ 半教師付き学習
- ・ 転移学習
- ・ マルチタスク学習
- ・ 潜在的配分法 (LDA) Topics や単語出現確率の推定をするモデル
- ・ Word2Vec (CBoW, Skip-gram)
 - ・ 単語分散表現 (Embedding) の素性 Tensor での加減や Analogy ができる加群モデル
 - ・ 未知語の素性や単語間の類似性が求められる Bag-of-Words 集合モデル

□ちょっと前は、「深層学習」 ⇨ 学習に注目したニューラル言語モデル

- ・ CNN 畳み込みとプーリングで画像処理が発展
- ・ RNN 再帰関数で時系列データを効率的に処理
- ・ LSTM Gate 付 RNN で、L と S の記憶セルを共に入力
- ・ GRU Gate 付 RNN で、LSTM の記憶セルなし (reset, update を内包)
- ・ Bandit 探索と知識の最適化問題（確率的、敵対的、最適腕、連続輪腕）
- ・ RL 強化学習 (TD, Q) 勾配法が要らない勾配法 BellmanEquation
- ・ DQN 強化学習の深層学習版
- ・ Attention 注目という分布とウエイト行列 (seq2seq に有効)

※隠れ層は、間違えた時に「何故、このような解答が出たのか？」という疑問に答えられない。