人工知能学会研究会資料
SIG-AGI-001-05

注意ベースモデルが結び付ける 脳と汎用人工知能そして産業応用

Attention-based models combine brain and artificial-general-intelligence for industrial applications

礼王懐成 [1]　山川宏 [2]

Author 1[1], Hiroshi Yamakawa[1,2]

[1] メタップス
[1]Metaps
[2]株式会社ドワンゴ ドワンゴ人工知能研究所
[2] DWANGO ARTIFICIAL INTELLIGENCE LABORATORIY, DWANGO CO LTD

Abstract: Recently Attention-based neural network is successfully applied to various tasks such as machine translation, caption generated from the image, the caption generation from video, speech recognition, generation of image from the caption. But those models are lacking in versatility.

In this paper, we review the various attention-based neural models, and propose a generalized model of attention-based neural models using the knowledge of the brain and artificial general intelligence.

１、はじめに

最近 Attention-based ニューラルネットを用いた機械翻訳、画像からキャプション生成、動画からキャプション生成、音声認識、キャプションから画像の生成、質問自動応答などで成果を上げており、それらの技術をレビューし、その限界を超えるべく、脳科学で得られている注意モデルや記憶モデルの知見を取り入れた汎用人工知能の枠組みを提案する。

その汎用人工知能の成果を、現在の産業で用いられている一般の人工知能に適用することで、学習データの不足の解決や精度の改善を目指す。

２、Attention-based ニューラルネットの紹介

注意メカニズムは、２つの目的で用いられる。

一つ目は、高次元の入力を部分要素に分けることで計算コストを下げる。

２つ目は、入力の異なる側面に注目し、入力のどの特徴がもっとも関連する情報を出力に寄与するかを計算することで出力のクオリティを改善することができる。

Neural machine translation by jointly learning to align and translate. [11]

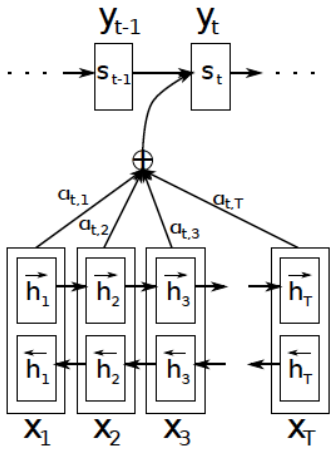

図１： Neural machine translation with soft attention

ソフトアテンションによる機械翻訳システム。

$$p(y_i|y_1,\ldots,y_{i-1},\mathbf{x}) = g(y_{i-1},s_i,c_i),$$

$$s_i = f(s_{i-1},y_{i-1},c_i)$$

$$c_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j$$

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x} \exp(e_{ik})},$$

$$e_{ij} = a(s_{i-1}, h_j)$$

翻訳ソース言語の内部状態 h_j と翻訳ターゲットの言語の内部状態 s_(i-1)との相関を softmax したものが注意信号の α である。

Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. [2]

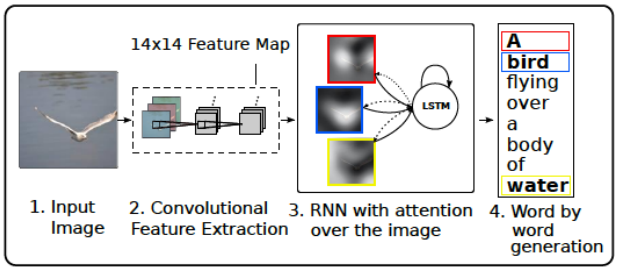

図 2： Neural image caption generation with visual attention.

$$\mathbf{i}_t = \sigma(W_i E\mathbf{y}_{t-1} + U_i\mathbf{h}_{t-1} + Z_i\hat{\mathbf{z}}_t + \mathbf{b}_i),$$
$$\mathbf{f}_t = \sigma(W_f E\mathbf{y}_{t-1} + U_f\mathbf{h}_{t-1} + Z_f\hat{\mathbf{z}}_t + \mathbf{b}_f),$$
$$\mathbf{c}_t = \mathbf{f}_t\mathbf{c}_{t-1} + \mathbf{i}_t \tanh(W_c E\mathbf{y}_{t-1} + U_c\mathbf{h}_{t-1} + Z_c\hat{\mathbf{z}}_t + \mathbf{b}_c),$$
$$\mathbf{o}_t = \sigma(W_o E\mathbf{y}_{t-1} + U_o\mathbf{h}_{t-1} + Z_o\hat{\mathbf{z}}_t + \mathbf{b}_o),$$
$$\mathbf{h}_t = \mathbf{o}_t \tanh(\mathbf{c}_t).$$

$$e_{ti} = f_{\text{att}}(\mathbf{a}_i, \mathbf{h}_{t-1})$$
$$\alpha_{ti} = \frac{\exp(e_{ti})}{\sum_{k=1}^{L}\exp(e_{tk})}.$$
$$p(s_{t,i} = 1 \mid s_{j<t}, \mathbf{a}) = \alpha_{t,i}$$
$$\hat{\mathbf{z}}_t = \sum_i s_{t,i}\mathbf{a}_i.$$

注意信号は、LSTM の隠れ状態の前の状態 h_t-1 と部分分解された画像のアノテーションベクトルとの相関関係から softmax を経て計算される。

Teaching Machines to Read and Comprehend. [3]

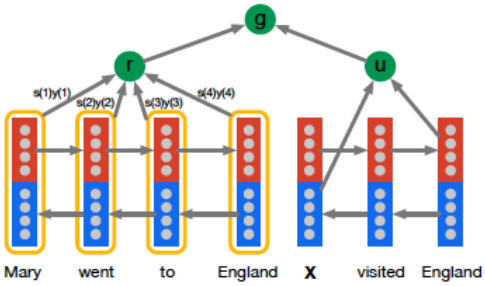

図 3： Attentive Reader.

PropBank フレームを用いて様々な質問形式を生成し、それを注意ベースで学習させる質問応答システム。

$$m(t) = \tanh\left(W_{ym}y_d(t) + W_{um}u\right),$$
$$s(t) \propto \exp\left(\mathrm{w}_{ms}^{\intercal}m(t)\right),$$
$$r = y_d s,$$

Reasoning about Entailment with Neural Attention [4]

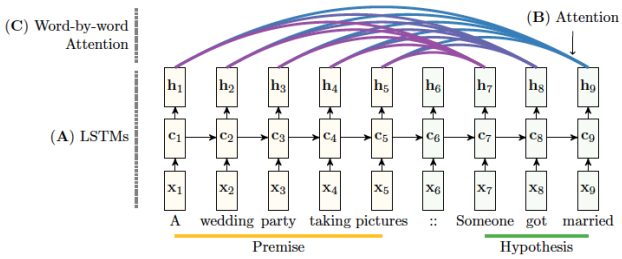

図 4： Recognizing textual entailment using two LSTMs.

図 4 のように Premise 文章が，Hypothis 文章と含意関係であることを学習するシステム。

$$\mathbf{M} = \tanh(\mathbf{W}^y\mathbf{Y} + \mathbf{W}^h\mathbf{h}_N \otimes \mathbf{e}_L) \qquad \mathbf{M} \in \mathbb{R}^{k\times L}$$
$$\alpha = \text{softmax}(\mathbf{w}^T\mathbf{M}) \qquad \alpha \in \mathbb{R}^L$$
$$\mathbf{r} = \mathbf{Y}\alpha^T \qquad \mathbf{r} \in \mathbb{R}^k$$

含意関係認識において、Premise の LSTM で処理された出力 Y と Hypothis の出力の相関を softmax 関数で注意信号に変換する。

以上に示したものは比較的に簡単なモデルである。より複雑なものとしては、注意ベースの質問応答や対話システムの研究を紹介する。

Ask Me Anything: Dynamic Memory Networks for Natural Language Processing. [6]

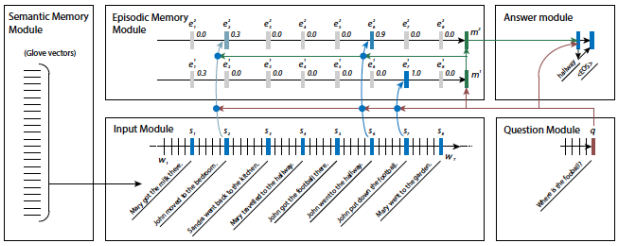

図 5： Dynamic Memory Networks

Dynamic Memory Networks は入力された文章に対して、質問応答形式で multi task を行うことも可能で、文章の内容だけでなく、文章の品詞、感情解析、言及参照などに答えることが可能である。

Semantic memory module:
単語の意味を参照のためのモジュールで辞書なども用いる。しかし今回は Glove を用いた埋め込みのベクトル化の実装のみである。
Input Module:
時間順序のタグをつけ Semantic memory module から埋め込みベクトル化をした結果を受け取って、Episodic Memory Module へ出力する。
Question Module:
RNN の一種である GRU で処理した結果を Episode Memory で処理する。

Answer Module:
Episodic Memory Module からのベクトルデータを単語に変換する。

Episodic Memory Module :
DMN の心臓部であり、質問に対して文章の入力系列から関連する事柄がある場合に注意信号が生成される。異なる事柄は異なるパスに保存し、それを組み合わせることで、transitive inference(推移的推論)を行うことができる。

$$z(c,m,q) = [c,m,q,c\circ q,c\circ m,|c-q|,|c-m|,c^TW^{(b)}q,c^TW^{(b)}m]$$
$$G(c,m,q) = \sigma\left(W^{(2)}\tanh\left(W^{(1)}z(c,m,q)+b^{(1)}\right)+b^{(2)}\right)$$
$$g_t^i = G(c_t,m^{i-1},q)$$
$$m^i = GRU(e^i,m^{i-1})$$
$$h_t^i = g_t^iGRU(c_t,h_{t-1}^i)+(1-g_t^i)h_{t-1}^i$$
$$e^i = h_{T_C}^i$$

ゲート g が注意制御に相当する。

Attention with Intention for a Neural Network Conversation Model. [12]

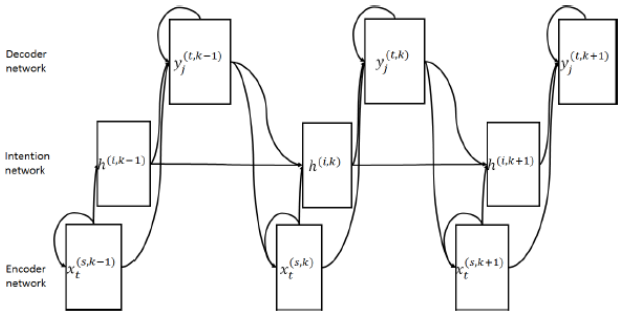

図 6： The attention with intention (AWI) model.

図 6 にあるような encoder network は発話されてた文章 x を RNN で処理し、Intention と Decoder network(それぞれ RNN)へ出力してから発話を生成する。
Intention は、encoder と decoder と過去の自身からの入力を受け、decoder へは、decoder の隠れ層の初期状態として出力される。
対話目的などを含むコンテキストの状態が保存されている Intention network は対話ターンごとに過去の自身の状態の入力を受けるため、それを保持し続けることができる。
アテンションは、decoder の隠れ状態 h と encoder の出力の相関を softmax したものになっている。
つまり、

$$\alpha_{jt} = \frac{\exp e_{jt}}{\sum_m \exp(e_{jm})}$$

$$e_{jt} = a(h_{j-1}^{(t)}, c_t^{(s)})$$

encoder と decoder の間に Intetion 層をつなぐことで、そのコンテキストごとに注意を生成することが可能となる。

Neural Programmer: Inducing Latent Programs With Gradient Descent.[7]

Neural Programmer は RDB のようなデータベースに対して SQL ではく自然言語の質問文を用いて検索を行うシステムである。

四則演算などの基本的なオペレータを別に用意する。

LSTM ではこのような処理は精度が悪い。

ある程度人間の手で用意する必要がある。

複雑な質問に対して、複数のオペレータを組み合わせることで対処できる。

テキスト情報と数字情報が混在する RDB のデータベースにおいて、テキストマッチングを行う場合に注意ベースモデルを使用している。

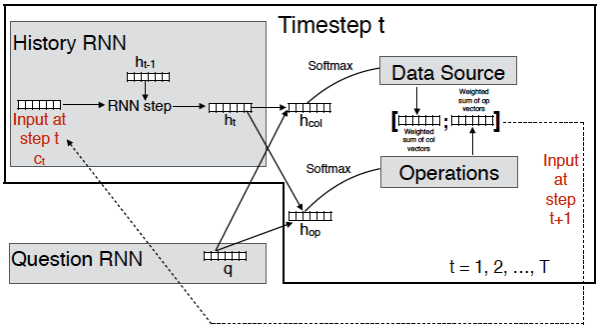

図 7： Neural Programmer

一つのクエリーを 4 つのタイムステップに分解して、一ステップごとに実行したオペレーターとデータを history RNN に保存し、次のオペレータと組み合わせることで、以下のような複雑な質問に答えることができる。

" What is the sum of elements in column B whose field in column C is word:1 and field in column A is word:7?"

Type	Operation	Definition
Aggregate	Sum	$sum_t[j] = \sum_{i=1}^{M} row\_select_{t-1}[i] * table[i][j], \forall j = 1, 2, \ldots, C$
	Count	$count_t = \sum_{i=1}^{M} row\_select_{t-1}[i]$
Arithmetic	Difference	$diff_t = scalar\_output_{t-3} - scalar\_output_{t-1}$
Comparison	Greater	$g_t[i][j] = table[i][j] > pivot_t, \forall (i, j), i = 1, \ldots, M, j = 1, \ldots, C$
	Lesser	$l_t[i][j] = table[i][j] < pivot_t, \forall (i, j), i = 1, \ldots, M, j = 1, \ldots, C$
Logic	And	$and_t[i] = min(row\_select_{t-1}[i], row\_select_{t-2}[i]), \forall i = 1, 2, \ldots, M$
	Or	$or_t[i] = max(row\_select_{t-1}[i], row\_select_{t-2}[i]), \forall i = 1, 2, \ldots, M$
Assign Lookup	assign	$assign_t[i][j] = row\_select_{t-1}[i], \forall (i, j) i = 1, 2, \ldots, M, j = 1, 2, \ldots, C$
Reset	Reset	$reset_t[i] = 1, \forall i = 1, 2, \ldots, M$

図 8： 用意されているオペレータの一覧

$$row\_select_t[i] = \alpha_t^{op}(\text{and})and_t[i] + \alpha_t^{op}(\text{or})or_t[i] + \alpha_t^{op}(\text{reset})reset_t[i] + \sum_{j=K+1}^{C} \alpha_t^{col}(j)(\alpha_t^{op}(\text{greater})g_t[i][j] + \alpha_t^{op}(\text{lesser})l_t[i][j]) + \sum_{j=1}^{K} \alpha_t^{col}(j)(\alpha_t^{op}(\text{text\_match})text\_match_t[i][j], \forall i = 1, \ldots, M$$

row 選択の式。

3. 脳における注意機構

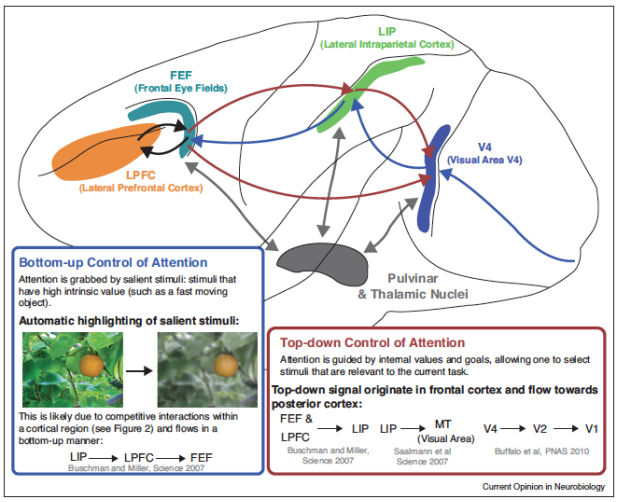

図 9： 脳のトップダウンとボトムアップ注意機構

脳の注意機構には、図 9 のように、ゴールや意図などの内部状態に起因するトップダウン注意と、外界の新規刺激や激しい刺激など予期しない刺激によるボトムアップ注意がある。[10]

次の 3 章ではこの知見を活かした認知アーキテクチャを紹介する。

3. 汎用人工知能

「汎用人工知能（Artificial General Intelligence, 以下 AGI）」という概念は「狭い AI」の反義語として現れ，この種の広範な汎化能力をもつシステムを指す. AGI のアプローチは「汎用知能」を，課題・問題個別の能力とは根本的に異なる性質として捉え，そのような性質の理解とそのような性質を呈するシステムの作成に直に取り組むものである. AI への既存のアプローチを記号的，創発的およびハイブリッドという三つのパラダイムに分けている.[8]

[7]では、リアルタイムの処理を行うために

図 9 のようにトップダウンとボトムアップの注意を別々に分けている。

また実行プロセス(オペレータ)を分けることで認識と実行を別々に分けることで、柔軟な問題解決が可能となる。

Goal-Driven Data Prioritizer (GDDP):

Novelty-Driven Data Prioritizer (NDDP):

Experience-Driven Process Prioritizer (EDPP):

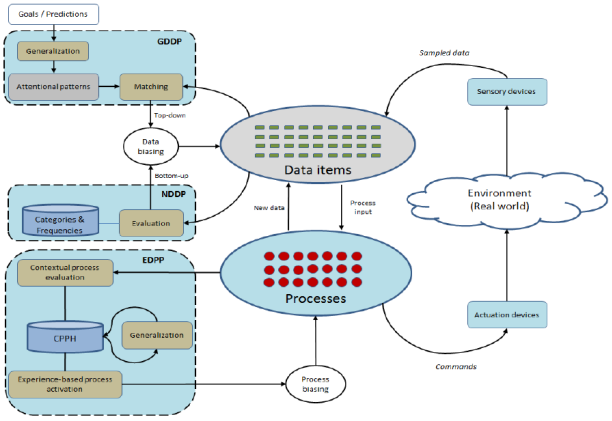

図 10： Attention driven cognitive architecture

4. 提案するモデルや研究課題

脳の機能や汎用人工知能で得られた知見により以下のモデルを提案する。

1) top-down, buttom-up アテンションの導入
2) 認知、推論、実行のモジュールを分ける。
3) メタ学習の導入
4) 注意信号を階層的に制御することで、multitask に対応できる。
5) LSTM など隠れ層の解析

1)については[9]を参考にリアルタイム処理を行ったり、例えば、対話システムの場合、外乱によって対話が脱線しても、外乱を buttom-up アテンションで処理することで、柔軟な対応が可能な対話システムを構築することができる。複数のオペレータを組み合わせる必要がある複雑なタスクでは、2)で列挙したようにモジュールが必要であり、それは２章で紹介した Neural Programmer [7]の研究でも示している。２）のようにモジュールに分けるにはあらけじめ人手で分けるか、学習して分類するかである。
人手を借りずに学習して分類する場合は 3)のような meta-learning の機構が必要である。
例えば、E2E memory-networks で行っているような演繹推論、帰納推論など様々な推論は、モジュールに分けることで異なるドメインで推論の再利用することだ出来る。また、脳のように解釈部と実行部を分けることで、task の分類を含めた、タスクの再利用のためのタスクの階層化やモニタリングなどを行うことができる。単純なオペレータを組み合わせ階層化する HTN ものと組み合わせることで、古典的な人工知能とニューラルネットの融合が行え、リアルの膨大なデータを用いて実装することが可能となる。

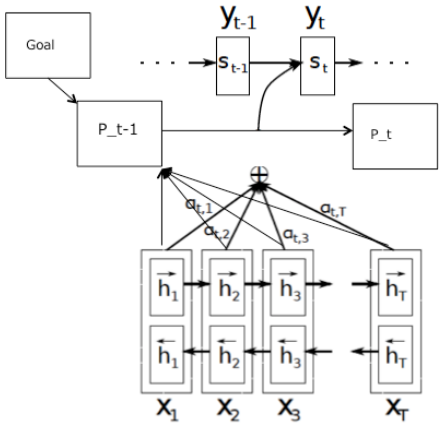

図 11： 汎用的な注意ベースモデル(提案)

4)において、脳では同じ感覚野(encoder)で異なるタスク(multi-task)を行えるが、今の注意ベースモデルでは、タスクごとに異なる encoder を用意する必要がある。
その制約の理由の一つは、既存の注意ベースモデルでは、decoder の隠れ層(h)と encoder からの入力の相関を正規化した変数を注意信号としているため、注意信号がタスク依存になってしまっている。
そこで、図 11 で示すように同じ encoder で multi-task が行えるよう、注意信号を階層的に制御を行うモデルを提案する。このモデルは、内部状態にゴールを持っており、それと外部の入力によってアテンションが生成される。
例えば、翻訳、要約、質問応答において共通の解釈層を使用することで、少ないデータでも効率的に学習できるのではないかと期待ができる。
また、5)では現状のニューラルネットで用いられている隠れ層は解析し難いものになっており、それに加え設計の指針がわかりにくい。それを解決する方法として隠れ層の構造を構築するにあたり脳や認知アーキテクチャを参考にして、個々の機能が相補的に上がるような汎用人工知能アーキテクチャの構築や時系列因子分析などを用いて内部のダイナミクスを解析が考えられる。

5. 結論

注意モデルを中心に、脳で得られた知見をもとに汎用人工知能と既存のデータ駆動のモデルとの融合モデルを提案した。また汎用化するにあたり様々な問題を明確化し、解決に向けての指針を示した。

参考文献

[１] Cho, Kyunghyun, Aaron Courville, and Yoshua Bengio. : Describing Multimedia Content using Attention-based Encoder–Decoder Networks, arXiv preprint arXiv:1507.01053, (2015)

[２] A., Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhutdinov, Richard Zemel, Yoshua Bengio.: Show, Attend and Tell: Neural Image Caption Generation with Visual Attention, ICML, (2015)

[３] Karl Moritz Hermanny, etc : Teaching Machines to Read and Comprehend, NIPS (2015)

[４] Rocktäschel, Grefenstette, Hermann, Kočiský and Blunsom: Reasoning about Entailment with Neural Attention. arXiv preprint arXiv:1509.06664. (2015)

[５] Sainbayar Sukhbaatar, Arthur Szlam, Jason Weston, Rob Fergus : End-To-End Memory Networks, NIPS (2015)

[６] Ankit Kumar, Ozan Irsoy, Peter Ondruska, Mohit Iyyer, James Bradbury, Ishaan Gulrajani, Richard Socher : Ask Me Anything: Dynamic Memory Networks for Natural Language Processing ,arXiv:1506.07285.(2015)

[７] A. Neelakantan, Q.V. Le, I. Sutskever : Neural Programmer: Inducing Latent Programs With Gradient Descent , arXiv:1511.04834v1, (2015)

[８] Ben Goertzel : 汎用人工知能概観, 人工知能　29 巻 3 号（2014 年 5 月）

[９] Helgi Páll Helgason, Kristinn R. Thórisson, Deon Garrett, Eric Nivel : Towards a General Attention Mechanism for Embedded Intelligent Systems, International Journal of Computer Science and Artificial Intelligence Vol. 4 Iss. 1, PP. 1-7 (Mar. 2014)

[１０] Earl K Miller, Timothy J Buschman : Cortical circuits for the control of attention. Current Opinion in Neurobiology, 23:216–222 (2013)

[１１] D Bahdanau Bahdanau, K Cho, Y Bengio : Neural machine translation by jointly learning to align and translate, ICLR 2015

[１２] Kaisheng Yao, Geoffrey Zweig, Baolin Peng : Attention with Intention for a Neural Network Conversation Model, arXiv:1510.08565v3 (2015)