

dwango

DWAL-TR-2015-004

Oct. 21 2015

第 11 回全脳アーキテクチャー勉強会 「Deep Learning の中身に迫る」 報告書

小林 輝行*¹ 上甲昌郎*² 中島 隆夫*³ 芦原 佑太*^{4,5} 山川 宏*⁶
Teruyuki Kobayashi Masao Joko Takao Nakajima Yuta Ashihara Hiroshi Yamakawa

*¹ 株式会社 COMPASS
COMPASS, Inc.

*² 日本アイ・ビー・エム株式会社
IBM Japan, Ltd.

*³ 三井情報株式会社
MITSUI KNOWLEDGE INDUSTRY CO., LTD.

*⁴ 電気通信大学
The University of Electro-Communications

*⁵ 株式会社クロスコンパス・インテリジェンス
XCompass Intelligence, Ltd.

*⁶ 株式会社ドワンゴ 人工知能研究所
DWANGO Co., Ltd. Artificial Intelligence Laboratory

 ドワンゴ
人工知能研究所
DWANGO ARTIFICIAL INTELLIGENCE LABORATORY

本報告書の概要

全脳アーキテクチャにおいて、中心的なテーマの一つである **Deep Learning** については、開発環境が充実しつつあるものの、現状では全脳アーキテクチャのための新しいアーキテクチャを創出するためには、ゼロからネットワークを構築することも想定される。そのためにも、**Deep Learning** の中身、例えば、中間層で何が起きているか等を理解することが重要である。第 11 回全脳アーキテクチャ勉強会として迎える今回は、物理学や、力学的解析の視点から **Deep Learning** の中身に迫る 2 件の講演と、**Deep Learning** を世界初の商用レベルでサポートされたオープンソースフレームワークである **DL4J** について講演が行われた。本報告書では、各講演の内容について報告する。

1. オープニング（ドワンゴ人工知能研究所 山川 宏氏）

<<NPO 法人全脳アーキテクチャ・イニシアティブ 創設賛助会員の募集について>>

今回の勉強会は、電通大の栗原聡氏の企画により実現されました。NPO 法人全脳アーキテクチャ・イニシアティブ(以下 WBAI)は 2015 年 8 月 21 日に創設され、現在は全脳アーキテクチャ・アプローチの理念に賛同/理解し、長期的にこの研究開発活動をご支援いただける会員を創設賛助会員として募集している。(詳しくは、<http://wba-initiative.org/support/> 参照) 創設賛助会員には、WBAI の活動に関する特典が提供される。

<<創設賛助会員 株式会社 Nextremer の紹介

今回の勉強会では、創設賛助会員である株式会社 Nextremer の紹介がオープニングで行われた。Nextremer 社では、AI-SAMURAI, 人“高知”脳, AI キツツキといった、多様性に重きを置いた対話エージェント型人工知能の研究開発ベンチャーである。(詳しくは <http://www.nextremer.com/> を参照)

2. 深層学習の学習過程における相転移（京都大学大学院情報学研究科 大関真之 氏）

<<講演内容>>

深層学習における中間層に何が起きているのか，統計力学的なアプローチを用いてその理解をしようとする取り組みについて講演された．このアプローチによって深層学習で扱われる教師なしデータによる pre-training，少量の教師ありデータによる fine-tuning について解析を行ったところ，統計力学・熱力学における相転移と同様の現象が発見されたというのが今回の講演内容であった．

<<統計力学的アプローチ>>

統計力学が扱う身近な現象の例として相転移現象がある．温度を上げると水が液体から気体に変化する現象，磁石が磁性を失う現象である．ミクロに見るとアボガドロ数個の分子の扱いが必要になり，解析が非常に困難な問題である．しかし，統計力学という処方箋を使って解析することで，温度という一つのパラメタを通じてマクロな現象として説明することが出来る．統計力学では「物質は自由エネルギーが最小となる平衡状態を実現する」という基本的なアプローチに基づいて考える．つまり，自由エネルギーを求めその最小値を算出することで，物質がどのような状態になるかを予言することができる．磁石の磁性を失う現象も本アプローチで予言される．

<<統計力学と機械学習の関係>>

統計力学は，ミクロの要素の集まりをマクロとして扱うという観点で，大量のデータを扱う機械学習分野とのアナロジーがある．大量のデータから意味のあるパターンを抽出する仕方を自動的に構築するという（深層学習に限らない）機械学習一般のフレームワークは，大量粒子の運動をマクロなパラメタ・観測量で記述する統計力学になぞらえることができる．つまり，物理学が築き上げてきた分析手法の“資産”を利用することができる．そもそも深層学習の中の代表的な手法である“Restricted Boltzmann Machine”（RBM）に名を残す Boltzmann は物理学者であっ

て、ニューラルネットの状態を確率モデルと捉えてパラメタ推定する際に用いられている最尤法は、統計力学における自由エネルギーの最小化(またはエントロピー最大化)と対応している。

<<深層学習における統計力学の関わり>>

深層学習のうち、Deep Belief Nets (DBN) のようなネットワークは、大量の教師無しデータによる Pre-Training, 少量の教師ありデータによる Fine-Tuning からなる。これら二つのステップにより、深層学習の性能がどのように実現されているのか、統計力学的に解析および実験を実施した。結論は「教師なし学習により識別器の潜在能力を高め、教師あり学習によってその識別器が持つ潜在能力まで到達させることで、高精度な学習が可能になっている」ということである。

<<実験:設定と結果>>

今回は、犬と猫の画像をわけるといった識別のしやすい構造を持つランダムなパターンについて単純パーセプトロン¹により識別する設定を考えた。

図1の左側は教師なしデータによる pre-training 後に少量の教師ありデータによる fine-tuning を行なった場合、右側が pre-training 後にやや多めの教師ありデータによる fine-tuning を行なった場合となっている。この実験では、教師なし学習後に教師あり学習を行っている。

レポーター註：下記、大関氏のサイトより引用

「Deep Learning という Deep じゃないと意味がないみたいに思われますが、扱っているデータをどのように表現するのが非自明だから多層にするのであって、データが表現できるような状況であれば多層化する必要はないわけです」

<http://www-adsys.sys.i.kyoto-u.ac.jp/mohzeki/Publish/BrandNew.html>

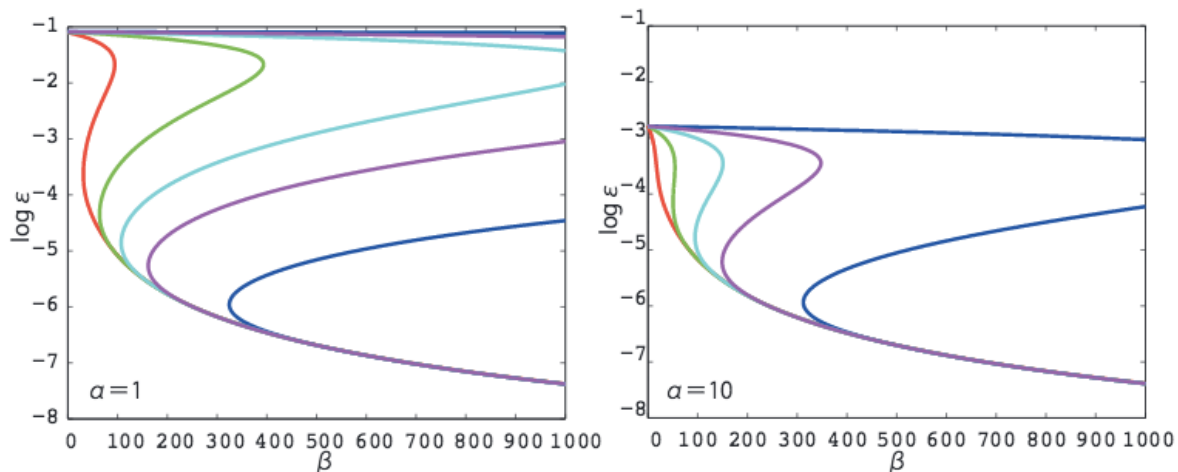


図 1

X 軸の値 β は教師なし学習によるデータ数（ノード数），Y 軸が誤差（値が小さいほうが優秀）， α が教師ありデータであり、色の違うプロットは赤→青の順に識別が難しくなっている．S 字のような曲線を描いているのが特徴で，これが相転移にあたる．ある時点を超えると急速に性能がよくなるのがわかる．

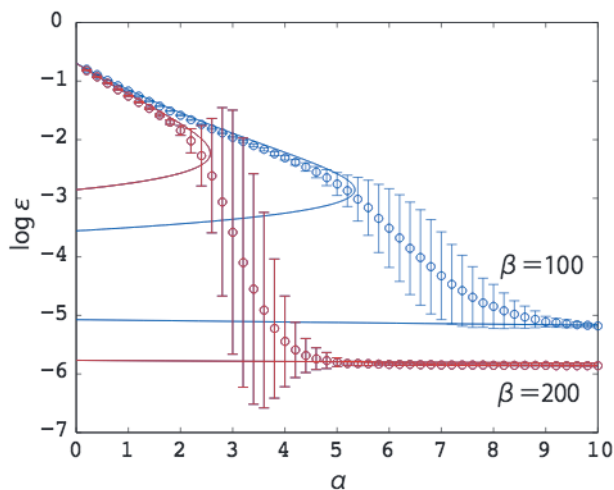


図 2

こちらの図は教師なし学習によるデータ数を一定値にし，X 軸を α （教師ありデータ数）としたものである．これらの結果から，教師なし学習によりポテンシャルが引き上がり，教師あり学習によりポテンシャルへ到達させることができるとわかる．

<<計算技術と機械学習の新時代>>

続いて、詳細釣り合いの破れた(非平衡系)での確率過程を用いて、マルコフ連鎖モンテカルロ法(MCMC法)、そしてコンストラスティブダイバージェンス法(CD法)について加速する(学習時間を短くする)研究について紹介された。深層学習で用いられる学習法にCD法という手法があり、これはデータを利用して尤度を近似した状態から、MCMC法を適用した手法であることから、MCMC法を加速することで、CD法も加速できる。つまり、統計力学によって機械学習のアルゴリズムが改良され、進化できるということが、機械学習の新時代であると大関氏は話している。

<<質疑応答>>

Q1: 近年、深層学習でCNNというネットワーク構造で、FineTuningだけで学習する方法があるが、統計力学的な観点で何か言えるか

A1: 今回は全結合型ネットワークで解析をしている。CNNは解析未実施のため何も言えない。PreTrainingの代わりに何が性能を良くしているのかには興味がある。

Q2: 教師あり学習と教師なし学習の順番は関係あるか

A2: 逆の順番にしても、教師あり学習がしっかりとできていれば問題はない。例えば局所安定会にはまる等の問題がある。

<<主な参考文献・参考URL(レポーターによる追記)>>

今日からできるスパースモデリング: 京都大学大学院情報学研究科システム科学専攻 大関 真之
<http://www-adsys.sys.i.kyoto-u.ac.jp/mohzeki/Presentation/lectureslide20150904-2.pdf>

Geoffrey E. Hinton (2012) “**A Practical Guide to Training Restricted Boltzmann Machines**” Volume 7700 of the series [Lecture Notes in Computer Science](#) pp 599-619
<https://www.cs.toronto.edu/~hinton/absps/guideTR.pdf>

3. Deep Neural Networks の力学的解析」(東京大学総合文化 研究科 本武陽一 氏)

<<講演内容>>

近年、機械学習分野において、「現実世界の高次元データは、本質的に低次元である」という多様体仮説を元にした学習方法が見られる。深層学習においても、多様体仮説に基づく学習方法がそうでない学習方法と比較して性能が良いことが示唆されている。本講演では、Deep Neural Network(以下 DNN)の学習過程において、層間での情報伝達を DNN のダイナミクスとして捉え、各層の出力のヤコビアンを解析を通じてダイナミクスを解析。ダイナミクスの解析を通じて、複雑な DNN においても、多様体が存在することについて講演された。

<<深層学習のダイナミクスの定義>>

深層学習において入力されたデータは、多段に組まれた層を登っていきながら逐次情報変換を受けることになる。そのような描像に立った時に、その情報伝達の素性を明らかにすることが研究の目的である。入力されるデータ(今回は画像)の 1 ピクセル 1 次元とし、DNN の層間での情報伝達を時間発展とする。各層はユニットの値によってデータが表現されるわけであるから、ユニットが張る特徴空間の遷移過程をモニタすればその具合がわかるだろう、というのが基本アイデアである。具体的には、隣接した 2 つの層のユニット値から **Jacobian** という行列を計算し、**Jacobian** の 1 より大きい特異値に対応する特異ベクトルの方向の情報は拡大され、1 より小さい特異値に対応する特異ベクトルの方向の情報は縮小されたものと定義している。

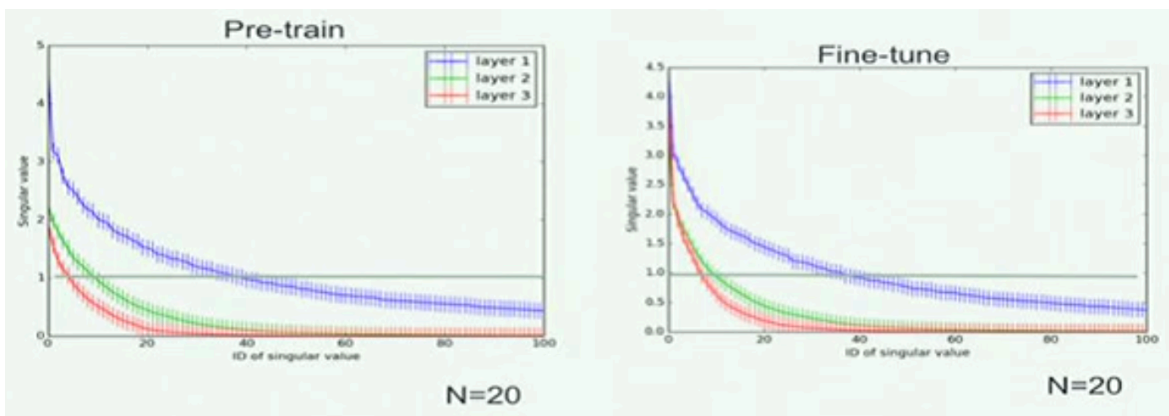
<<情報量の多様体仮説>>

Bengio らによれば、カテゴリの情報は、低次元の多様体に表現されているという(多様体仮説)。この仮説によれば、データは特徴空間上に一様に散らばっているのではなくて、カテゴリ(直線・曲線、イヌ・ネコの類)ごとに別の部分多様体上に局在している。これにより「空間の伸縮 = 情報の取捨選択 = 多様体の構造(拡大さ

れる特異ベクトル：多様体の接線方向，縮小される特異ベクトル：多様体の法線方向)」が主張でき，空間伸縮と抽出された情報（つまり学習結果）を対比することで，仮説の evidence を得ることができる可能性がある。

<<実験：設定と結果>>

今回は，Deep Belief Network（データセット：MNIST、入力：28×28 ピクセル、隠れ層：3層×1,000 ユニット）およびさらに巨大な AlexNet（データセット：ImageNet、入力：224×224×3、隠れ層：7層）という“実践的な系”において仮説の評価を行った．結果は非常に明快で，pre-training においては下流になるほど Jacobian の特異値が 1 を超える次元が非常に少数になることが観測され，ある狭い部分空間が抽出されていることが強く示唆される．またその部分空間からサンプリングすると元々の画像 feature とよく対応しているように見える．（図左部分）さらに興味深いことは，fine-tuning を行った後では Jacobian の特異値が 1 を超える次元が pre-training に比べ増加し，（図 1 右部分）feature がむしろぼやけて見えることである．多様体仮説における多様体以外の情報が追加されたためだろうとのことだが，ボトムアップによるパターン抽出をトップダウンで一部修正する，というようなメカニズムが垣間見えて深層学習の挙動理解という点では示唆的だった．



図， pre-training 時と fine-tuning 時の Jacobian の特異値

さらに多様体仮説を裏付けるために，入力に摂動（微小変化）を加えた時の出力の影響を見る実験を行った．抽出される次元・捨棄される次元はわかっているので，抽出される次元方向に摂動を入れると出力も攪乱され，逆に捨棄される次元に摂動を入れてもロバストになれば多様体仮説が支持されることになる．果たして結果は

そうであった。決定的結果と言えるまでにはまだ研究を深める必要があるようだが、ブラックボックスの中身が透けて見えてきたことで、最終結果を見ながら職人技でチューニングをするのが常であったハイパーパラメタ決定に Jacobian 特異値情報が使えるかもしれないということだ。

<<質疑応答>>

Q1 : DBN が FineTuning をすることで、表象は捉えられなくなっているにもかかわらず、なぜ精度は上がるのか?

A1 : 教師なし学習は多様体構造を捉まえる、教師あり学習は識別面を捉まえる(精度を上げる)、これらの二つの異なる情報が合わさったものであるためだと思われる。特異値/特異ベクトルを見ることは、精度の観点とは別として、多様体構造をとらまえているかの確認をする上で有用である。

Q2: ヤコビアンはどの単位で計算をしているのか

A2: インプットの画像毎にヤコビアンを計算している。特異値分布においては、各画像毎のヤコビアンおよびその特異値/特異ベクトルを計算して、特異値 ID 毎の特異値の平均値と分散をプロットしている。どのような画像が与えられたとしても情報が拡大される次元数にばらつきが小さい、つまり与えられたデータを表現する多様体は一定の次元数で表現可能であることが示唆される。

<<主な参考文献・参考 URL (レポーターによる追記)>>

Salah Rifai, Yann N. Dauphin, Pascal Vincent, Yoshua Bengio, Xavier Muller (2011) “The Manifold Tangent Classifier” Advances in Neural Information Processing Systems

<http://papers.nips.cc/paper/4409-the-manifold-tangent-classifier.pdf>

本武 陽一, 池上 高志(2015)“Deep Neural Networks の力学的解析” The 29th Annual Conference of the Japanese Society for Artificial Intelligence,

http://sacral.c.u-tokyo.ac.jp/pdf/mototake_jsai_2015.pdf

Yhoichi Mototake, Takashi Ikegami (2015) “**The dynamics of deep neural networks**,” Twentieth International Symposium on Artificial Life and Robotics

http://sacral.c.u-tokyo.ac.jp/pdf/mototake_AROB_2015.pdf

4. RECRUIT Challenge のご案内（株式会社リクルートホールディングス 加藤真吾氏）

今回の勉強会の会場を提供していただいたリクルートホールディングスの加藤氏より、RECRUIT Challenge に関する案内があった。同社は 2015/04/15 に 5 人の AI の世界的権威をアドバイザーに招聘して AI 研究を開始している。

<<Kaggle の紹介>>

データサイエンティストコミュニティプラットフォーム。このプラットフォームを通じて、データ分析に関するコンペティションが開催されており、世界中のデータサイエンティストと競える場を提供している。また、単なるコンペのプラットフォームにとどまらず、教育の場としても活用されている。

<<RECRUIT Challenge の紹介>>

ポンパレで販売したクーポンの購買予測のコンペを、7/16～9/30 の期間で開催している。成績優秀者には賞金やインタビューなど様々な特典があるようだ。

<<参考 URL (レポーターによる追記)>>

kaggle <https://www.kaggle.com/>

株式会社リクルートホールディングスのプレスリリース

http://www.recruit.jp/news_data/release/2015/0716_15946.html

RECRUIT CHALLENGE <http://challenge.recruit.ai/>

ポンパレ <http://ponpare.jp/>

5. Skymind の Deep Learning への取り組み (Skymind 社 CTO

Adam Gibson 氏)

<<講演内容>>

Gibson 氏による、世界初の商用レベルでサポートされたオープンソースディープラーニングフレームワークである Deeplearnig4J や、科学計算用の ND4J など、Java 仮想マシン向けのオープンソースライブラリについて解説された。

<<deep learningの現状>>

日本でも多くの方が Deep Learning が多く認知され、使われ始めているように、世界でも Deep Learning を使う試みは増えてきている。シリコンバレーを中心にアメリカでは、世界に先んじて様々な取り組みが進んでいるようだ。しかしながら、Deep Learning が一部の資本力の大きな企業に独占されている現状では、Deep Learning を使いたい企業が出てきても、その要望を叶えるのは必ずしも容易では無いといった現状がある。実際日本国内では、Deep Learning を導入している実用レベルの技術は一部の企業、大学や研究者にとどまっていることから、Deep Learning の技術を広く利用する環境は今後さらに整えていく状況にあると言える。

<<Skymind社のDeep Learningへの試み>>

そうした現状を受けて、Skymind社では、今日にも扱われている Deep Learning の主要なアルゴリズムの実装を可能とする。DeepLearning4J (以下 DL4J) を開発した。DL4J は Java, Scala で書かれたオープンソースの分散 Deep Learning ライブラリで、研究や調査目的だけでなく、ビジネスにも活かすことを目指して開発されている。各種アルゴリズムをミックスすることが可能で、Hadoop や Spark との親和も図れる。Deep Learning ではしばしば、使いたいデータをそれぞれのライブラリにあったデータ型に変換することが困難ではあるが、DL4J では入力されるデータは CANOVA というライブラリによってベクトル化され、データ型の変換を容易にする。また、行列演算は ND4J という線形代数ライブラリで行う。そして、モデルを一度コーディングすると最適化を続ける。

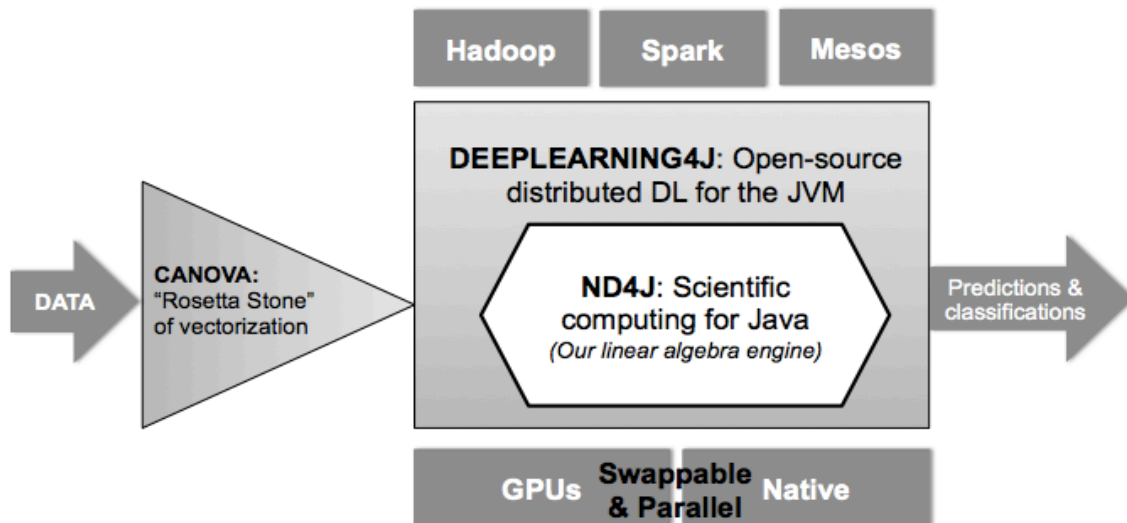


図1, DL4Jの概略図, 入力されたデータがCanovaによってベクトルに変換され, DL4J内部で Prediction問題(予測問題)やClassification問題(分類問題)に応じて処理される.

<<なぜJavaで書かれたのか>>

Javaという言語は世界で最も普及しているAndroidのシステムを書いている言語でもあり, 企業でも, 日常生活のハードにも重要な位置を占めている. そのため, Javaで書かれたDL4Jは, HadoopおよびSparkと統合されており, DL4Jのフレームワークをプラットフォームに用いることで, 学習済みのDeep Learningのモデルは様々なハードウェアの上に載せる, 例えばモバイル端末にも使用することができる. 問題となるのは学習だが, DL4Jでは並列分散的に処理を行うことができ, また, クラスタで学習を行えば大量のデータを高速に学習することができる.

<<山川氏との対談セッション>>

講演後には, ドワンゴ人工知能研究所 山川氏との対談セッションも組まれた. 山川氏は全脳アーキテクチャの目指す汎用人工知能のアプローチには, 複数の機械学習機を組み合わせた複合学習機の実現が必要であるとする一方で, 実現することは簡単ではないと主張する. 一方, Gibson氏は, 実際にDeep Learningでは画像の畳み込みニューラルネットワークと再帰型ニューラルネットワークを合わせた学習機

を実現しているが、これは個々のネットワークごとに別々に学習させることによって可能であるということを主張した。双方のアプローチはDeep Learningをどのように見ているかが異なっているようであった。

<<主な参考文献・参考URL(レポーターによる追記)>>

DL4J Documents <http://deeplearning4j.org/>

Skymind社 Homepage <http://www.skymind.io/>

6. レポーター後記

冒頭のあいさつでも触れられているが、今回の勉強会は募集開始数時間でキャンセル待ちが発生、最終的に約500名が応募（参加は200名）という状況となったことに驚くばかりである。深層学習というと「その仕方はよくわからないが、なぜかうまくいく」という理解が少なくとも非専門家の間では相場だったと思われるが、大関氏のように協働現象と喝破して秩序パラメタで特徴づけてしまう統計力学的手法は、統計力学と機械学習の協奏を展開されており、分野を越えたアプローチによって深層学習に示唆を与えていた。また、本武氏のように空間分離過程を丹念にトレースしていく方法は、深層学習における層から層への情報を伝播する過程を捉える上で興味深いアプローチであった。今回の講演から、「その仕方」がおぼろげながら姿を現してきたという印象だ。これらは、前回（第10回）の勉強会で、東京大学の松尾豊氏が理論研究も盛んになってきているとおっしゃっていたことを実感する機会になった。加えてSkymind社のフレームワークのようなビジネスに向けた深層学習アプリケーションが登場してきたことで、新しい価値を出すということが非専門家レベルでも実践できるフェーズに入ってきたと感じた。そして、全脳アーキテクチャが目指すアプローチから見たDeep Learningは、海外で考えられているDeep Learningを使っていく目的とは、異なっているように見えた。これはチャンスである。汎用人工知能を実現するアプローチとして、全脳アーキテクチャのチャレンジは加速していくべきであり、今後の研究開発にも期待がかかる。

以上