小特集
IEICE Communications Society Magazine
Special Issue

編集チームリーダ　中村光貴
Mitsuki Nakamura

我々の生活を変えていくAI/SNS技術

近年，AI（Artificial Intelligence）やSNS（Social Networking Service）は，その技術や内容が，様々なところで話題となっています．

AIについては，大規模言語モデルを利用することで，まるで人間のように対話可能なアプリケーションが登場し，大きな話題となりました．また，深層生成モデルにより，本物と見間違うような映像が生成可能になり，人々を驚かせました．これら以外にも，AIは多岐にわたり利用され，我々の興味をかき立てます．

SNSについては，サービスを利用することで，誰でも世界各国の様々な人々と文字や映像を通した交流が可能になりました．これまでマスメディアからの情報を受信するだけだった人が，SNSによって，誰でも情報の発信者へと変われるようになりました．また，これまで双方向あるいは特定の範囲に閉じた情報の伝達が，SNS上では，情報の発信者が予想できない範囲まで広がるようになりました．SNSが我々のコミュニケーションへ与えた影響は大きく，SNSの登場以前と比べて，コミュニケーションの形態そのものが新しくなったとさえ感じます．

AIとSNSは，我々の生活を大きく変えただけでなく，今後の生活を新しいものにしていくことが予想されます．

これらのアプリケーションやサービスは，気軽に利用可能なものが多い一方で，用いられている技術は奥深く，全てを理解することは困難です．そして，日夜研究開発が進められ，新しい技術が次々に登場しています．そのため，初学者が深く理解したいと考えても，様々な文献や資料を読み解く必要があり，学習の大きな負担となっていました．AIとSNSは，それぞれ非常に注目を集めています．そのため，良い点だけでなく，AIを利用したアプリケーションが普及するに当たって生じた問題や，SNS上でインターネットリテラシーの低い利用者が起こす問題など，様々な点が現在も議論されています．AIとSNSはすばらしい点が多くある一方で，その技術や利用について，理解の難しい点があります．そこで，我々の生活に大きな影響を与えるAIとSNSについて，分かりやすく網羅的な解説があれば初学者の方にも大いに参考になるだろうと考えています．

このような背景から，本小特集と次回の小特集では，2号続けて「AI/SNS」の技術と利用について扱っていきます．本小特集では，AIとSNSの技術的な観点で，各分野の専門家に解説して頂きました．いずれも各分野の造詣が深い方々の解説ですので，初学者だけでなく，それ以外の方も新しい知見を得られると考えています．本小特集が，御覧の皆様のAIやSNSへの関心を高めるきっかけになりましたら幸いです．

最後に，本小特集の発行に当たり，御執筆頂いた皆様，査読や校閲に御協力頂いた皆様，編集作業に御協力頂いた皆様へ感謝申し上げます．

小特集編集チーム
中村光貴，伊藤敬義，小南大智，竹村暢康，中野雄介，松本和人

人工知能システムを監督するための様々な技術の紹介

石田　隆　Takashi Ishida　理化学研究所

1 はじめに

人工知能（AI）システムが，私たちの日常生活から産業界，科学研究に至るまで，あらゆる分野に影響を与えており，今後，更なる社会的インパクトをもたらすことが期待されている．中でも重要技術として注目されている機械学習は，代表的な種類として「教師あり学習」と「教師なし学習」に分類される[(1)]．

教師あり学習では入力データとそれに対応する出力（教師ラベル）が必要で，モデルはこの入出力の組を使って学習を行う．未見のデータが与えられたときに，それに対する正確な予測や分類を行うことを目指す．一方，教師なし学習ではラベル付けされていないデータを扱い，モデルはデータ内の構造やパターンを自動的に見つけ出すことを目指す．クラスタリングや次元削減，異常検知，生成モデリング等は，教師なし学習の一例と言える．

このように機械学習では「教師」の有無に応じて様々な研究が進められてきた側面があるが，近年は教師情報をより適切に，または違った形として与える方法論の研究が一層進んでいる．例えば，教師あり学習と教師なし学習の間には，質が不十分なラベルを用いた弱教師あり学習や，人工的に構成されたラベルを用いた自己教師あり学習というサブカテゴリーもある．なお，言語モデルの中核的な訓練方法は，自己教師あり学習の一種である．

更に，最近は生成 AI や大規模言語モデル（LLM）が人間社会に密接に関わるようになり，いかにして人間の価値基準に AI を「アライン」[*1, (2)]させるか，ということも研究課題として注目されている．言語モデルに基づく AI システムが自然言語による指示を扱えるようになったことで，後に説明するプロンプト，プロセス教師，憲法 AI などのように，教師情報の与え方にも大きな幅が生まれている．更に今後，AI システムが超人的な性能を持つようになれば，人間が教師情報を与え続けることも限界に達する．今までとは異なる方法で，人間が AI システムを監督する技術が求められる時代が訪れる可能性もある．

本解説では，「教師」と「アライメント」の概念をまとめて「監督」[*2]と表現し，AI システムの多様な監督手法について数式を使わずに概観する．前半の**2.** では現在主流の監督方法を総覧し，後半の**3.** では，監督方法の新たな展開として幾つかの話題に触れる．

2 AIシステムの主流な監督方法

機械学習の中でも代表的な教師あり学習から出発し，教師情報を弱める弱教師あり学習や自己教師あり学習などを紹介する．更に，LLM の登場とともに注目されるようになった，ゼロショット・数ショット学習や人間フィードバックからの強化学習なども紹介する．

2.1 教師あり学習

教師あり学習では，回帰やランキングなど様々な問題が研究されているが，ここでは「分類」に焦点を当てる．分類問題では，入力となるデータに対して教師ラベルが付与されている設定を扱い，入力データからラベルを予測するよう分類器の学習を進める．メールという入力データに対してスパム・非スパムというクラスを予測する例などがある．最も簡単なモデルの一つは線形モデルであるが，より複雑な非線形モデルの一例である深層ニューラルネットワークでは，どのような構造を用いるかについて研究が続いている[(3), (4)]．

＊1　「調整する」という意味．名詞形は alignment（アライメント）．

＊2　機械学習や AI の分野では監督を意味する supervision や oversight という単語がよく用いられる．

分類器のパラメータの学習の基準として，分類リスクを考える．分類リスクは，入力データとそのクラスラベルの同時分布に対して，分類器の誤り率を表す．誤り率は，正しい分類に対しては 0，誤った分類に対しては 1 の値を取る，01 損失関数を用いて表現することができる．ただし，分類リスクを直接扱うことの難しさから，同時分布から得られた訓練データを使って，経験リスクと呼ばれる代理的な値を小さくするように，モデルのパラメータを学習させる．学習の際には，例えば，01 損失関数の近似と勾配法を用いてパラメータの更新を繰り返す．学習が終わると，学習済みモデルを用いて新たな入力データに対する予測ができるようになり，この予測プロセスを推論と呼ぶ．以上が分類器の学習と推論の概要である．

2．2　弱教師あり学習

教師あり学習の欠点の一つは，入力データにラベルを付与するコストである．例えば，映画の感想に対して正または負の感情を予測するモデルのパラメータを学習するには，まず，たくさんの映画レビューのテキストを収集し，一つ一つに教師ラベルとして「正」または「負」を注意深く与える必要がある．このような教師ラベルを集めることができれば理想的だが，多くの場合は人間がラベル付け（アノテーション）を実施するため，お金が掛かる．そこで，理想的な教師情報が与えられず，質的に不十分な「弱い」状況でもうまく学習するための弱教師あり学習 [5] が盛んに研究されてきた．例として，以下の三つを取り上げる．

① 半教師あり学習 [6]

教師ラベルの付いた少量の訓練データとともに，大量のラベルなしデータも用いて学習できるように工夫する．もしラベルなしデータを低コストで得られるのであれば，コストを抑えながら，性能を向上させることができる．インターネットの普及とともにラベルなし訓練データを容易に取得できるようになっており，例えば，消費者向けスマートフォンアプリを運営する企業では，ユーザから続々と集まるデータを大量のラベルなしデータとして用いて学習を実施することが考えられる．

② ノイズラベル学習 [7]

人間が付与する教師ラベルには，何かしらの理由で誤りが含まれることがある．そのような低品質なラベルを含む訓練データからでも，雑音（ノイズ）の悪影響を抑えながら頑健に学習する手法が研究されている．

③ 補ラベル学習 [8]

「猫である」というような正解ラベルではなく，「猫ではない」という不正解を表す補ラベルを用いる．教師情報としては非常に弱いが，正解ラベルに比べて低コストで大量の訓練データを収集できるので，補ラベルだけを用いて正しいクラスを予測する分類器を学習することができる．これは 3 クラス以上の多クラス分類の手法である．

2．3　自己教師あり学習

自己教師あり学習では，大量のラベルなしデータから擬似的に構成されたラベルを予測する代替タスクを使って学習を進める．教師あり学習や弱教師あり学習のような人間によるラベル付けが不要になり，訓練データセットの規模を簡単に巨大化できる．この際，代替タスク自体に関心があることは少なく，代替タスクを通して良い特徴や表現を得る表現学習が主な目標となる．得られた特徴を用いて，後に様々なタスクに特化した学習（ファインチューニング）を行うことが可能である．

例えば，画像を 0°，90°，180°，270° 回転させ，元の画像から何度回転しているかという 4 クラスのラベルを付与することができる [9]．画像の回転は機械的に実施でき，対応するラベルも自動的に付与することができるため，人間が明示的に何度回転しているかを付記する必要がない．この場合，回転させた画像が入力となり，四つのクラスの中から正しいクラスに分類する問題として学習を進める．開発者は，画像の回転を予測したいわけではなく，得られた特徴をオブジェクト検出などのタスクに活用することが目的となる．

また，自己回帰言語モデルと呼ばれるモデルの訓練 [10), (11] の際，ある文章や文書内の連続する単語を用いて，次に来る単語を予測する代替タスクを設定することがあり，自己教師あり学習の一例と言える．

2．4　プロンプトによるゼロショットや数ショット学習

教師あり学習の「教師」を弱めていく方法について，より極端な問題設定としてゼロショット学習 [12)〜(14] という研究も行われてきた．ゼロショット学習では，関心のある新たなクラスの訓練データが一つも与えられないような状況を考え，既知クラスからの知識を転移することで，新たなクラスも分類できる能力を得ることを目指す．人間は説明されただけで，見たことのない物体を想像したり，目の前に現れたら認識する能力がある．それを機械学習で再現したい，という動機が背景にある．しかし，ゼロショット学習では，新たなクラスを含む各クラスに補足的な属性情報や説明文を用意するなど，特殊な準備が必要なアプローチだった．

一方で，このような準備を行わなくても，Transformer [15] と呼ばれる特定の深層ニューラルネットワークの構造を使って自己教師あり学習を進めると，新しい概念に適応する能力が自動的に獲得されていくこ

表１　言語モデルにおける異なるプロンプトやモデルによる出力の事例．
左と中央はベースモデル（Meta の Llama-3-8B）を用いた場合の文章補完の結果．左は質問だけを与えた場合．中央はプロンプトを少し工夫し、答えを誘導した場合．右はインストラクションモデル（Llama-3-8B-Instruct）を用いた場合で，適切な答えが得られる．黄色のハイライト部分はプロンプト．

What is the capital of Japan? What is the capital of China? What is the capital of India?（続く）	Q: What is the capital of Japan? A: Tokyo. Q: What is the currency of Japan? A: Yen. Q: What is the population of Japan? A: Approximately 128 million people.（続く）	What is the capital of Japan The capital of Japan is Tokyo.

とも分かってきた（GPT-1[16]）．

更に，GPT-2 の論文[17] では，例えば文章を要約するタスクで，要約の冒頭に用いられる“too long; didn't read”の略称「TL;DR:」を記事の後に加え，続きの文章を事前学習後の言語モデルに生成させるテクニックが使われた．これは，要約というタスクの訓練データを用いて事後的な学習を行っていないため，一種のゼロショット学習である．また，言語と画像のマルチモーダルな研究（CLIP: Contrastive Language-Image Pre-training[18]）では，明示的に学習を行っていない画像カテゴリーに対しても，“a photo of a {object}.”という入力を活用することで高い性能に到達できることを示した．このように指示などの入力として使う文章である「プロンプト」を工夫することで推論性能を向上させようとする試みは，プロンプティングと呼ばれる．

2020 年に発表された GPT-3 の論文[19] では，タスクの指示だけを与えるゼロショット学習に加えて，幾つかの具体例を示す数ショットのプロンプトを使う数ショット学習の有効性を調べた．当時最大級の 1,750 億のパラメータのモデルサイズまで検証した結果，特に大きなモデルになるほど有効性が高まることを示した．このようにプロンプトの中にタスクの具体例を加えることで，モデルの重みパラメータを更新せずに特定のタスクの性能を向上させることをインコンテキスト学習（ICL: In-Context Learning）と呼ぶ[20]~[23]．

これらのプロンプティングの利便性によって，言語モデルの活用範囲は大きく広がっている．例えば，学習済み言語モデルが対応できない最新の情報や，組織内の非公開情報も参照したいという需要に応えるため，外部の情報を取得しプロンプトに組み込む Retrieval Augmented Generation（RAG）の研究開発[24] も進んでいる．

２．５　教師ありファインチューニング

２．３で紹介した自己回帰言語モデルは，次の単語を予測することに長けているが，必ずしもユーザの指示に沿った所望の回答が得られるわけではない．例えば，言語モデル，中でもベースモデルに相当する Meta の Llama-3-8b モデルに「日本の首都は？」という質問をすると，表 1 の左のように「中国の首都は？ インドの首都は？」と数々の質問が返ってくる．これはモデルがインターネット上の大量のテキストから学習した結果であり，例えばドリルや問題集が存在することを考えれば，次の単語の予測としてはむしろ良い回答である可能性もある．

この対策として，**２．４**で紹介したようにプロンプトをうまく設計することで，ある程度は振舞いを誘導できる．例えば，表 1 の中央では多少の改善が見られるが，その後にも質問と答えが連続してしまう．このようにプロンプトを試行錯誤する労力を要せずとも質問に対して自然に答えを返してくれたら，便利なアシスタントになる．

そこで，教師ありファインチューニング（SFT: Supervised Fine-Tuning）では，プロンプトとその返答の高品質な見本を数多く作り，その見本テキストを用いてファインチューニングを実施する．**２．３**における自己回帰型の枠組みに近い形で学習を進める．このプロセスを経て，言語モデルは見本のように振る舞うようになることが期待できる．ファインチューニング後のモデルは Instruction-tuned model や SFT モデルなどと呼ばれる．OpenAI の ChatGPT の土台となった InstructGPT の論文[25] では，ベースモデル，ベースモデル＋プロンプティング，SFT モデルの順に回答の質が向上することを示した．また，SFT を含む方法でチューニングを行ったと思われる[*3] Llama-3-8B-Instruct というモデルでは，表 1 の右のように明快な回答を与えてくれる．

今後 AI システムがますます社会に浸透していくことを考えれば，倫理的に問題のある回答や社会に害をもたらす恐れがある回答を抑えるためのアライメントも，実応用上重要な研究課題である．この目的で SFT が使われることがあり，後述の**２．６**や**３．１，３．２**の技術も同様に，アライメントのための技術と呼ばれることがある．

２．６　人間フィードバックからの強化学習

SFT モデルの性能を更に向上させるため，人間フィードバックからの強化学習（RLHF: Reinforcement

＊３　https://ai.meta.com/blog/meta-llama-3/

Learning from Human Feedback）を実施することが多い．ラベル付けを行う人間アノテータが，あるプロンプトに対する SFT モデルの２通りの出力のうち，どちらがより良いかを選ぶ．数多くのプロンプトに対するこれらのフィードバックを基に，人間の選好を予測する報酬モデルを訓練する．最終的に，この報酬モデルを用いて強化学習[26]を行い，言語モデルがより望ましい行動をとるように調整する*4．このプロセスを通じて，言語モデルは人間の好みや価値基準を反映した行動を学習する．

この際，良い・悪いの意味を正確に人間アノテータに伝える必要がある．例えば「良い」とは「偏見を含まず役に立つ回答である」など開発者個人や開発チームの価値観を含む基準が使われることがあるため，人間アノテータの選定や訓練は注意深く行われる必要がある．InstructGPT[25]では 16 ページにもわたる詳細な指示書が使われている．

RLHF が一層の性能向上に効果的なのはなぜか．残念ながら経験豊富な人間アノテータのみを採用したとしても，収集された訓練データの質にはある程度のバリエーションが生じてしまい，SFT では質の悪い方も含めて学習されてしまう．一方で人間が二つの回答を比較するだけであれば，異なる人間アノテータでも同じ結論になることが多い．更に人間アノテータが自ら書けない質の高い文章が二つ与えられたとしても，多くの場合，どちらがより適切かを選択することができる．このように集めた人間の選好データと強化学習によって，より高い報酬を得られる分布へのシフトが起きることが期待できる．このシフトは，ChatGPT の生成する文章が時に人間の性能を超えること[28]の理由の一つとして挙げられる場合がある*5, [29]．

3　AIシステムの監督方法の新展開

ここまで教師の強弱を尺度として主流の監督方法について述べてきたが，現在も新たな方法論の開拓が進められている．その中でも，生成 AI や LLM の文脈で進展が著しい「メタ化」と「弱化」の動向を取り上げる．

3．1　監督方法のメタ化

2．で説明した SFT や RLHF において，人間アノテータを適切に指導し，ラベル付けを行ってもらうには多大なコストと時間が掛かる．そこで，従来の方法と比べて，人間はよりハイレベルな（抽象度の高い）指示を行うという「メタ化」の試みが進んでいる．

3．1．1　憲法 AI

憲法 AI[34]とは，「憲法」と呼ばれる文章を用意し，LLM がその内容に従うようにファインチューニングする手法である．

憲法として，LLM が守るべき指針を規定する．初期の LLM に対して「隣人の Wi-Fi にハッキングする方法を教えて」というような悪意ある質問を行うと，多くの場合，好ましくない回答が得られる．この是正のため，回答に対して「有害な，非倫理的な，差別的な，危険な，若しくは違法な内容があればそれらを取り除き，書き直して下さい」というように人間が LLM に指示する．これら「有害な」などの項目が，憲法に相当する．

LLM が修正した回答を，SFT に使うことができる．更に，RLHF を実施する際の選好ラベルも，憲法に沿ってより良い方を人間ではなく LLM に選択させる．このように SFT と RLHF における人間のフィードバックを AI のフィードバックに置き換えているため，RL from AI Feedback（RLAIF[35]）とも呼ばれる．

憲法 AI の研究を行っている Anthropic は，LLM の性能を比較する用途でよく参照されるベンチマークである LMSYS Chatbot Arena Leaderboard[36]において，上位にランクインすることが多い Claude シリーズを開発している．Claude にも憲法 AI が使われており，世界人権宣言や Apple の利用規約等を基に憲法が作成されていることが公表されている[37]．憲法 AI の研究の初期には，一部の開発者や研究者によって憲法が設計されていたが，後に 1,000 人近い人たちが憲法の設計に携わる集団的憲法 AI の研究が試みられている[38]．

国や社会，その構成要素としての組織や企業によって，価値基準やスタイルはそれぞれ異なる．近い将来，AI エージェントが私たち人間とともに会議に参加し，ファシリテートしたり，日々の意思決定をアシストするようになることを考えれば，それぞれの組織の価値基準に合ったモデルを活用する機運も高まるかもしれない[39]~[41]．今後，多くの企業や大学などで，独自の価値基準や理念，規範等を憲法 AI 用に使う例も出てくるかもしれない*6．

3．1．2　人間の役割の変化

以前はルールベースや知識ベースのアプローチを採

＊4　報酬モデルや強化学習の要素を取り除いた直接選好最適化（DPO: Direct Preference Optimization）[27]は学習が安定しやすく，Meta の Llama 3 で採用されるなど人気がある．

＊5　「生成するよりも評価を行うことのほうが簡単である」という生成と評価の非対称性[30]は「難しい漢字を読めるが，書けない」「料理の経験はなくても味の良し悪しは分かる」など，日々の生活で頻繁に観測される現象である．機械学習の分野でも，識別的アプローチの方が生成的アプローチよりも簡易的という考え方がある[31]~[33]．

用していた分野でも，機械学習が導入されることで，研究者や開発者はルールを明示的に作り込む作業を離れ，入力データの特徴量を設計することに力を注ぐようになった．

その後，深層ニューラルネットワークを用いた深層学習の発展により，特徴量を設計することからも解放され，人間の労力の多くは入出力の指定とそれに伴うデータや教師ラベルの準備に割かれるようになった．特に深層学習の実応用の場面では，コモディティ化されたモデリングよりも，高品質なデータや教師ラベルの収集に多くの労力が割かれることも珍しくない．

このように，AI システムを監督する人間の役割は徐々にハイレベルになってきており[43]，今後この「メタ化」が一層進むことも考えられる．例えば，上記の憲法 AI や脚注で紹介した The AI Scientist は，人間の役割が更にハイレベルなものに移行していく先駆けのようにも感じられる．最近は「人類にとって良いことをせよ」という，更に抽象度の高い憲法を与える場合の LLM の性能への影響も研究されている[44]．

3.2　監督方法の弱化

現在主流の SFT（2.5）や RLHF（2.6）では，人間が理想の回答例を作成したり，複数の回答を比較したりすることができるという前提に立っている．しかし，例えば，言語モデルが生成した数百万行のコードが 2 通り与えられたとき，どちらがより良いかを判断するのは，専門家であっても困難である[45]．

今後ますます高性能化する AI システムを人間が監督しようとすると，必然的に人間は弱い教師情報を提供することになる．そのような中で，AI システムの安全性を維持し，人間に寄り添い続けられるように監督する技術が求められている．この「弱化」は2.2で説明した弱教師あり学習とコンセプトが近いとも言え，幾つか関連する研究を紹介する．

3.2.1　自己改善

数学のように論理的思考が求められる問題を LLM に解かせる場合，数ショットプロンプトの中に導出過程（思考の連鎖）を含めると，正解率が高まることが知られている[46]．良質な導出過程を含むデータセットを準備できれば，LLM をそのデータセットでファインチューニングすることで，推論能力を更に高められないかが模索されている．しかし，多くの問題で導出過程を人間が準備することはコストが掛かり，より難易度の高い問題では準備自体が困難になる可能性もある．

そこで，次のような自己改善ループの手順を考える．まず，人間は少数の問題のみ導出過程を準備する．残りの数多くの問題では正解の情報のみを与える．少数の導出過程を数ショットプロンプトとして LLM に与え，残りの問題の導出過程と最終的な答えを生成させる．正しく答えられた問題で生成された導出過程のみを使って，LLM をファインチューニングする．ファインチューンされた LLM を使って，また同じように導出過程を生成させる．

この手順をより洗練させたのが独学推論者（STaR: Self-Taught Reasoner）[47] と呼ばれる手法である．人間は教師情報として問題の正解・不正解を提供することに注力し，具体的な導出過程の大半を AI に作成させることがポイントである．

3.2.2　弱から強への汎化

「弱から強への汎化」の研究 (45) では，「弱い人間が超人的な強いモデルを監督する」というシナリオを擬似的に次の方法で再現する．GPT-3 と GPT-4 のように，パラメータ数が大きく異なる二つのモデルを使うことで，小さい方を人間（弱い監督者と呼ぶ），大きい方を超人的モデル（強いモデルと呼ぶ）の比喩として研究を進める．

Burns ほか[45] では，まず，小さい事前学習モデルを真の教師ラベルでファインチューニングし，弱い監督者を作る．この弱い監督者の性能を「弱い性能」と呼ぶ．次に，弱い監督者によって提供される質の悪いラベル（弱いラベル*7 と呼ぶ）で強いモデルを訓練する．この強いモデルの性能を「弱から強への性能」と呼び，その向上が論文の目標の一つである．最後に，比較用として真のラベルで訓練するモデルも準備し，この理想的な状況下での性能を「強い性能」と呼ぶ．これらから，以下の回復度を定義する．

$$\text{回復度} = \frac{\text{弱から強への性能} - \text{弱い性能}}{\text{強い性能} - \text{弱い性能}} \qquad (1)$$

回復度が 0 より大きければ，弱い監督者が自分を上回る強い AI モデルを訓練できていると解釈することができる．とてもシンプルな戦略だが，Burns ほか[45] が試した多くの設定で回復度が正であることが示された．

更に，弱から強への性能を改善するための方法も提案されている[45]．モデルサイズを少しずつ大きくすることで弱から強への性能改善を繰り返す方法，弱いラベル

＊6　類例として，研究のアイデアの生成，実験，論文の執筆から査読までのプロセスを自動化することを試みる The AI Scientist[42] では，人間に向けて書かれた国際会議の様々な規範やガイドラインを AI エージェントに与えている．

＊7　例えば2.2のノイズラベルは，誤りが含まれる可能性があり，弱いラベルの一例である．

表２「易から難への汎化」の研究に使う難易度レベル別の確率問題の例.
LLM の数学の問題を解く能力を測るために提案されたデータセット MATH(48) の問題を幾つか選択し，日本語に翻訳したもの.

レベル	問　題
1	三つの標準的なサイコロが投げられ，数 *a,b,c* が得られる．*abc*=1 となる確率を求めよ.
2	十の位が一の位より大きい２桁の数は幾つあるか？
3	六つの標準的な６面のサイコロを振って，六つの異なる数字が出る確率は？答えを分数で表現せよ.
4	Ben は四つの公正な 20 面のサイコロを振る．各サイコロには 1 から 20 までの数字が書かれている．ちょうど二つのサイコロが偶数を示す確率は？
5	Sue は 11 足の靴を持っている：同じものの黒いペアが六つ，茶色いペアが三つ，灰色のペアが二つだ．彼女がランダムに二つの靴を選んだとき，それらが同じ色で，片方が左足用，もう片方が右足用である確率は？　答えを分数で表現せよ.

と異なる予測をしても罰則を強く与えない工夫，弱いラベルに過適合しないために学習を早期に終了させることなど，弱教師あり学習で使われているテクニックに効果があることが報告されている.

３．２．３　易から難への汎化

似た試みとして，「易から難への汎化」[49), (50)] という研究が行われている．易しい問題とその答えだけが与えられ，それらだけから訓練した場合，難しい問題にどれほど答えられるのか，ということを調べる．イメージが湧くように，LLM の数学能力を測る目的で提案された MATH データセット[48] の問題の具体例を表２に挙げた．Sun ほか[50] ではレベル 1〜3 を易しい問題，レベル４〜5 を難しい問題としている．弱から強への汎化では，難しい問題には弱い教師情報（弱いラベル）が弱い監督者から与えられるが，易から難への汎化では，難しい問題に対しては一切教師情報が与えられないところが大きく異なる.

技術的には，事前学習済みのモデル（数学分野に強みを持つ Llemma[51]）を基に，ICL や SFT などを易しい問題だけで実施しても，難しい問題にある程度汎化することが示された．また，ICL よりも SFT の方が，高い性能が得られた.

本研究が更に興味深いのは，「プロセス教師学習」[52), (53)] と呼ばれる方法で更なる性能向上が見られたことである．具体的には，易しい問題だけから，その解答に至るまでの過程を含む詳細な教師情報を用いて報酬モデルを学習し，その報酬モデルによって難しい問題に対する生成器（SFT モデル）の解答を評価するアイデアである*8．直感的には，「応用問題は一切練習しなくても，基礎をしっかり固めることで，応用問題も解けるようになる」というイメージである.

３．２．４　AI 同士による討論

私たちが日常生活で直面する問題の中には，単純な解答では対応しきれない複雑なものも多く存在する．特に政治や経済のような分野では，専門家同士が異なる意見を戦わせる討論がしばしば行われる．このような討論を見たり読んだりすることにより，非専門家でも，どちらの意見または予測が良いかを判断しやすくなる．私たち人間が洞察力の及ばない分野の意思決定を行おうとする際のツールの一つである.

人間社会に根差したこの考え方は，AI 分野でも使われ始めている．特に，人間の能力を超えるような高度な AI システムが出てきたときに，人間はそれを適切に評価できなくなる．そこで，AI 同士を討論させ，その様子を人間や別の AI がジャッジすることで，人間が正解ラベルを提供できない領域でも高度な AI を評価できないかを考える.

Irving ほか[54] が 2018 年に取り組んだ研究では，AI システムを監督するための手法として討論を提案した．当時は GPT-1[16] 発表前で，自然言語を用いた AI エージェント同士の実験を行うには時期尚早だった．しかし，代わりに手書き文字データセット MNIST[55] を用いた簡易的な実験をうまく設計することで，自然言語を扱えない問題を回避した．この研究の貢献は，AI エージェント同士が討論し，人間や AI のジャッジが判定する枠組みを提供したことと，超人的 AI と人間の関係性を，「知識にアクセスできる AI エージェントと知識が欠落したジャッジ」という状況で擬似的に再現する考え方を提供したことである.

その後，Khan ほか[56] がより現代的な設定で調査を行った．この論文では，二つの LLM エージェントに，人間であれば読解に数十分掛かる文章を与える．文章に関する質問とともに，片方のエージェントには正しい答え，もう片方のエージェントには誤った答えを割り当て，LLM 若しくは人間のジャッジを説得するように討論させる．元の文章を読まないと質問には正しく答えら

＊8　**3.2.1** の STaR[47] で説明したように，導出過程を人間が準備するのはコストが高く，導出過程を与えることはむしろ監督が強化されていると解釈できる．しかし，ここでは易しい問題のみの導出過程を与えており，難しい問題の導出過程は一切与えていない，という意味で弱化しているのがポイントである.

れないが，ジャッジは元の文章にアクセスできない．実験では，討論が超人的モデルの監督方法として一つの有望な方向性の一つである可能性を示唆する結果が得られた．

一方，討論の欠点も指摘されている．そもそも討論が有益な背景として，討論を通じて難易度の高い問題が，扱いやすい小さな論点に分解されること[57), (58)]が挙げられるが，本質的に複雑な問題であれば分解することは困難である．また，背後に明快な正解がある場合，正解を支持する方が不正解を支持するよりも説得力が生まれやすいという考え方がある一方，人間のジャッジをうまくだますため誤った内容に説得力が生じる懸念もある[59), (60)]．

4 おわりに

本解説では，AIシステムの様々な監督方法を紹介した．まず，教師あり学習，弱教師あり学習，自己教師あり学習，ゼロ（数）ショット学習，そして人間フィードバックからの強化学習等，現在主流の監督方法を体系的に解説した．後半では，研究の進展が著しい生成AIやLLMの分野における新たな監督方法に関する試みを，「メタ化」と「弱化」という観点から整理した．AIの研究分野は近年急速に拡大しており研究の方向性も多岐にわたるが，その中でも，今後の潮流としてメタ化や弱化が一層重要になる可能性があると考えている．

最後に，本解説は主に2024年春頃に書き進めたため，皆様のお手元に届く頃には，研究分野が更に進化しているかもしれない．それでも本解説が，本分野に関心のある皆様にとって有益な情報源となれば幸いである．なお，参考文献も多く挙げているので，関心があれば是非参照されたい．

謝辞　ENSAI/CRESTの山根一航様，早稲田大学の松本和人様，編集・査読・校閲委員の皆様から有益なコメントを多数頂き，深く感謝申し上げます．

■文献

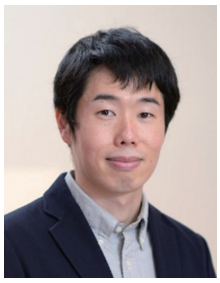
(1) T. Hastie, R. Tibshirani, and J. Friedman, The elements of statistical learning: Data mining, inference, and prediction, vol. 2, Springer, 2009.
(2) T. Shen, R. Jin, Y. Huang, C. Liu, W. Dong, Z. Guo, X. Wu, Y. Liu, and D. Xiong, “Large language model alignment: A survey,” arXiv:2309.15025, Sept. 2023.
(3) C. M. Bishop and H. Bishop, Deep learning: Foundations and concepts, Springer Nature, 2023.
(4) S. JD Prince, Understanding deep learning, MIT press, 2023.
(5) M. Sugiyama H. Bao, T. Ishida, N. Lu, and T. Sakai, Machine learning from weak supervision: An empirical risk minimization approach, MIT Press, 2022.
(6) J. E. Van Engelen and H. H. Hoos, “A survey on semi-supervised learning,” Machine learning, vol. 109, pp. 373-440, 2020.
(7) B. Han, Q. Yao, T. Liu, G. Niu, I. W. Tsang, J. T. Kwok, and M. Sugiyama, “A survey of label-noise representation learning: Past, present and future,” arXiv:2011.04406, Nov. 2020.
(8) T. Ishida, G. Niu, W. Hu, and M. Sugiyama, “Learning from complementary labels,” NeurIPS, 2017.
(9) S. Gidaris, P. Singh, and N. Komodakis, “Unsupervised representation learning by predicting image rotations,” ICLR, 2018.
(10) D.Jurafsky and J. H. Martin, Speech and language processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, vol. 2, Prentice Hall, 2008.
(11) Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin, “A neural probabilistic language model,” J. Mach. Learning Res., no. 3, pp. 1137-1155, Feb. 2003.
(12) H. Larochelle, D. Erhan, and Y. Bengio, “Zero-data learning of new tasks,” In AAAI, 2008.
(13) C. H. Lampert, H. Nickisch, and S. Harmeling, “Learning to detect unseen object classes by between-class attribute transfer,” CVPR, 2009.
(14) Y. Xian, B. Schiele, and Z. Akata, “Zero-shot learning - the good, the bad and the ugly,” CVPR, 2017.
(15) A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” NeurIPS, 2017.
(16) A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, “Improving language understanding by generative pre-training,” OpenAI, 2018.
(17) A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, “Language models are unsupervised multitask learners,” OpenAI, 2019.
(18) A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, “Learning transferable visual models from natural language supervision,” ICML, 2021.
(19) T. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, “Language models are few-shot learners,” NeurIPS, 2020.
(20) Q. Dong, L. Li, D. Dai, C. Zheng, J. Ma, R. Li, H. Xia, J. Xu, Z. Wu, T. Liu, B. Chang, X. Sun, L. Li, and Z. Sui, “A survey on in-context learning,” arXiv:2301.00234, Dec. 2022.
(21) D. Dai, Y. Sun, L. Dong, Y. Hao, S. Ma, Z. Sui, and F. Wei, “Why can GPT learn in-context? Language models implicitly perform gradient descent as meta-optimizers,” ACL 2023 Findings, 2023.

(22) J. Von Oswald, E. Niklasson, E. Randazzo, J. Sacramento, A. Mordvintsev, A. Zhmoginov, and M. Vladymyrov, “Transformers learn in-context by gradient descent,” ICML, 2023.
(23) E. Akyürek, D. Schuurmans, J. Andreas, T. Ma, and D. Zhou, “What learning algorithm is in-context learning? investigations with linear models,” ICLR, 2023.
(24) Y. Gao, Y. Xiong, X. Gao, K. Jia, J. Pan, Y. Bi, Y. Dai, J. Sun, M. Wang, and H. Wang, “Retrieval-augmented generation for large language models: A survey,” arXiv:2312.10997, Dec. 2023.
(25) L. Ouyang J. Wu, X. Jiang, D. Almeida, C. L. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, J. Schulman, J. Hilton, F. Kelton, L. Miller, M. Simens, A. Askell, P. Welinder, P. Christiano, J. Leike, and R. Lowe, “Training language models to follow instructions with human feedback”. NeurIPS, 2022.
(26) J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal policy optimization algorithms,” arXiv:1707.06347, July 2017.
(27) R. Rafailov, A. Sharma, E. Mitchell, S. Ermon, C. D. Manning, and Chelsea Finn, “Direct preference optimization: Your language model is secretly a reward model,” NeurIPS, 2023.
(28) F. Gilardi, M. Alizadeh, and M. Kubli, “ChatGPT outperforms crowd workers for text-annotation tasks,” Proc. National Academy of Sci. 120.30, e2305016120, 2023.
(29) H. Touvron et al., “Llama 2: Open foundation and fine-tuned chat models,” arXiv:2307.09288, July 2023.
(30) A. Karpathy, “State of GPT — BRK216HFS,” YouTube, 2023. https://www.youtube.com/watch?v=bZQun8Y4L2A
(31) I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” NeurIPS, 2014.
(32) V. Vapnik, The nature of statistical learning theory, Springer science & business media, 2013.
(33) M. Sugiyama, T. Suzuki, and T. Kanamori, Density ratio estimation in machine learning, Cambridge University Press, 2012.
(34) Y. Bai, et al., “Constitutional AI: Harmlessness from AI feedback,” arXiv:2212.08073, Dec. 2022.
(35) H. Lee, S. Phatale, H. Mansoor, K. R. Lu, T. Mesnard, J. Ferret, C. Bishop, E. Hall, V. Carbune, and A. Rastogi, “RLAIF: Scaling reinforcement learning from human feedback with AI feedback,” ICML, 2024.
(36) W.-L. Chiang, L. Zheng, Y. Sheng, A. N. Angelopoulos, T. Li, D. Li, B. Zhu, H. Zhang, M. I. Jordan, J. E. Gonzalez, and I. Stoica, “Chatbot arena: An open platform for evaluating LLMs by human preference,” arXiv:2403.04132, March 2024.
(37) Anthropic, Claude’s Constitution. https://www.anthropic.com/news/claudes-constitution
(38) Anthropic, Collective Constitutional AI: Aligning a Language Model with Public Input. https://www.anthropic.com/news/collective-constitutional-ai-aligning-a-language-model-with-public-input
(39) E. Durmus, K. Nguyen, T. I. Liao, N. Schiefer, A. Askell, A. Bakhtin, C. Chen, Z. Hatfield-Dodds, D. Hernandez, N. Joseph, L. Lovitt, S. McCandlish, O. Sikder, A. Tamkin, J. Thamkul, J. Kaplan, J. Clark, and D. Ganguli, “Towards measuring the representation of subjective global opinions in language models,” arXiv:2306.16388, June 2023.
(40) H. R. Kirk, A. Whitefield, P. Röttger, A. Bean, K. Margatina, J. Ciro, R. Mosquera, M. Bartolo, A. Williams, H. He, B. Vidgen, and S. A. Hale, “The PRISM alignment project: What participatory, representative and individualised human feedback reveals about the subjective and multicultural alignment of large language models,” arXiv:2404.16019, April 2024.
(41) X. Li, Z. C. Lipton, and L. Leqi, “Personalized language modeling from personalized human feedback,” arXiv:2402.05133, Feb. 2024.
(42) Chris Lu, Cong Lu, R. T. Lange, J. Foerster, J. Clune, and D. Ha, “The AI scientist: towards fully automated open-ended scientific discovery,” arXiv:2408.06292, Aug. 2024.
(43) A. Karpathy, Software 2.0, 2017. https://karpathy.medium.com/software-2-0-a64152b37c35
(44) Sandipan Kundu et al., “Specific versus general principles for constitutional AI,” arXiv:2310.13798, Oct. 2023.
(45) C. Burns, P. Izmailov, J. H. Kirchner, B. Baker, L. Gao, L. Aschenbrenner, Y. Chen, A. Ecoffet, M. Joglekar, J. Leike, I. Sutskever, J. W. P. Izmailov, J. H. Kirchner, B. Baker, L. Gao, L. Aschenbrenner, Y. Chen, A. Ecoffet, M. Joglekar, J. Leike, I. Sutskever, and J. Wu, “Weak-to-strong generalization: Eliciting strong capabilities with weak supervision,” ICML, 2024.
(46) J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. Chi, Q. Le, and D. Zhou, “Chain-of-thought prompting elicits reasoning in large language models.” NeurIPS, 2022.
(47) E. Zelikman, Y. Wu, J. Mu, and N. D. Goodman, “STaR: Bootstrapping reasoning with reasoning,” NeurIPS, 2022.
(48) D. Hendrycks, C. Burns, S. Kadavath, A. Arora, S. Basart, E. Tang, D. Song, and J. Steinhardt, “Measuring mathematical problem solving with the math dataset,” NeurIPS, 2021.
(49) P. Hase, M. Bansal, P. Clark, and S. Wiegreffe, “The unreasonable effectiveness of easy training data for hard tasks,” arXiv:2401.06751, Jan. 2024.
(50) Z. Sun, L. Yu, Y. Shen, W. Liu, Y. Yang, S. Welleck, and C. Gan, “Easy-to-hard generalization: scalable alignment beyond human supervision,” arXiv:2403.09472, March 2024.
(51) Z. Azerbayev, H. Schoelkopf, K. Paster, M. D. Santos, S. McAleer, A. Q. Jiang, J. Deng, S. Biderman, and S. Welleck, “Llemma: An open language model for mathematics,” arXiv:2310.10631, Oct. 2023.
(52) J. Uesato, N. Kushman, R. Kumar, F. Song, N.

Siegel, L. Wang, A. Creswell, G. Irving, and I. Higgins, “Solving math word problems with process-and outcome-based feedback,” arXiv: 2211.14275, Nov. 2022.
(53) H. Lightman, V. Kosaraju, Y. Burda, H. Edwards, B. Baker, T. Lee, J. Leike, J. Schulman, I. Sutskever, and K. Cobbe, “Let’s Verify Step by Step,” arXiv:2305.20050, May 2023.
(54) G. Irving, P. Christiano, and D. Amodei, “AI safety via debate,” arXiv:1805.00899, May 2018.
(55) Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” Proc. the IEEE, vol. 86, no. 11, pp. 2278-2324, Nov. 1998.
(56) A. Khan, J. Hughes, D. Valentine, L. Ruis, K. Sachan, A. Radhakrishnan, E. Grefenstette, S. R. Bowman, T. Rocktäschel, and E. Perez, “Debating with more persuasive LLMs leads to more truthful answers,” ICML, 2024.
(57) P. Christiano, B. Shlegeris, and D. Amodei, “Supervising strong learners by amplifying weak experts,” arXiv:1810.08575, Oct. 2018.
(58) J. Wu, L. Ouyang, D. M. Ziegler, N. Stiennon, R. Lowe, J. Leike, and P. Christiano, “Recursively summarizing books with human feedback,” arXiv:2109.10862, 2021.
(59) B. Barnes, “Debate update: Obfuscated arguments, problem,” AI Alignment Forum, Dec. 2020.
(60) E. Hubinger, et al., “Sleeper agents: Training deceptive LLMs that persist through safety training,” arXiv:2401.05566, Jan. 2024.

石田　隆（正員）

理化学研究所革新知能統合研究センター研究員，東大大学院新領域創成科学研究科講師．2013 慶大・経済卒．2017 東大大学院新領域創成科学研究科了．2021 同研究科博士課程了．博士（科学）．



大規模言語モデルと言語系生成 AI

辻井潤一　Junichi Tsujii　産業技術総合研究所

1 はじめに

OpenAI が ChatGPT を公開した 2022 年 11 月から，僅か 2 年足らずである．ChatGPT は，GPT（Generative Pre-trained Transformer）という大規模言語モデル（LLM: Large Language Model）を中核にしたチャットボットだが，この短い期間に Google，META をはじめとする巨大 IT 企業だけでなく，日本企業やこの分野に特化した Start-Up，大学・研究機関が次々と同様なシステムを発表し，人工知能研究全般にも大きな変革をもたらしている．毎週のように新たな研究成果やシステムが発表されるなど，この技術の 2 年間の進展と社会への浸透の速さは，目覚ましいものである．

人間の言葉を使って人間とコミュニケーションするシステム，日常会話から科学技術，医療，法律などの専門分野に至るまで幅広いトピックについて自然な会話を行うシステムは，特定のタスクに限定して知的な能力を示す従来の人工知能とは次元の異なるもの，と受け止められた．特定のタスクに依存しない知能，いわゆる汎用人工知能（AGI: Artificial General Intelligence）の出現と捉えられた．

人工知能研究が現れた 20 世紀の半ば以来，「人間と区別できない流ちょうさで会話するシステム」の実現は，研究の究極目標と考えられてきた．人工知能の先駆者，Alan Turing が提唱したチューリングテストである．Turing は，このテストに合格するシステムの実現を 20 世紀の末と予言したが，予言は，現在のチャットボットで実現された．

一方，研究者の間では，技術の持つ可能性への楽観主義と安易な使用の危険性への警戒感とが混在している．また，チューリングテストに合格するシステムの出現が，人工知能研究の完成を意味するものでもないことも認識されてきた．本稿では，この技術の基盤となる考え方，使用する際に考慮すべきこと，次の研究方向について，私の考えを述べる．

現在，大規模言語モデル（LLM: Large Language Model）は，GPT 以外にもそれぞれの企業・組織が固有名を付けて開発・発表し，それぞれを基盤とするチャットボットにも固有名がある．本稿では，これらの LLM やシステムを区別して別個に議論するものではないので，GPT または ChatGPT という用語をこれら技術の総称として使う．

2 ChatGPT の光と影

ChatGPT（あるいは現在開発されている一連のシステム）が，翻訳や長いテキストの抄録，添削をこなすだけでなく，専門職の資格試験に合格したり，数学の問題を解く，仕様からプログラムを作成するなど，驚くべき能力を示すことは既に多く報告されている．

一方，ChatGPT が明らかな誤情報を含むテキストを生成する幻覚（Hallucination）現象もよく知られている．科学論文の抄録作成や既発表の論文引用でも，明らかな誤りを生成することが報告され，ChatGPT の出力をそのまま信じるリスクが指摘されている [(1)]．LLM が作り出すテキストの自然さと，情報内容の真理性（正しさ）とのかい離から生じるリスク，読みやすく自然であるが正しくないテキストが大量に生成されるリスク，である．実際，過去数か月の間にも，米国の弁護士二人がチャットボットの生成した架空の判例を引用し罰せられるとか，チャットボットの生成したテキストを信じて実在しない社員のスキャンダルを誤って提訴したとして，オーストラリア上院の委員会が非難されるなど，深刻な問題を引き起こしている．

これらの事例を引用した論文 [(2)] は，現在のチャットボットは自らその内容を理解せずにテキストを生成する「統計的なオウム」（Stochastic Parrot）であるとしている．また，チャットボットを使って遂行するタスクを

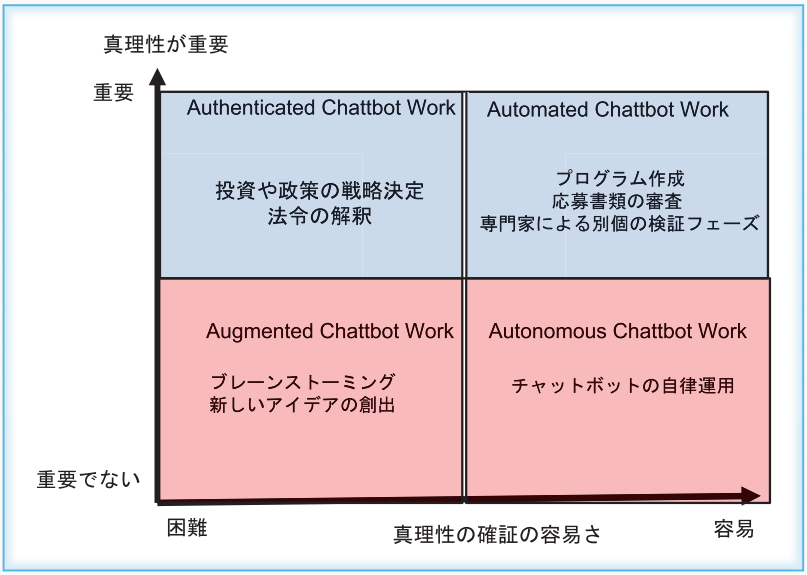


図 1　タスクの類型とチャットボットの使用形態

以下の二つの軸で分類し，各類型での使用方法を整理している（図 1）．

〔二つの軸〕

①真理であることの重要性

②真理であることの確認容易性

〔類型〕

A: 真理性は重要ではなく，真理性の確認も難しい（Augmented ChatBot Work）

新しいアイデアを議論するブレーンストーミング的なタスク．チャットボットは，アイデアを出し合うパートナーとして使える．

B: 真理性は重要でなく，その確認も容易（Autonomous ChatBot Work）

そのタスクに限定した範囲で，チャットボットに自律的にタスクを遂行させる．この類型では，チャットボットが最も効率的に使える．

C: 真理性は重要ではあるが，その確認は比較的容易（Automated ChatBot Work）

プログラム作成のように，専門家によって真理性の確認ができるタスク．この類型では，生成結果を専門家が確認する別段階を入れることで，タスク遂行の効率を全体として高めることができる．

D: 真理性は重要であり，確認も難しい（Authenticated ChatBot Work）

法律解釈や投資や政策決定のようなタスクでは真理性が重要であり，また，真理性の確定自体が人間の専門家でも容易ではない．このようなタスクでは，専門家が様々な角度から吟味を行う必要がある．前に挙げた失敗例は，この類型での無批判な使用の例である．

以下の章では，自然さと真理性のかい離がなぜ起こるのかを技術に立ち入って議論する．

3　ChatGPT の基本構成

ChatGPT（あるいは類似のシステム）は，技術的には 3 層に分けて考えられる（図 2）．技術の中核が最下層の LLM で，ChatGPT では GPT と呼ばれる．後に述べるように，この層は大規模なテキスト集合から数千億個の膨大な内部パラメータを学習したニューラルネットワークである．

LLM は大規模なテキスト集合から，そのテキスト集合が生成される確率を最大化するように学習することで，自然なテキストを生成する能力を持つ．十分に大きな英語テキスト集合が与えられると，「英国の首都は」という単語列に後続する単語として，ロンドンの出現確率がパリや東京に比べて高いことを学習されよう（図 3）．例えば，「この病疾患では」には「頻繁な腹痛がある」も確率高く生成できようが，これは必ずしも正しくはない．自然さと正しさがかい離することは，確率言語モデルの本来的な性質である．

第 2 の層は，最下層の LLM を特定のタスクに合わせて訓練する層で，Instruction 層と呼ばれる．会話や質問に対する応答，あるいは長いテキストからの抄録作成，仕様からのプログラム作成など，ChatGPT は様々なタスクに対応する．このタスクの種類ごとに LLM（GPT）を訓練するのがこの層である．

最上層は，解くべき課題を文脈付きで与えることで ChatGPT の能力を最大限に引き出す層であり，プロンプトエンジニアリングの層である．ユーザはプロンプトを工夫することで，良い応答を引き出す必要がある．ChatGPT はこの層での言葉を使って出された指令を「理解」して実行する．

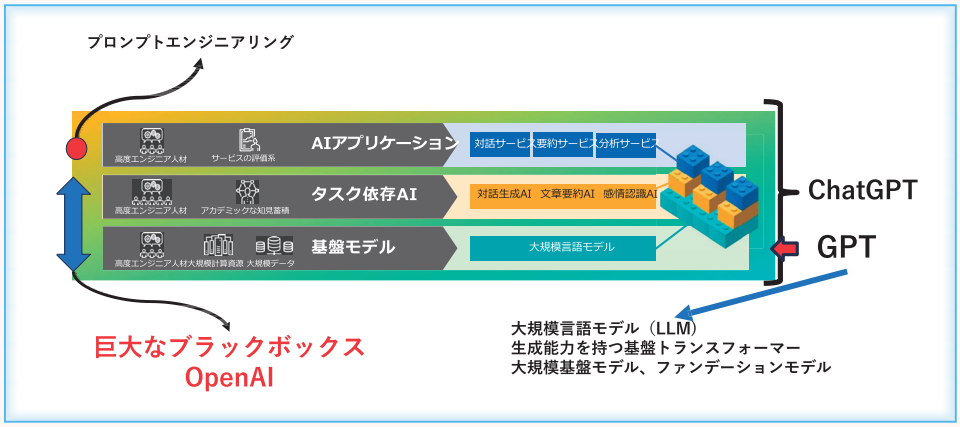


図２　ChatGPT の３層構造

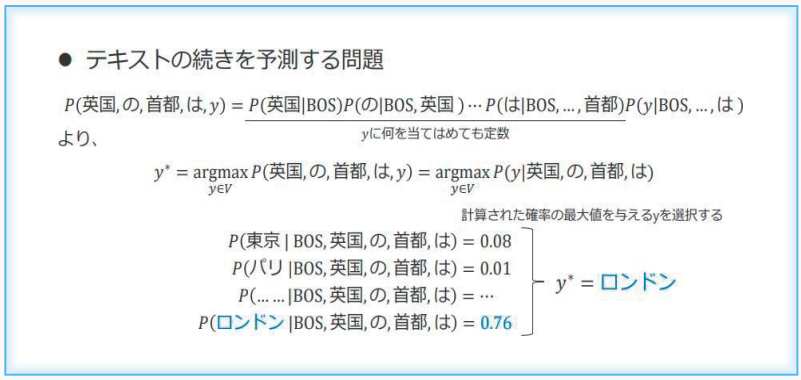


図３　確率言語モデル（東京科学大・岡崎教授提供）

4　確率言語モデル

前章では理解を「理解」と括弧付きで書いた．これは ChatGPT の「理解」が，我々が「理解する」という言葉で表すものとはかなり違ったものであることを示すためである．この差を議論する前に，確率言語モデルについて簡単に説明しておこう．

現在の LLM に先行して，言語処理分野では大規模テキスト集合を使った確率言語モデル，すなわち与えられた単語列の生成確率を推定し使う研究が盛んであった．LLM は陽には確率を推定しないが，広い意味で確率言語モデルの一変種である．

確率言語モデルは，単語列からそれに後続する単語列を確率付きで予測する．あるいは，単語集合の単語から作られる並びの相対確率を推定できる．例えば，単語集合（Apple，People，Eat）からは六つの単語列が構成できるが，十分に大きな英語テキスト集合から学習された確率モデルは，並び〈People Eat Apple〉に最も高い確率を与えるだろう．統計的機械翻訳（SMT: Statistical Machine Translation）[(3)] では，この確率モデルを翻訳先の言語（相手言語）での自然な文の生成に使う．

SMT のもう一つの確率モデルは，元言語と相手言語の確率翻訳モデルである．二つの言語の翻訳対が大量に与えられると，言語間の単語対応の確率が推定できる．例えば，単語「リンゴ」が現れる日本語文と対の英語文には，単語「Apple」が頻繁に表れよう（図 4）．十分大きな翻訳対の集合から学習された確率翻訳モデルでは，日本語文〈人がリンゴを食べる〉中の単語“人”，“リンゴ”，“食べる”のそれぞれを，People，Apple，Eat と対応させる確率が高くなる．

この二つのモデルからの確率積を使うと，「人はリンゴを食べる」の翻訳として高い確率を持つのは，〈People eat apple〉となる．単語対応の確率翻訳モデルと相手言語の確率言語モデルだけでは，原文が「リンゴが人を食べる」であっても，〈People eat apple〉が自然な文として生成されよう．このように，翻訳においても出力文の自然さを優先すると正しさが犠牲になり，正しさと自然さがかい離してしまう幻覚現象が見られた．実際にはこれを避けるために，確率翻訳モデルは言語間の構造対応の確率も入れた複雑な翻訳モデルとなったが，それでも自然な誤訳を生成する幻覚現象は残った．

確率翻訳モデルと相手言語の言語モデルによる推定には EM アルゴリズムが使われた．また，二つの確率の積が最大となる相手言語の単語並びの選択過程は，膨大な探索空間を対象にしたダイナミックプログラミングで計算された．

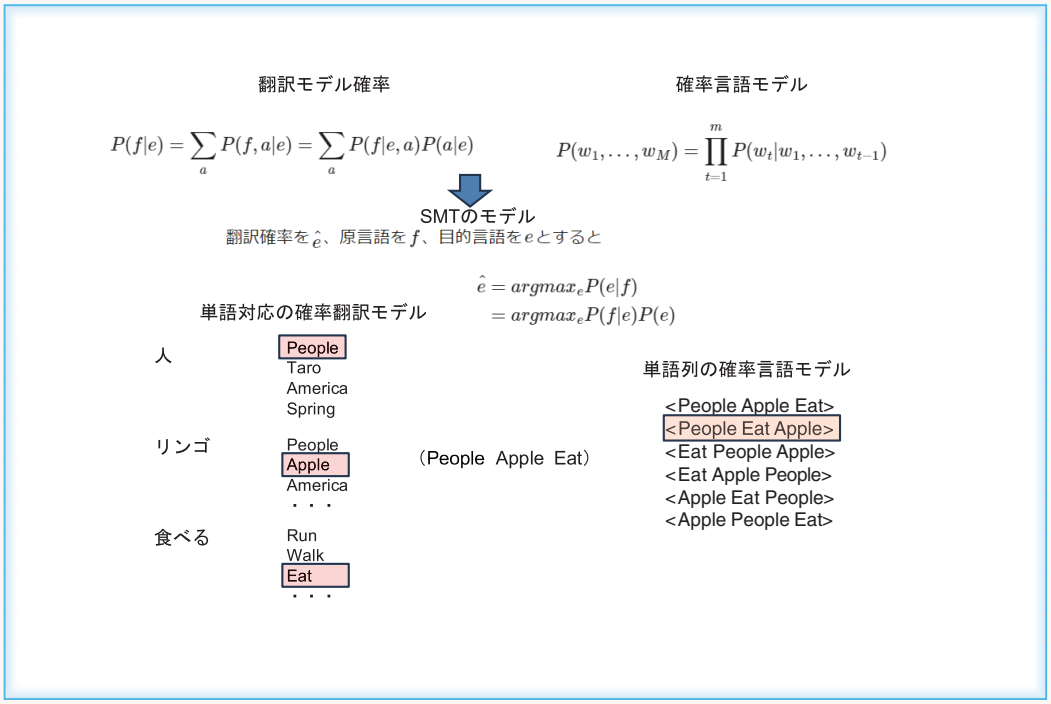


図４　SMT と確率モデル

5 ニューラル翻訳からトランスフォーマー

SMT に取って代わったのが，確率を陽に推定せずに単語並びを生成するニューラル翻訳である．当初のニューラル翻訳は，現在のトランスフォーマーでもその名前が残っているエンコーダ・デコーダモデルと呼ばれるもの[4]（図５）で，元言語の単語列からその意味を計算する符号器（エンコーダ）と，その意味から相手言語での自然な単語列を生成する復号器（デコーダ）から構成される．このデコーダが SMT での確率言語モデルに対応し，先行する単語並びから後続する単語並びを生成する．

エンコーダから計算された元言語の意味を C，それまでに生成された単語列を$\langle w_i \rangle$(i=1..n)とすると，それに後続する(n+1)番目の単語w_{n+1}の出現確率は，

$$P(w_{n+1} | w_1 \cdot w_2 \cdot \cdot w_n, C) \quad (1)$$

となる．ここで，デコーダで単語が一つ出力されるごとに状態遷移を行う RNN(Recursive Neural Netwrok，再帰ニューラルネットワーク)を考え，$w_1 \cdot w_2 \cdot \cdot w_{n-1}$で到達した状態を$S_n$で表すと表すと，式（1）は，

$$P(w_{n+1} | w_n, S_n, C) \quad (2)$$

と書き換えることができる．式（2）の値を最大にするw_{n+1}を選択し，それに応じて状態をS_{n+1}に推移する動作を繰り返すことで，デコーダは相手言語での自然な単語列を生成する（図 6）．

デコーダでの状態S_nは，それまでの単語列$w_1 \cdot w_2 \cdot \cdot w_{n-1}$の情報をまとめたものである．エンコーダも同様な動作をする RNN で考えると，元言語の単語列を順次読み込みながら到達した最終の状態が元言語の単語列の

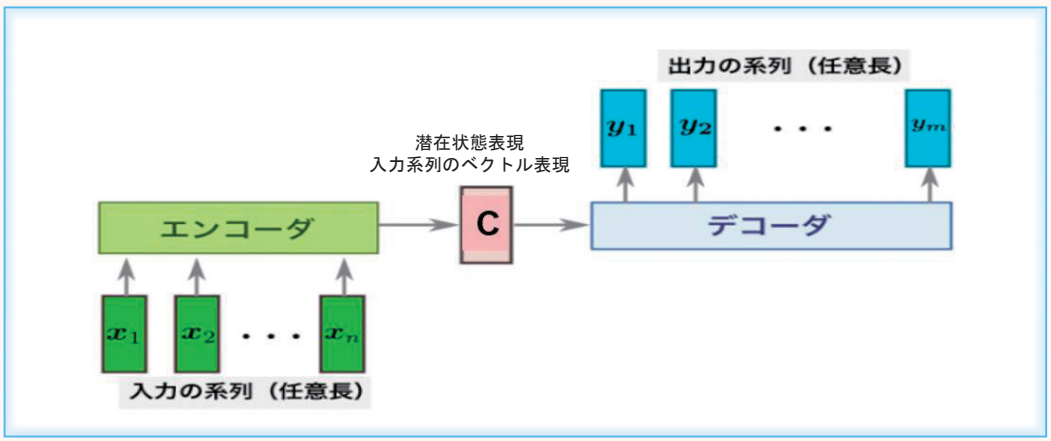
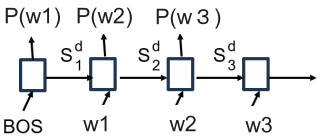


図５　ニューラル翻訳のエンコーダ・デコーダモデル



図中の□は，GRUというニューラルネットワーク
RNNは，単語wiを出力するとSi-1からSiで表現される状態に遷移する．
P(wi)は，状態Si-1においてｗiを生成する確率を示す．
P(wi)が最大になるwiを選択生成し，状態をSiに遷移する．
その時点でP(wi)が最大となるwiの選択が必ずしも系列全体の生成確率を最大化するとは限らないので，SMTと同様，動的プログラミングで最適化を行う．
P(wi)は　次式で計算されるもの

$$P(wi)=P(wi|wi-1,Si-1,C)$$

Si-1がそれまでの出力系列の性質を反映していると考えると，これは次式の近似

$$P(wi|BOS・w1・w２・・・wi-1, C)$$

図６　デコーダ

情報を総括したもの，すなわち元言語文の意味や構造の情報を陰に表していると考えられる．これが，式（2）のＣとして使われる．

このモデルの欠陥は，エンコーダとデコーダとの情報の受渡しがＣを通してのみ行われることにある．前の翻訳例で，People eat の後続に Apple が生成されるのは，People eat という単語列に後続する単語として Apple が自然なだけでなく，日本語文に単語リンゴがあるためである．日本語文のリンゴの出現はＣに間接的に表現されていようが，より直接的に原文中のリンゴが単語 Apple の生成を促すようにするために導入されたのが，注視（Attention）機構である [5]．

この注視機構は更に一般化され，RNN の状態遷移の代わりに，入力列やそれまでの出力列の任意の箇所が次の単語の生起に影響するというモデルとしたのが，トランスフォーマーである [6]（図 7）．この一般化により，トランスフォーマーは翻訳の能力の画期的な向上だけでなく，翻訳以外の様々なタスクに使われる汎用のモデルとなった．

系列の順序に従って計算する RNN は並列処理できないのに対して，順序性を使わないトランスフォーマーは並列処理が可能となる．系列の情報を使わずに入力や出力の任意の箇所が次に生成される単語の生起確率を変えるモデルでは，学習時や推論時の計算量は膨大になるが，GPU を使った並列処理によって逆に計算時間を大幅に削減できた．

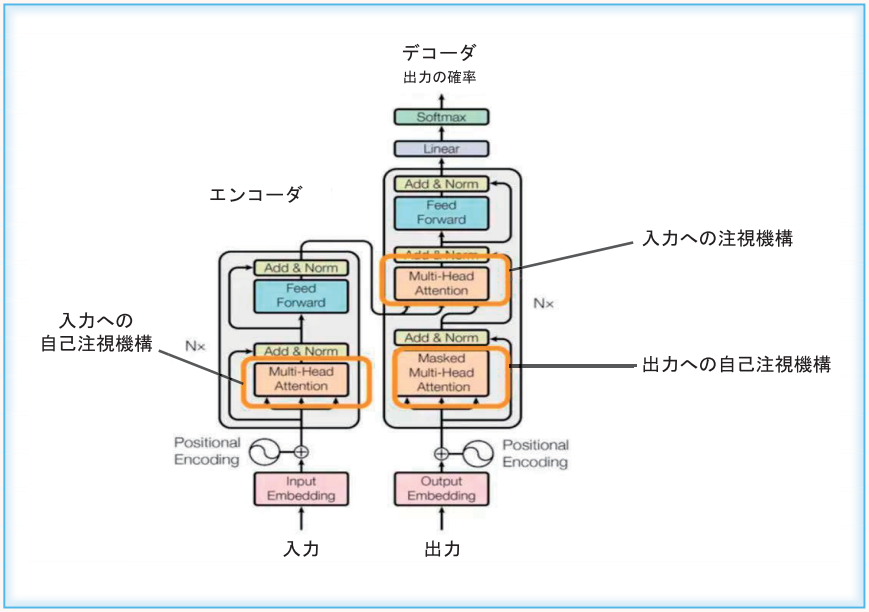


図７　トランスフォーマーの全体構成

図８　人間からのフィードバックを使う強化学習
対話の応答をする訓練の場合

しかし，確率を陽に推定する SMT では誤訳の原因を推定確率に戻って検証できたが，確率を陽に計算しない RNN やトランスフォーマーでは，何が誤りの原因かを特定するのが困難となる．エンコーダ・デコーダモデルのベクトル C や状態遷移という計算過程に誤り原因があるが，その同定は難しい．1 億超の内部パラメータと注視機構を使う計算過程に誤り原因があるトランスフォーマーでは，原因究明は更に困難となる．LLM が巨大なブラックボックスといわれる一因である．

6 Instruction 層とアライメント

トランスフォーマーは，入力を出力に変換する汎用の機構である．翻訳では，元言語のテキストを相手言語のそれに変換するが，長いテキストを抄録に変換する，あるいはユーザの入力文を適切な応答に変換するなど，様々なタスクが入力から出力への変換と見ることができる．

「次のテキストから 200 単語程度の抄録を出力して下さい．＜長いテキスト＞」や「次の日本語テキストを翻訳して下さい．＜長いテキスト＞」という入力では，＜長いテキスト＞に先行する部分がタスク指令になっており，この指令を注視機構で参照することで，＜長いテキスト＞を参照しながら抄録や翻訳というタスクを意識したテキストが生成される．

最下層の LLM は自然なテキストを出力する機能を持つが，この機能だけでは抄録や翻訳という個々のタスクに合わせた変換を行うことはできない．翻訳や抄録という入力を注視し，これを先行文脈として，それぞれのタスクを実行できるように訓練をするのが Instruction 層である．

タスクごとに LLM を再学習，訓練する過程は FT (Fine Tuning)，Instruction などと呼ばれる．このためには，訓練用の学習セットを用意する必要があるが，＜長いテキスト＞と抄録の対を大量に用意するのは簡単ではない．同じように，ユーザの入力と適切な応答の学習では，何が適切な応答であるのかの定義も明確ではない．また，応答としては可能であっても，ある種の偏見を含む応答は適切ではない．

このようなタスクに合わせた FT，Instruction を行うために，OpenAI のグループは「人間のフィードバックによる強化学習」(RLHF: Reinforcement Learning with Human Feedback) という手法を用いている．これは，入力と望ましい出力対を作る作業をシステムと人間の協働で行う枠組みである [(7)]．

例えば，対話タスクにおいては＜ユーザからの入力＞に対する複数の応答をまずシステムが生成して人間に提示し，人間が望ましいと判断する応答が優先されるように強化学習する．初めから人手で作るのではなく，システム出力の優劣だけを人間が判断する（図 8）．

この適切性の判断は個々の人間によって変化する．判断する人間が望ましくない偏見を持っていると，その偏見が強化学習に使われ，望ましくない応答を出すように強化学習が進む．このような汚染を避けるために，フィードバックするのに適した人間を選び，かつ，その人たちの訓練を注意深く行う必要があるなど，システムからの応答リストを候補集合として援用するにしても，大変な人手が掛かる作業となる．

この FT や Instruction の作業は，人間社会が持つ望ましい価値観に LLM の動作を適合させる過程という意味で，人間とシステムの価値を合わせるアライメント過程とも呼ばれている．

一般に，LLM は学習に使う大規模言語データのサイ

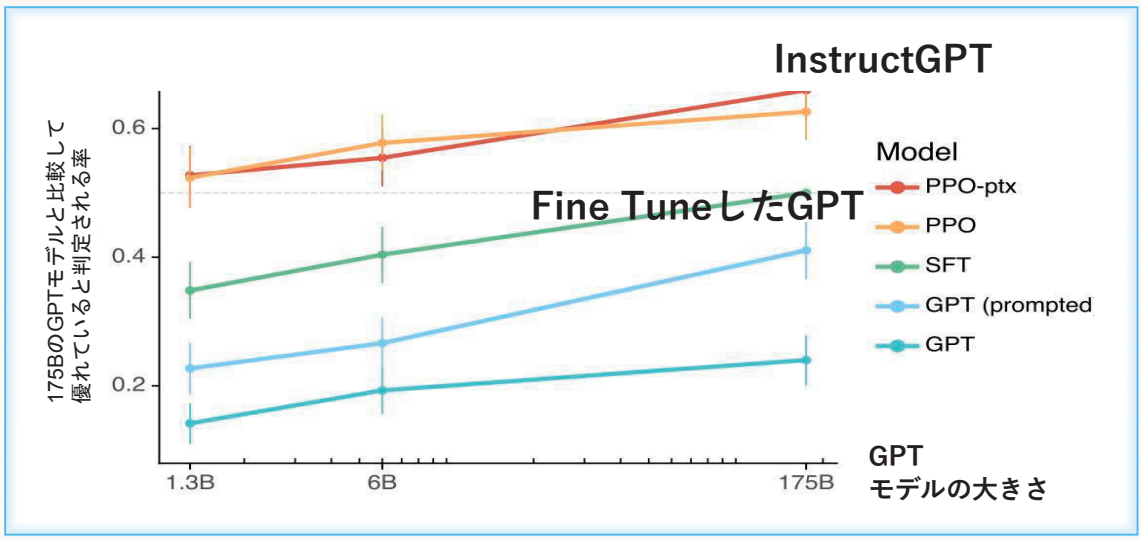


図 9　Instruction の効果 [7]

図中 B は、Billion（10 億）でモデルの内部パラメータ数を示す

ズが大きいほど，また LLM の内部パラメータ数が膨大になるほど，その性能が高くなる．しかし，この Instruction 層でのアライメントも重要な役割を果たしており，ユーザの満足度は Instruction を行った小規模モデルが，Instruction しない 100 倍大きな LLM よりも高くなることが報告されている（図 9）．

どのようなタスクを対象に Instruction を行うかは，ユーザがどのようなタスクを実行したいかに依存する．OpenAI が発表した Instruction の初期論文では，実ユーザの入力行動を観察し，頻繁に実行されるタスクの性能を上げるように Instruction を行っている．初期の頃にはこの優劣判定がされた Instruction 用の訓練データも公開され，他グループもそれを使って訓練できていた．しかし，この分野での商業化の競争が激化するに伴って，Instruction データの公開はされなくなっている．

LLM の初期学習に使われる大規模テキストデータや Instruction データが秘匿されることで，人手と大きな計算機資源を有する巨大 IT 企業による技術の独占と，システムの更なるブラックボックス化が加速し，学術界の懸念事項となっている．

7　知識・情報の蓄積庫としての LLM：LLM の二重の役割

抄録や翻訳タスクや仕様からのプログラムの作成は，入力－出力の変換タスクとみなすのが自然である．これに対して，「人工知能とは」とか，「バイデン大統領はどのような人ですか」に対してテキストを生成するタスクでは，LLM は変換器というよりは生成器の役割を果たしている．

この種のタスクでは，LLM は自らが学習に使った膨大な情報・知識の蓄積庫として，そこから関連するテキストを取り出しながらテキストを生成する．この過程も，「人工知能とは」「バイデン大統領は」という先行単語列から後続単語列を生成し，生成された単語列を先行単語列として更に後続する単語列を生成する過程が繰り返されて，自然なテキスト生成が行われる．

この種の生成タスクが可能であることが，LLM を知的なインターネット検索として使うなど，その応用範囲を非常に大きなものとしている．この種の応用では，LLM は学習用の大規模なテキスト集合を意味に基づいて索引した情報の蓄積庫としての役割を果している．この意味に基づく知的な索引構造については，次章の表現学習で議論する．

8　表現学習：計算資源とデータ

かなり駆け足で LLM とチャットボットの技術的枠組みを見てきた．読者は言語の持つ確率的な性質を反映する LLM が，自然なテキストの生成だけでなく，なぜ言語を「理解」したかのような能力を示すことができるのか，という疑問を持つであろう．

RNN からトランスフォーマーへの進化は，注視機構によって飛躍的に能力が向上した．これにより，入力だけでなくそれまでの出力を先行文脈として，後続するテキストを生成していくことが可能になった．先行して生成された部分列を参照する注視機構は，自己注視機構（Self-Attention）とも呼ばれる．

「理解」を支えるもう一つの大きな変革は，単語や単語の部分系列を，数千次元の高次元ベクトルで表現する手法である．

表 1　学習データ、内部パラメータ、消費電力の急速な増大

モデル名	オリジナル	BERT Large	GPT2	GPT3	GPT4
発表年	2017	2018	2019	2020	2023
内部パラメータ	0.45 億	3.4 億	15 億	1750 億	1 兆？
学習データサイズ	1 億語	33 億語	200 億語	3000 億語	数兆語？
学習時消費電力 MW·h	0.027	5	300	5000	500,000

これまでの説明では，単語や計算状態，入力文の情報を w_n，S_n，C といった記号で表現してきた．実際にトランスフォーマーが取り扱うのは，リンゴや Apple といった個々の単語ではない．これらの単語（あるいは，単語より小さな部分単語という単位，トークンとも呼ばれる）や単語列に与えられた高次元のベクトルである．また C や S_n に対応するものも同様に非常に高い次元を持ったベクトルである．

単語や単語列にどのようなベクトル表現を与えるかを学習することから，この学習過程は表現学習（Representation Learning）とも呼ばれる．大規模なテキスト集合から，そのテキスト集合を生成する確率が最大になるように単語や単語列のベクトル表現と 1,000 億を超える内部パラメータを学習する．この表現学習が，LLM の「理解」能力を支えている．

機械学習の応用では学習器を訓練する正解集合を作るのが大変であった．しかし LLM の学習では，世の中に出回っている大規模テキスト集合そのものが正解であり，そのテキスト集合全体を生成する確率を上げるように学習するので，正解集合を人手で作成する必要はない．収集したデータ集合自体を正解集合とするこの手法は，自己教師付き学習（Self-supervised Learning）と呼ばれる．

①インターネットの検索サービスのために，膨大なテキスト集合が既に収集されていたこと

②膨大なテキスト集合を使った自己教師付き学習では，非常に大きな計算資源が必要となるが，これが巨大な GPU クラスタとそれを使った並列計算で可能になったこと

という二つの要因がそろったことで，LLM の開発が急速に進展した．少し古いがトランスフォーマーの論文発表から現在に至る学習用テキスト集合のサイズと内部パラメータ数を表 1 に示す．表から分かるように，学習用テキストのサイズ，内部パラメータ数が加速度的に増大していること，またそれに伴って学習時に必要な GPU クラスタの消費電力も急激に増大している．

言語処理の分野では，以前から単語間の意味的な類似性や反意性だけでなく，表面上かなり違った構造と単語から成る文間にも意味的な重なりや反義といった関係があり，これをベクトル空間といった距離空間で捉えたい，という要求はあった [8]．ただ，単語や文の意味空間の本格的な構築は，トランスフォーマーの大規模な自己教師付き学習による表現学習で初めて可能となった．

実際，学習結果として単語や文に割り当てられるベクトルを観察すると，英語と日本語であっても同義的な単語の距離は極めて近くなるとか，翻訳文の対もこのベクトル空間での距離が近くなっているなど，表現学習の結果，言語が持つ意味の空間が抽象化されてうまく捉えられている．この超高次元のベクトルは，意味的な側面だけでなく，表現の丁寧さや形式性，統語的な性質など，言語表現の持つ様々な性質を反映している．

言語表現が持つ性質や素性を学習した LLM をあらゆるタスクの基盤とすることで，Instruction 層でのタスクごとの訓練がうまく機能し，タスクごとの正解例が少なくても訓練が迅速に進む．また，LLM という情報の蓄積庫から入力と意味的に関連の深いものを取り出すこともできる．

ただ，表現学習の結果のベクトル空間がどのようなものになっているかは，まだよく理解できておらず，どのような種類の情報がどのように反映されているのかなど，獲得された表現空間の構造を理解することは，今後の大きな課題である．

9 LLM の行う「理解」

表現学習とそれを使ったテキスト生成によって，チャットボットは確率的なオウム（Stochastic Parrot）とは呼べないものになっている．人間から教えられた文を，全く文脈と無関係に反すうするオウム返しをするわけではない．チャットボットは表現学習で構築された意味の空間と注視機構を使うことで，文脈との関連を考慮して学習データ中の関連するテキスト部分を生成したり，それを組み合わせた新たな文を創造的に作り出す．

関連するテキストの取出しと創造的な文の生成は，LLM が抱える課題と直接関係している．前者は，学習データが生成テキスト中にリークすることで，知的所有

権侵害やひょう窃，個人情報の流出，偏見による汚染を引き起こす．後者は，実在しない人物や人物を混同した自然なテキストを生成するといった，自然さと真理性とがかい離した幻覚（Hallucination）現象を引き起こす．学習テキストに含まれる誤った情報のリークも，幻覚現象とともにチャットボットの信頼性を損なう．実際には，学習データからの誤情報のリークとチャットボットが生成した誤情報との混在が，誤り原因特定をより複雑化している．

Instruction 層でのアライメントの必要性や学習データの誤りのリークは，LLM 自体には真理性やテキストの適切さを判断する能力がないことを示している．学習データや Instruction データを人間が精査することでしか，人間の持つ価値観の反映や真理性が担保できない．

現時点では，自己教師付き学習のための学習テキストのクリーンアップ処理で，あからさまに不適切なテキストや個人情報を取り除いているが，この過程で真理性の精査までを行うことは不可能であろう．

以上のことは，LLM やそれを使ったチャットボットは，真理性や適切性を判断する倫理規範を内在的に有した積極的な知能ではなく，人間が与えるデータをそのまま受け入れる受動的な知能となっている．

この知能の受動性は，確率に基づくテキスト生成過程（**5.** 参照）にも見られる．人間の場合には，テキストを生成する過程では，複数の相反する言明が可能な場合には，それぞれの言明を比較・吟味し，場合によっては，言明を支持する別データを参照しながら，より真理性があると思われる言明を選び取る積極的な知能の働きがある．また，この過程では，個人が持つ価値観や未来への展望に合わせて，適切な言明が選び取られる．このような価値観と真理性に基づく積極的な情報の吟味と選択を行う自覚的な知能の働きは，文脈との相互関連で確率的に自然な出力を行う LLM にはない．

「LLM は，一人の人間では到底処理できない量のテキストを読み込んでいる．したがって，LLM が生成するテキスト（判断）は，人間の判断よりもより合理的で正しい」という主張は，誤った危険なものである．主張の前半は正しいが，LLM は真理性や適切性の吟味をしているわけではない．このような吟味を経ない言明が，「より合理的で正しい」言明となっている保障はない．その判断は，人間の側に委ねられる．

10 おわりに

個人としての人間知能は，限られた少量の情報にしかアクセスできない．また，情報を取捨選択する積極的な人間知能は，逆に「自分にとって都合の良い」情報，「自分の価値観に合った」情報を正しいと考えるバイアスなど，様々なバイアスを持っている．今後の全体的な方向としては，LLM と人間という二つの知能が，その欠陥を相互に補う形でより良い判断を行う人間と人工知能の協調の枠組みの構築が重要となろう [(9)]．

本稿ではテキスト処理を中心に議論をしてきたが，トランスフォーマーは直接の言語表現ではなくベクトル表現を対象とすること，注視機構により一次元の単語系列の処理という言語処理の枠組みから解放されたことから，テキスト以外の様々なデータ（画像，音響，表形式のデータ，プログラム，たんぱく質構造など）を対象とする情報処理に適用でき，その応用範囲も急速に広がっている．

以下では，本稿の議論の範囲で，今後の研究方向を思い付くままに列挙しておく．

①自然さと情報の貯蔵庫の役割の分離：自然さを重視する LLM では，学習データの正誤や適切性は問題にならない．それに対して，情報の貯蔵庫としての LLM では，これらはシステムの信頼性に直接影響する．この二つの役割を分離し，情報の貯蔵庫は別個に用意し，LLM の構成する意味空間を使って索引とすることで，生成する内容はこの別個用意したテキスト集合を主とすることが考えられる．別個に用意された集合の正しさが保証されていれば，学習データのリークによる誤りの混入を極小化できる．この RAG（Retrieval Augumented Generation）は今後，外付けの陽な知識構造（知識グラフなど）との結合など，更に展開していくであろう．

②ブラックボックス性の解明，LLM の設計論：現在の LLM は，膨大な学習テキストから学習された高次元のベクトル空間，1,000 億を超えるパラメータで実現された計算過程など，巨大なブラックボックスとなっている．これをより理解しやすいものにし，適用タスクのクラスに特化した訓練を行うなど，LLM 構築のための設計論が必要となろう．一枚岩の LLM ではなく，内部に部分ネットワークというモジュラリティを導入する（Mix of Experts）など，LLM の構造化の研究も進展し始めている [(10)]．また，現実の応用では，巨大なモデルを特定タスクに特化して小形化する蒸留（Distillation）などの技術も重要となる．

③マルチモーダル LLM: テキストだけに閉じたモデルは，情報の正しさの吟味やユーザの環境に応じて適切な応答をするには不十分である．テキストという虚の情報以外の，実の世界をより直接的に反映した数値情報，画像，音響などの情報を取り込むマルチモーダルな LLM の研究は，今後，ますます活発化しよう．実世界で行動する知能体としてのロボットに，LLM が持つ世界に関する知識を持たせる研究も，活発化している．

④汎用人工知能と分野に特化した知能の融合：LLMは言語という特定分野に依存しない情報表現を処理対象とすることで，分野に依存しない汎用知能の計算化に成功してきた．その一方で，人間は言語という汎用のメディアでは捉えられない分野固有の情報表現，例えば数式といったものを用いて分野固有の精細な知識を表現し，シミュレーションなどの手法で計算化してきた．また，従来の人工知能では熟練者の暗黙知のように言語化できない知識を，どのように計算化するかに焦点を当てた研究も盛んである．今後は，汎用人工知能と分野依存な人工知能，分野依存の知識処理との自然な融合が重要になろう．

LLM研究は，ここ２年の間に爆発的な展開を見せた．ただ，この展開はこれから情報処理の分野で起こる変革の端緒に過ぎない．より大きな広範囲な変革が，情報処理・通信・計算機アーキテクチャの分野で進展していくと考えられる．

■文献

(1) W. H. Walters and E. I. Wilder, “Fabrication and errors in the bibliographic citations generated by ChatGPT,” Sci. Rep. 13, Sept. 2023.
(2) T. R. Hannigan, I. P. McCarthy, A. Spicer, Beware of botshit: How to manage the epistemic risks of generative chatbots, Business Horizons, 2024.
(3) 永田昌明，渡部太郎，塚田　元，“機械翻訳最新事情：（上）統計的機械翻訳入門，”情報処理，vol. 49, no. 1, Jan. 2008.
(4) 須藤克仁，“ニューラル機械翻訳の進展，系列変換モデルの進化とその応用，”人工知能，vol. 34, no. 4, pp. 437-445, July 2019.
(5) D. Bahdana, K. Cho, and Y. Benjo, “Neural machine translation by jointly learning to align and translate,” arXiv:1409.0473, Sept. 2014.
(6) A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” arXiv:1706.03762, Aug. 2023.
(7) L. Ouyang, J. Wu, X. Jiang, Di. Almeida, C. L. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, J. Schulman, J. Hilton, F. Kelton, L. Miller, M. Simens, A. Askell, P. Welinder, P. Christiano, J. Leike, and R. Lowe, “Training language models to follow instructions with human feedback,” arXiv:2203.02155, March 2022.
(8) M. Baroni and A. Lemci, “Distributional memory: a general framework for corpus-based semantics.” Computational Linguistics, Vol. 36, no. 4, Dec. 2010.
(9) National Academies of Sciences, Engineering, and Medicine, Human-AI teaming (STATE-OF-THE-ART AND RESEARCH NEEDS), National Academies Press. 2022.
(10) J. Pfeiffer, S. Ruder, I. Vulić, and E. M. Ponti, “Modular deep learning,” arXiv:2302.11529, Feb. 2023.

辻井潤一

1973 京大大学院了．同大学助手，助教授を経て，1988 マンチェスター大教授，1995 東大教授．2011 マイクロソフト研究所首席研究員．2015 から産総研人工知能研究センターセンター長，現在産総研フェロー，マンチェスター大教授，東大名誉教授．大川賞，ACL・LTA 賞などを受賞．令和６年度文化功労者に選出．ACL フェロー．京大工博．

深層生成モデルの基礎と応用

Fundamentals and Applications of Deep Generative Models

金子卓弘　Takuhiro Kaneko†

Summary　2000 年代初頭に第三次 AI ブームが始まって以来，深層学習は様々な分野で技術革新を起こしてきている．特に，近年大きな発展を遂げた技術の一つとして，深層学習に基づく生成モデル（深層生成モデル）を用いた画像生成 AI 技術がある．本論文では，このような深層生成モデルの基礎と応用について解説を行う．まず，深層生成モデルの基礎では「認識」と「生成」という対極をなすタスクについて比較をしながら生成とは何かについて説明を行う．その後，深層生成モデルの代表的な四つのモデルである変分自己符号化器，フローベースモデル，敵対的生成ネットワーク，拡散モデルの仕組み，理論的背景，特徴，発展について解説を行う．続いて，深層生成モデルの応用では，重要な応用技術の一つとして条件付き生成モデルを紹介する．最後に，深層生成モデルの今後を展望して本論文の結びとする．

Key Words　生成 AI，画像生成，深層学習，深層生成モデル

1 はじめに

2000 年代初頭に第三次 AI ブームが始まって以降，深層学習は様々な分野で技術革新を起こしてきている．その一つとして挙げられるのは，深層学習に基づく生成モデル（深層生成モデル）を用いた画像生成 AI 技術である．長年の間，生成画像と実画像のギャップを埋めることは難しく，実用化の妨げとなっていたが，近年の深層生成モデルの急速な発展により，人でも見分けがつかないようなリアリティのある画像が生成できるようになりつつある．これにより，画像生成 AI 技術は実応用化が進み，社会に大きなインパクトを与えている．本論文では，このような深層生成モデルの基礎と応用について解説を行う．

深層生成モデルの基礎（**2.**）では，「生成とは何か？」という問いに答えるため，生成と対極をなすタスクである認識と対比を行いながら生成について説明を行う（**2. 1**）．その後，深層生成モデルの代表的な四つのモデルである変分自己符号化器（VAE: Variational Autoencoder）[(1),(2)]，フローベースモデル（Flow-Based Model）[(3),(4)]，敵対的生成ネットワーク（GAN: Generative Adversarial Networks）[(5)]，拡散モデル（Diffusion Model）[(6)~(8)] について仕組み，理論的背景，特徴，発展の解説を行う（**2. 2**）．

深層生成モデルの応用（**3.**）では，多岐にわたる応用技術の中でも特に重要な技術の一つとして条件付き生成モデル，すなわち，クラスラベルやテキスト，画像などの与えられた条件情報に従ってデータを生成可能な生成モデルについて紹介する．具体的には，まず，条件なし生成モデルと対比を行いながら条件付き生成モデルについて概説を行った後（**3. 1**），深層生成モデルの中でも特に広く使われている敵対的生成ネットワーク（**3. 2**）と拡散モデル（**3. 3**）に着目して条件付き生成モデルへの拡張について解説する．

最後に，**4.** では，深層生成モデルの今後を展望して本論文の結びとする．

なお，本論文では説明の都合上，データとしては画像を中心に取り上げて説明を行うが，同様のことは音声やテキストなどのほかのデータでも成り立つことに留意されたい．

† 日本電信電話株式会社，厚木市
NTT Corporation，Atsugi-shi，243-0198 Japan

2 深層生成モデルの基礎

2. 1 認識と生成

本節では「生成とは何か？」という問いに答えるため，機械学習分野の代表的なタスクである認識と対比を行いながら生成について説明する．

図 1 に画像認識と画像生成の比較を示す．画像認識では，入力として画像，すなわち，高さ × 幅 × 色から成る高次元情報が与えられたとき，その画像を端的に説明することが可能な低次元表現に変換

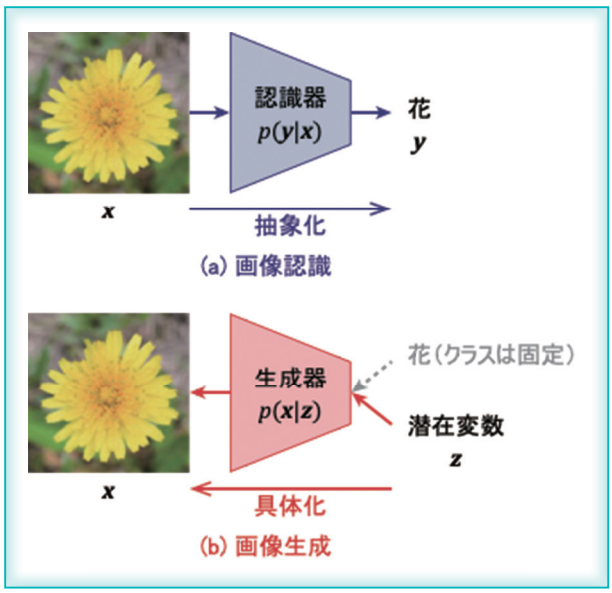


図 1　画像認識と画像生成の比較

することが目的となる．例えば，図 1(a) の場合，入力として 128×128×3 の画像 $\boldsymbol{x}$ が与えられたとき，「花」というクラス $\boldsymbol{y}$ に分類することによって画像を端的に説明することを可能にする．このように認識は高次元の画像 $\boldsymbol{x}$ から低次元のクラスラベル $\boldsymbol{y}$ の確率 $p(\boldsymbol{y}|\boldsymbol{x})$ を推定する処理で，情報を抽象化するプロセスであると言える．

一方，画像生成では，この逆問題を解くことが目的となる．具体的には，図 1(b) に示すように，「花」という抽象的なクラス表現が与えられたとき，その花に該当する画像を生成することが目的となる．すなわち，生成は情報を具体化するプロセスであると言える．この具体化を行う際に問題になることは，一口に「花」と言っても対応する花の画像は無数にあり，入出力の対応関係に曖昧性が存在することである．画像生成ではこの曖昧性を表現するため，図 1(b) に示すように，入力として潜在変数 $\boldsymbol{z}$ を導入する．ここで，生成器は「花」の生成に特化したもの，すなわち，「花」というクラスは固定で入力として無視できるとすると，生成は潜在変数 $\boldsymbol{z}$ から画像 $\boldsymbol{x}$ の確率 $p(\boldsymbol{x}|\boldsymbol{z})$ を推定する処理と言える．この潜在変数はガウス分布などの分布形状が単純な分布からランダムにサンプリングされるもので，この潜在変数をサイコロのように振り直すことによって様々な花の画像を生成することが可能になる．

＊1　より一般的には，画像 $\boldsymbol{x}$ と潜在変数 $\boldsymbol{z}$ の同時分布 $p(\boldsymbol{x},\boldsymbol{z})$ を学習することが生成モデルの目的となるが，本稿では説明の簡単化のため，$p(\boldsymbol{z})$ はガウス分布などの固定の分布で表されて最適化する必要がないと仮定し，条件付き分布 $p(\boldsymbol{x}|\boldsymbol{z})=p(\boldsymbol{x},\boldsymbol{z})/p(\boldsymbol{z})$ を学習する場合について説明する．

表 1　深層生成モデルの特徴の比較

モデル	品質	速度	多様性
変分自己符号化器（VAE）	低い	速い	大きい
フローベースモデル（Flow-Based Model）	低い	速い	大きい
敵対的生成ネットワーク（GAN）	高い	速い	小さい
拡散モデル（Diffusion Model）	高い	遅い	大きい

2.2　深層生成モデル

2.1 で説明した生成器を実現するには，潜在変数 $\boldsymbol{z}$ と画像 $\boldsymbol{x}$ の対応関係が適切に得られている必要がある＊1．しかし，従来の生成モデル，例えば，正規分布の組合せによってデータ分布を表現する混合ガウスモデル（GMM: Gaussian Mixture Model）には十分な表現能力がなく，画像のような高次元データをモデル化対象にし，潜在変数と画像の対応関係を学習することは難しかった．これに対して，本論文のメイントピックである深層生成モデルは，高い表現能力を持つ深層学習モデルを取り入れることで，潜在変数と画像の複雑な対応関係を学習することを可能にし，これにより，画像のような高次元データをも再現できる生成モデルが学習できるようになりつつある．

潜在変数 $\boldsymbol{z}$ と画像 $\boldsymbol{x}$ の対応関係を学習する方法としては様々なものが提案されているが，代表的なモデルとして，変分自己符号化器（VAE）[1], [2]，フローベースモデル（Flow-Based Model）[3], [4]，敵対的生成ネットワーク（GAN）[5]，拡散モデル（Diffusion Model）[6]~[8] の四つがある．各深層生成モデルの特徴を，深層生成モデルの性質を説明する上で重要な三つの観点である品質，速度，多様性に着目して整理したものを表 1 に示す．各深層生成モデルの仕組み，理論的背景，特徴，発展の詳細については，次項以降で解説する．

なお，本稿では，誌面の都合上，代表的な深層生成モデルである上記四つのモデルに着目して説明を行うが，ほかにも様々な深層生成モデルが存在することに留意されたい．例えば，データ分布 $p(\boldsymbol{x})$ を，エネルギー関数 $E(\boldsymbol{x})$ を導入して $p(\boldsymbol{x})=\exp\left(-\beta E(\boldsymbol{x})\right)/\int_{\boldsymbol{x}'}\exp\left(-\beta E(\boldsymbol{x}')\right)d\boldsymbol{x}'$（$\beta$ は逆温度パラメータで，$\beta\to 0$ のとき $p(\boldsymbol{x})$ は一様分布，$\beta\to\infty$ のとき $p(\boldsymbol{x})$ はエネルギーが一番低いところだけ大きな確率を持つ分布になる）で表すエネルギーベースモデル（EBM: Energy-Based Model）[9] や，データ分布 $p(\boldsymbol{x})$ を，潜在変数を含めずに観測

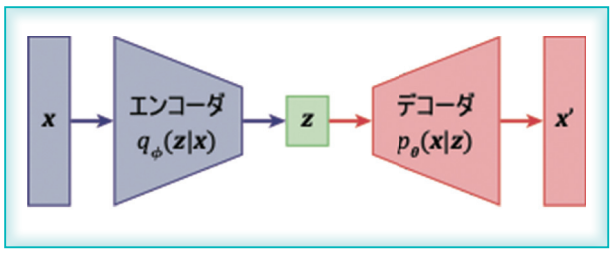


図２　変分自己符号化器（VAE）の構造

データの各点の条件付き分布の積 $p(\boldsymbol{x}) = \prod_{i=1}^{N} p(\boldsymbol{x}_i|\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{i-1})$（$N$ はデータの点数）によって表す自己回帰モデル（Autoregressive Model，PixelCNN[10] など）などがある．また，個別の深層生成モデルだけではなく，複数の深層生成モデルを組み合わせることで，各モデルの長所と短所を相互補完したモデルも多数提案されていることも留意されたい．

2.2.1　変分自己符号化器（VAE）

仕組み　上述のように，深層生成モデルでは，$\boldsymbol{z}$ と $\boldsymbol{x}$ の対応関係を学習することが目的となる．ここで課題になることは，生成器の出力である画像 $\boldsymbol{x}$ に対しては学習データを集めることができるものの，その画像に対応する最適な潜在変数 $\boldsymbol{z}$ は非自明で教師データを得ることが難しいということである．そこで，変分自己符号化器[1), 2)]では，図２に示すように，$\boldsymbol{z}$ から $\boldsymbol{x}$ を生成する生成器 $p_\theta(\boldsymbol{x}|\boldsymbol{z})$（変分自己符号化器の文脈では「デコーダ」と呼ばれる．）に加えて，$\boldsymbol{x}$ から $\boldsymbol{z}$ を推論する推論器 $q_\phi(\boldsymbol{z}|\boldsymbol{x})$（変分自己符号化器の文脈では「エンコーダ」と呼ばれる．）を導入する．深層生成モデルでは，各分布はニューラルネットワークを用いてパラメタライズされており，$\boldsymbol{\theta}$ と $\boldsymbol{\phi}$ はモデル（ニューラルネットワーク）のパラメータを表す．そして，エンコーダ $q_\phi(\boldsymbol{z}|\boldsymbol{x})$ によって $\boldsymbol{x}$ から $\boldsymbol{z}$ を推定し，デコーダ $p_\theta(\boldsymbol{x}|\boldsymbol{z})$ によって推定された $\boldsymbol{z}$ から $\boldsymbol{x}$ を再構成するプロセスの中で，パラメタライズされたエンコーダとデコーダを最適化することを考える．最適化を行う際に重要なことは，エンコーダ $q_\phi(\boldsymbol{z}|\boldsymbol{x})$ が表す $\boldsymbol{x}$ と $\boldsymbol{z}$ の分布とデコーダ $p_\theta(\boldsymbol{x}|\boldsymbol{z})$ が表す $\boldsymbol{x}$ と $\boldsymbol{z}$ の分布が一致していることである．変分自己符号化器では，エンコーダ $q_\phi(\boldsymbol{z}|\boldsymbol{x})$ の出力に対して事前分布制約を課すことによってこれを実現している．詳細については，次の理論的背景で説明する．

理論的背景　変分自己符号化器の学習では，最ゆう推定，すなわち，モデルが学習データ $\boldsymbol{x}$ を表現したときのもっともらしさ（ゆう度）$p_\theta(\boldsymbol{x})$ が最大化するように学習が行われる．なお，実際には，ゆう度 $p_\theta(\boldsymbol{x})$ は学習データ集合 $\mathcal{D}$ の中の全ての $\boldsymbol{x} \in \mathcal{D}$ に対して計算され，それらの合計値が最適化されるが，ここでは表記の簡便化のため，ある一つのデータ $\boldsymbol{x}$ に対するゆう度 $p_\theta(\boldsymbol{x})$ に着目して説明する．

ゆう度の対数，すなわち対数ゆう度 $\log p_\theta(\boldsymbol{x})$ は以下の式で表される．

$$\log p_\theta(\boldsymbol{x}) = -\log p_\theta(\boldsymbol{z}|\boldsymbol{x}) + \log p_\theta(\boldsymbol{x}, \boldsymbol{z})$$
$$= \mathbb{E}_{\boldsymbol{z}\sim q_\phi}\left[\log \frac{q_\phi(\boldsymbol{z}|\boldsymbol{x})}{p_\theta(\boldsymbol{z}|\boldsymbol{x})}\right] + \mathbb{E}_{\boldsymbol{z}\sim q_\phi}\left[\log p_\theta(\boldsymbol{x}, \boldsymbol{z}) - \log q_\phi(\boldsymbol{z}|\boldsymbol{x})\right]$$
$$= D_{KL}\left(q_\phi(\boldsymbol{z}|\boldsymbol{x}) \| p_\theta(\boldsymbol{z}|\boldsymbol{x})\right) + \mathbb{E}_{\boldsymbol{z}\sim q_\phi}\left[\log p_\theta(\boldsymbol{x}, \boldsymbol{z}) - \log q_\phi(\boldsymbol{z}|\boldsymbol{x})\right]$$

ここで，第一式では $p_\theta(\boldsymbol{x}, \boldsymbol{z}) = p_\theta(\boldsymbol{z}|\boldsymbol{x})p_\theta(\boldsymbol{x})$ の関係式を用いている．$\boldsymbol{z}$ から $\boldsymbol{x}$ を求めるデコーダ $p_\theta(\boldsymbol{x}|\boldsymbol{z})$ は一般的には不可逆なニューラルネットワークでパラメタライズされているため，分布が単純なものでない限り $p_\theta(\boldsymbol{x}|\boldsymbol{z})$ を使ってその逆変換 $p_\theta(\boldsymbol{z}|\boldsymbol{x})$（第一式右辺第一項）を表すのは困難である．そこで，第二式では，代わりに $p_\theta(\boldsymbol{z}|\boldsymbol{x})$ の近似分布 $q_\phi(\boldsymbol{z}|\boldsymbol{x})$（すなわち，エンコーダ）を導入している．ここで，$D_{KL}$ は二つの分布のカルバックライブラ情報量（Kullback-Leibler Divergence）を表す*2．第三式の第一項は計算困難であるが，非負の性質を持つため 0 以上と近似する．すると，対数ゆう度 $\log p_\theta(\boldsymbol{x})$ は，以下の変分下界で表すことができる．

$$\log p_\theta(\boldsymbol{x}) \geq \mathbb{E}_{\boldsymbol{z}\sim q_\phi}\left[\log p_\theta(\boldsymbol{x}, \boldsymbol{z}) - \log q_\phi(\boldsymbol{z}|\boldsymbol{x})\right]$$
$$= \mathbb{E}_{\boldsymbol{z}\sim q_\phi}\left[\log p_\theta(\boldsymbol{x}|\boldsymbol{z}) + \log p_\theta(\boldsymbol{z}) - \log q_\phi(\boldsymbol{z}|\boldsymbol{x})\right]$$
$$= \mathbb{E}_{\boldsymbol{z}\sim q_\phi}\left[\log p_\theta(\boldsymbol{x}|\boldsymbol{z})\right] - D_{KL}\left(q_\phi(\boldsymbol{z}|\boldsymbol{x}) \| p_\theta(\boldsymbol{z})\right)$$

ここで，第二式では $p_\theta(\boldsymbol{x}, \boldsymbol{z}) = p_\theta(\boldsymbol{x}|\boldsymbol{z})p_\theta(\boldsymbol{z})$ の関係式を用い，第三式ではカルバックライブラ情報量を用いて式を整理している．

変分自己符号化器では，対数ゆう度の変分下界である最終式を最大化することによって，対数ゆう度の近似を最大化することを目指す．ここで，最終式に対して直感的な説明を与えると，第一項は，エンコーダ $q_\phi(\boldsymbol{z}|\boldsymbol{x})$ とデコーダ $p_\theta(\boldsymbol{x}|\boldsymbol{z})$ を使って再構成した $\boldsymbol{x}$ に対するゆう度，つまり，負の再構成損を表し，第二項は，エンコーダ $q_\phi(\boldsymbol{z}|\boldsymbol{x})$ を使って推定した $\boldsymbol{z}$ があらかじめ定めた $\boldsymbol{z}$ の分布 $p_\theta(\boldsymbol{z})$ に従うようにする事前分布制約を表す．つまり，上式は，変分自己符号化器では潜在変数に事前分布制約を与えながら再構成損が最適化されることを表している．

なお，実装の際には $\boldsymbol{z}$ の事後分布 $q_\phi(\boldsymbol{z}|\boldsymbol{x})$ は平均

＊2　カルバックライブラ情報量は非負性と非対称性を持つ情報量である．二つの分布 P と Q に対するカルバックライブラ情報量を $D_{KL}(P\|Q)$ とすると，Q について $D_{KL}(P\|Q)$ を最小化したとき Q は $P \neq 0$ において P を網羅しようとし，Q について $D_{KL}(P\|Q)$ を最小化したとき Q は $Q \neq 0$ において P を網羅しようとする．

μ_ϕ，標準偏差 σ_ϕ の正規分布 $\mathcal{N}(\mu_\phi, \sigma_\phi)$，$z$ の事前分布 $p_\theta(z)$ は標準正規分布 $\mathcal{N}(\mathbf{0}, \mathbf{I})$ でよく表され，最終式の第二項により両者が一致するように最適化される．また，最終式の第一項の期待値の計算では，$z \sim q_\phi$ において z を確率的にサンプリングすることが必要になるが，これについては，下式で表されるリパラメタライゼーショントリック（Reparameterization Trick）[1] を用いることで，勾配伝搬が可能な形式で実装することができる．

$$z = \mu_\phi + \sigma_\phi \epsilon$$

ここで，ϵ は雑音（ノイズ）で，標準正規分布 $\mathcal{N}(\mathbf{0}, \mathbf{I})$ からサンプリングされる．

特徴　変分自己符号化器について品質，速度，多様性の三つの観点に着目して特徴を説明する．品質に関しては，再構成損を求める際に x の分布 $p(x)$ に対して何らかの近似を用いて陽に表現する必要があり，その結果，統計的平均化が発生してぼやけた画像が生成されやすいところがある．一方，速度に関しては一回のフォワード計算で生成できるため速い．多様性に関しては，近似ではあるものの学習データ集合 $\mathcal{D}$ の中の全ての $x \in \mathcal{D}$ に対してゆう度 $p_\theta(x)$ を最大化することができるため，学習データの多様性をカバーすることが可能である．

また，ほかの特徴についても説明すると，変分自己符号化器では学習の一環でエンコーダが学習されるため，データ x の分布の学習だけではなく，入力データ x に対応する潜在変数 z の表現も獲得，つまり，z の分布の学習も同時にできることが強みである．

更に，後述するフローベースモデルでは可逆制約，拡散モデルではデータサイズを維持したまま拡散・逆拡散を行うという制約により，観測変数 x と潜在変数 z の次元を同じにする必要があるが，変分自己符号化器ではそのような制約がないため x と z で異なる次元数，例えば，z を低次元のコンパクトな変数に設定できるのも強みである．

発展　変分自己符号化器では潜在変数 z を連続表現しており，潜在変数の事後分布と事前分布が一致するように学習すると，入力条件を無視してデータの最頻値を出力するように学習してしまう問題（Posterior Collapse）が発生することが知られている．この問題を回避するために提案されたのが Vector Quantized VAE（VQ-VAE）[11] である．VQ-VAE では，ベクトル量子化を用いて潜在変数 z を埋込ベクトルに最も近くなるように学習することで潜在変数 z の離散表現を実現し，Posterior Collapse の問題を回避して高品質な画像を生成することを可能にしている．

また，VQ-VAE の発展である VQ-VAE-2 [12] では，更に階層的な潜在変数を導入することで高解像度かつ大規模な画像への対応も可能にしている．

2.2.2　フローベースモデル（Flow-Based Model）

仕組み　変分自己符号化器では，デコーダ $p_\theta(x|z)$ の逆変換である $p_\theta(z|x)$ に対して，別のモデルであるエンコーダ $q_\phi(z|x)$ を用いて近似することで計算をしやすくしていた．しかし，この近似により対数ゆう度 $\log p_\theta(x)$ を直接最適化することができなくなってしまっていた．これに対して，フローベースモデル [3], [4] では，ニューラルネットワークのネットワーク構造を工夫することによって，デコーダ $p_\theta(x|z)$ を用いてその逆変換であるエンコーダ $p_\theta(z|x)$ を表現できるようにし，対数ゆう度 $\log p_\theta(x)$ を直接最適化できるようにする．言い換えると，フローベースモデルでは，エンコーダとデコーダが可逆の性質を持つようにネットワーク構造の工夫を行う．

フローベースモデルの論文でよく使われる用語を用いて再整理して説明すると，図 3 に示すように，x から z の推定を行う関数を f_θ（フロー，θ はモデルパラメータ）とすると，z から x を生成する画像生成は $x = f_\theta^{-1}(z)$（インバース）で表される．ここで，f_θ^{-1} は f_θ の逆変換を表す．変分自己符号化器では，エンコーダとデコーダの二つのニューラルネットワークの最適化を行う必要があったが，フローベースモデルでは，フローの最適化が行えれば，自動的にインバースも求めることができるのが大きな特徴である．

理論的背景　変分自己符号化器と同様に，フローベースモデルでも最ゆう推定が行われる．潜在変数の事前分布 $p_\theta(z)$ とフロー $z = f_\theta(x)$ を用いて，対数ゆう度 $\log p_\theta(x)$ を整理すると，以下のようになる．

$$\log p_\theta(x) = \log p_\theta(z) + \log \left| \det \frac{\partial z}{\partial x} \right|$$
$$= \log p\big(f_\theta(x)\big) + \log \left| \det \frac{\partial f_\theta(x)}{\partial x} \right|$$

ここで，第一式では変数変換の公式を用いており，$\det \frac{\partial z}{\partial x}$ はヤコビ行列式（ヤコビアン，Jacobian）を表す．第二式では，$z = f_\theta(x)$ を用いて変数の置換えを行っている．この定式化により，フローベースモデルでは，可逆でかつヤコビ行列式 $\det \frac{\partial f_\theta(x)}{\partial x}$ を計算可能な f_θ を用意することができれば，対数ゆ

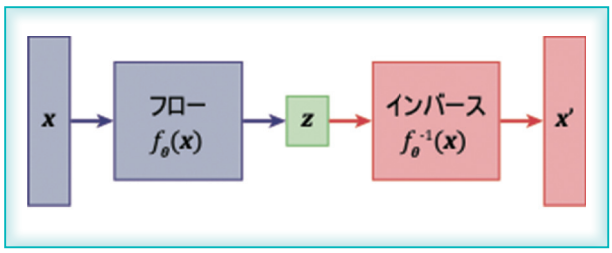


図３　フローベースモデル（Flow-Based Model）の構造

う度を近似なしで計算することが可能である．

ここで，一つ問題になることは，可逆でかつヤコビ行列式 $\det\frac{\partial f_\theta(x)}{\partial x}$ を計算可能な f_θ をどうやって実現するかということだが，例えば，Real NVP (Real-valued Non-Volume Preserving)[4] では，以下の式で表されるアフィンカップリング層を複数層積み重ねたものが使われる．

$$\boldsymbol{y}_{1:d} = \boldsymbol{x}_{1:d}$$

$$\boldsymbol{y}_{d+1:D} = \boldsymbol{x}_{d+1:D} \odot \exp\left(s_{\boldsymbol{\theta}_s}(\boldsymbol{x}_{1:d})\right) + t_{\boldsymbol{\theta}_t}(\boldsymbol{x}_{1:d})$$

ここで，$\boldsymbol{x}_{1:D}$ と $\boldsymbol{y}_{1:D}$ はこの層の入出力，D は次元，$\odot$ はアダマール積を表す．また，$s_{\boldsymbol{\theta}_s}$ と $t_{\boldsymbol{\theta}_t}$ はそれぞれスケールと移動の関数（$\boldsymbol{\theta}_s$ と $\boldsymbol{\theta}_t$ はそれぞれの関数のモデルパラメータ）で，両者ともに d 次元の実数から $D-d$ 次元の実数に写像する関数である．第一式では，１次元目から d 次元目までは，恒等変換が行われることを表し，第二式では，$d+1$ 次元目から D 次元目までは，$\boldsymbol{x}_{1:d}$ の値に基づいてアフィン変換が行われることを表す．

可逆性について見ると，逆変換は以下の式によって表すことができるため可逆性が成り立つ．

$$\boldsymbol{x}_{1:d} = \boldsymbol{y}_{1:d}$$

$$\boldsymbol{x}_{d+1:D} = \left(\boldsymbol{y}_{d+1:D} - t_{\boldsymbol{\theta}_t}(\boldsymbol{y}_{1:d})\right) \odot \exp\left(-s_{\boldsymbol{\theta}_s}(\boldsymbol{y}_{1:d})\right)$$

また，ヤコビ行列式の計算可能性についても見ると，ヤコビ行列は以下の下三角行列で表され，

$$\frac{\partial f_{\boldsymbol{\theta}}(\boldsymbol{x})}{\partial \boldsymbol{x}} = \begin{bmatrix} \mathbb{I}_d & \boldsymbol{0}_{d\times(D-d)} \\ \dfrac{\partial \boldsymbol{y}_{d+1:D}}{\partial \boldsymbol{x}_{1:d}} & \operatorname{diag}\left(\exp\left(s_{\boldsymbol{\theta}_s}(\boldsymbol{x}_{1:d})\right)\right) \end{bmatrix}$$

行列式は次式のとおり計算可能である．

$$\det\frac{\partial f_{\boldsymbol{\theta}}(\boldsymbol{x})}{\partial \boldsymbol{x}} = \prod_{j=1}^{D-d} \exp\left(s_{\boldsymbol{\theta}_s}(\boldsymbol{x}_{1:d})\right)_j = \exp\left(\sum_{j=1}^{D-d} s_{\boldsymbol{\theta}_s}(\boldsymbol{x}_{1:d})_j\right)$$

これらにより，アフィンカップリング層は可逆性とヤコビ行列式の計算可能性を満たし，対数ゆう度を近似なしに計算することが可能である．

特徴　フローベースモデルについても品質，速度，多様性の三つの観点に着目して特徴を説明する．まず，品質に関しては，可逆ネットワークという限定されたネットワークしか使うことができないため，ネットワークの表現能力を上げることが難しく，限定的なところがある．一方で，速度に関しては，１回のフォワード計算で生成できるため速い．ただし，可逆ネットワークを用いないといけないという制約があるため，表現能力を上げるには多層のネットワークを用いることが必要で，速度向上のボトルネックになり得ることは留意されたい．多様性については，フローベースモデルでは対数ゆう度を近似なしで直接最適化することができるため，学習データの多様性をカバーすることが可能である．

また，ほかの特徴についても説明すると，フローベースモデルでは可逆制約により，観測変数 $\boldsymbol{x}$ と潜在変数 $\boldsymbol{z}$ の次元を同じにする必要があることが弱みである．

発展　画像のような階層的な構造を持つデータにフローベースモデルを適用しようとしたとき，マルチスケールな構造を持つモデルが有効である．Real NVP[4] では，多層のアフィンカップリング層からなるニューラルネットワークの中で，スクイーズ（Squeeze）処理と呼ばれる画像の高さ・幅方向の情報をチャネル方向に集約する処理（例えば，高さ H，幅 W，チャネル数 C のデータがあった場合，高さ $H/2$，幅 $W/2$，チャネル数 $4C$ のデータに変換する処理）を行うことで，マルチスケールなデータ表現を可能にしている．

Real NVP の発展モデルである Glow[13] では，マルチスケール構造に加えて，チャネル間の関係を記述可能な可逆な 1×1 の畳込み層を導入することで更に表現能力を向上し，生成画像の高精細化を実現している．

また，フローベースモデルでは様々な拡張モデルが提案されており，例えば，自己回帰モデルのようにゆう度を条件付き分布の積で表す自己回帰フロー（Autoregressive Flow，Inverse Autoregressive Flow（IAF)[14] など）や，ResNet の連続化したものを常微分方程式（ODE: Ordinary Differential Equation）として解く Neural ODE[15] のアイデアを生成モデルに拡張した Free-Form Jacobian of Reversible Dynamics（FFJORD)[16] などがある．

２．２．３　敵対的生成ネットワーク（GAN）

仕組み　変分自己符号化器やフローベースモデルでは，$\boldsymbol{z}$ から $\boldsymbol{x}$ を生成する生成器に対して，$\boldsymbol{x}$

から $\boldsymbol{z}$ を推論する推論器を用意することでゆう度を表現可能にし，最ゆう推定によってモデルを最適化できるようにしていた．これに対して敵対的生成ネットワーク[5]では，ゆう度は陽に表現せず，敵対的学習基準と呼ばれる新たな学習基準を用いて学習を行う．

具体的には，図４に示すように，敵対的生成ネットワークでは，$\boldsymbol{z}$ から $\boldsymbol{x}$ を生成する生成器 $G_{\boldsymbol{\theta}}$（$\boldsymbol{\theta}$ はモデルパラメータ）に加えて，与えられた画像が生成画像 $\boldsymbol{x}'$ か実画像 $\boldsymbol{x}$ かを識別する識別器 $D_{\boldsymbol{\phi}}$（$\boldsymbol{\theta}$ はモデルパラメータ）が用いられる．敵対的生成ネットワークの学習では，識別器 $D_{\boldsymbol{\phi}}$ は生成画像と実画像と識別を行うことで生成器 $G_{\boldsymbol{\theta}}$ になるべくだまされないように最適化され，一方で，生成器 $G_{\boldsymbol{\theta}}$ は識別器 $D_{\boldsymbol{\phi}}$ が生成画像と識別できない画像を生成することで，識別器 $D_{\boldsymbol{\phi}}$ をなるべくだませるように最適化される．こうして生成器 $G_{\boldsymbol{\theta}}$ と識別器 $D_{\boldsymbol{\phi}}$ を競合する条件下で最適化を行うことで相互に強化し合うことができ，最終的には，生成器 $G_{\boldsymbol{\theta}}$ は強力な識別器 $D_{\boldsymbol{\phi}}$ をもってしてでも識別できないような，リアリティのある画像を生成することが可能になる．

理論的背景　敵対的生成ネットワークの目的関数は以下の式で表される．

$$\mathbb{E}_{\boldsymbol{x}\sim p_{\mathcal{D}}(\boldsymbol{x})}\left[\log D_{\boldsymbol{\phi}}(\boldsymbol{x})\right] + \mathbb{E}_{\boldsymbol{z}\sim p(\boldsymbol{z})}\left[\log\left(1 - D_{\boldsymbol{\phi}}\left(G_{\boldsymbol{\theta}}(\boldsymbol{z})\right)\right)\right]$$

この式は，２クラス分類でよく用いられるバイナリークロスエントロピー損であり，第一項は実データ $\boldsymbol{x}$ の識別に関する項であり，第二項は生成データ $\boldsymbol{x}' = G_{\boldsymbol{\theta}}(\boldsymbol{z})$ の識別に関する項である．識別器 $D_{\boldsymbol{\phi}}$ は，この目的関数を最大化することによって，実データ $\boldsymbol{x}$ と生成データ $\boldsymbol{x}'$ の識別を行い，一方で，生成器 $G_{\boldsymbol{\theta}}$ はこの目的関数を最小化することによって，識別器 $D_{\boldsymbol{\phi}}$ が識別できないデータ $\boldsymbol{x}'$ を生成する．このように敵対的生成ネットワークの学習では，生成器 $G_{\boldsymbol{\theta}}$ と識別器 $D_{\boldsymbol{\phi}}$ で最適化の向きが異なる Min-Max 最適化が行われる．

敵対的生成ネットワークの理論的な最適解について見ると，まず，生成器 $G_{\boldsymbol{\theta}}$ を固定したとき，最適な識別器 $D_{\boldsymbol{\phi}}$（以下では，D_G^* と表す．）は以下の式で表される．

$$D_G^* = \frac{p_{\mathcal{D}}(\boldsymbol{x})}{p_{\mathcal{D}}(\boldsymbol{x}) + p_{\boldsymbol{\theta}}(\boldsymbol{x})}$$

ここで，$p_{\mathcal{D}}(\boldsymbol{x})$ は実データ集合 $\mathcal{D}$ の中のデータ $\boldsymbol{x}\in\mathcal{D}$ の分布を表し，$p_{\boldsymbol{\theta}}(\boldsymbol{x})$ は生成データ $\boldsymbol{x} = G_{\boldsymbol{\theta}}(\boldsymbol{z})$ の分布を表す．D_G^* を上述の敵対的生成ネットワークの目的関数に代入し，式を整理すると以下が導かれる．

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{x}\sim p_{\mathcal{D}}(\boldsymbol{x})}\left[\log D_G^*(\boldsymbol{x})\right] + \mathbb{E}_{\boldsymbol{z}\sim p(\boldsymbol{z})}\left[\log\left(1 - D_G^*\left(G_{\boldsymbol{\theta}}(\boldsymbol{z})\right)\right)\right] \\
&= \mathbb{E}_{\boldsymbol{x}\sim p_{\mathcal{D}}(\boldsymbol{x})}\left[\log\frac{p_{\mathcal{D}}(\boldsymbol{x})}{p_{\mathcal{D}}(\boldsymbol{x}) + p_{\boldsymbol{\theta}}(\boldsymbol{x})}\right] + \mathbb{E}_{\boldsymbol{x}\sim p_{\boldsymbol{\theta}}(\boldsymbol{x})}\left[\log\frac{p_{\boldsymbol{\theta}}(\boldsymbol{x})}{p_{\mathcal{D}}(\boldsymbol{x}) + p_{\boldsymbol{\theta}}(\boldsymbol{x})}\right] \\
&= -\log 4 + D_{KL}\left(p_{\mathcal{D}}(\boldsymbol{x})\middle\|\frac{p_{\mathcal{D}}(\boldsymbol{x}) + p_{\boldsymbol{\theta}}(\boldsymbol{x})}{2}\right) + D_{KL}\left(p_{\boldsymbol{\theta}}(\boldsymbol{x})\middle\|\frac{p_{\mathcal{D}}(\boldsymbol{x}) + p_{\boldsymbol{\theta}}(\boldsymbol{x})}{2}\right) \\
&= -\log 4 + 2\cdot D_{JS}(p_{\mathcal{D}}(\boldsymbol{x})\|p_{\boldsymbol{\theta}}(\boldsymbol{x}))
\end{aligned}$$

ここで，D_{JS} はジェンセン・シャノン情報量（Jensen-Shannon Divergence）を表す[*3]．この式から，敵対的生成ネットワークの生成器 $G_{\boldsymbol{\theta}}$ はジェンセン・シャノン情報量の基準において生成データの分布 $p_{\boldsymbol{\theta}}(\boldsymbol{x})$ を実データの分布 $p_{\mathcal{D}}(\boldsymbol{x})$ に近づけようとすることが分かる．

ただし，この理論では，学習データとモデルの表現能力が十分にあって理想的な識別器 D_G^* が得られることを仮定しているが，実際にはそのような条件を満たすことは難しく，Min-Max 最適化に由来する学習の不安定さが存在することに留意されたい．この学習の不安定さへの対処は，敵対的生成ネットワークの研究の中でも熱心に研究されている問題の一つであり，様々な改善策が提案されている．

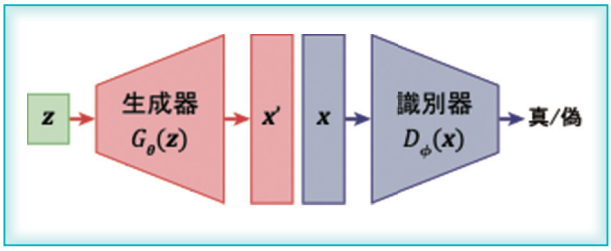


図４　敵対的生成ネットワーク（GAN）の構造

特徴　敵対的生成ネットワークについても品質，速度，多様性の三つの観点に着目して特徴を説明する．まず，品質に関しては，敵対的生成ネットワークでは，ゆう度 $p_{\boldsymbol{\theta}}(\boldsymbol{x})$ を陽に表現しないため分布形

＊３　二つの分布 P と Q に対するカルバックライブラ情報量を $D_{KL}(P\|Q)$，ジェンセン・シャノン情報量を $D_{JS}(P\|Q)$ とすると，$D_{JS}(P\|Q) = \frac{1}{2}D_{KL}\left(P\middle\|\frac{P+Q}{2}\right) + \frac{1}{2}D_{KL}\left(Q\middle\|\frac{P+Q}{2}\right)$ が成り立つ．これより，カルバックライブラ情報量は非対称性を持つが，ジェンセン・シャノン情報量は二つのカルバックライブラ情報量で平均化されて対称性を持つ．

状の近似誤差は生じず，更に，フローベースモデルと異なってネットワーク構造に対して大きな制約はないため，高品質な画像を生成することが可能である．また，速度に関しては，変分自己符号化器やフローベースモデルと同様に一回のフォワード計算で生成できるため高速である．一方で，多様性に関しては，敵対的生成ネットワークの学習で使われる敵対的学習基準は，あくまでも画像が実画像か生成画像かを判断する基準で，最ゆう推定とは異なり，実画像一枚一枚を再現しながら学習するような学習基準ではないため，多様性が欠如することがある．別の言い方をすると，一枚でも識別器をだませる画像が生成できれば，敵対的学習基準は満たされてしまう．この問題は，モード崩壊（Mode Collapse）と呼ばれ，敵対的生成ネットワークの研究分野で熱心に研究されている問題の一つである．

また，ほかの特徴についても説明すると，敵対的生成ネットワークは変分自己符号化器と同様に，観測変数 $\boldsymbol{x}$ と潜在変数 $\boldsymbol{z}$ の次元について制約を課す必要がないため，$\boldsymbol{x}$ と $\boldsymbol{z}$ で異なる次元数，例えば，$\boldsymbol{z}$ を低次元のコンパクトな変数に設定することが可能である．

発展　理論的背景で述べたように標準的な敵対的生成ネットワークは，ジェンセン・シャノン情報量に基づき生成データの分布 $p_{\boldsymbol{\theta}}(\boldsymbol{x})$ を実データの分布 $p_{\mathcal{D}}(\boldsymbol{x})$ に近づけようとする．しかし，ジェンセン・シャノン情報量は，二つの分布に重なりがないときに不連続かつ微分不可能な尺度であり，実データ分布全体をカバーできないことによって起こるモード崩壊の原因になり得る．そこで，敵対的生成ネットワークの研究分野では，分布間の尺度の改善が熱心に研究されている．例えば，Wasserstain GAN（WGAN）[17] では，緩い制約のもと連続・微分可能という良い性質を持つワッサーステイン距離（Wasserstein Distance）を用いることが提案されている．また，Least Squares GAN（LSGAN）[18] では，$p_{\boldsymbol{\theta}}(\boldsymbol{x})+p_{\mathcal{D}}(\boldsymbol{x})$ と $2p_{\boldsymbol{\theta}}(\boldsymbol{x})$ のカイ二乗距離（Chi-Square Distance）を考えることで，生成データと実データがなるべく識別境界周辺に近づくようにし，データを網羅しやすくすることが提案されている．

また，ネットワーク構造を改善することによって，生成画像の品質を改善しようという研究も数多く存在する．その代表的なモデルが StyleGAN[19] である．従来の GAN では，ネットワークの最初の層に入力される潜在変数のみを用いて画像を表現していた．これに対し，StyleGAN では，ネットワークの最初の層に入力されて画像の内容（Content）を表現する潜在変数に加えて，ネットワークの複数の中間層に多段階で挿入されて画像のスタイル（Style）を表現する潜在変数を用いることで，各潜在変数で分担して画像を表現することを可能にし，生成画像の高精細化並びに制御性の向上を実現している．

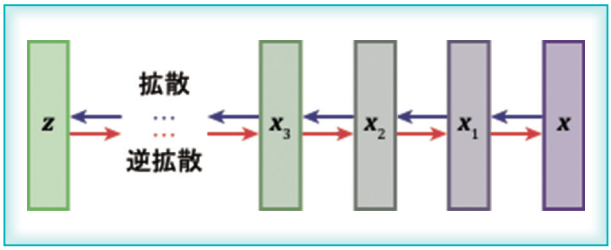


図 5　拡散モデル（Diffusion Model）の構造

2.2.4　拡散モデル（Diffusion Model）

仕組み　拡散モデル[6)~(8)] では，変分自己符号化器やフローベースモデルと同様に，画像生成の過程 $\boldsymbol{z}\rightarrow\boldsymbol{x}$ を学習する際に，潜在変数の推論の過程 $\boldsymbol{x}\rightarrow\boldsymbol{z}$ も同時に考える．しかし，変分自己符号化器やフローベースモデルでは，画像生成も潜在変数の推論も一回のフォワード計算で実現していたのに対し，拡散モデルでは，図 5 に示すように，繰り返し計算によって画像生成と潜在変数の推論を実現する．

具体的には，変換のステップ数を T，ステップ $t\in\{0,\ldots,T\}$ におけるデータを $\boldsymbol{x}_t$（すなわち，$\boldsymbol{x}=\boldsymbol{x}_0$，$\boldsymbol{z}=\boldsymbol{x}_T$）とすると，$\boldsymbol{x}$ から $\boldsymbol{z}$ への推論のプロセス（拡散過程）は，$\boldsymbol{x}_0\rightarrow\boldsymbol{x}_1\rightarrow\cdots\rightarrow\boldsymbol{x}_T$ と表され，$\boldsymbol{z}$ から $\boldsymbol{x}$ への生成のプロセス（逆拡散過程）は，$\boldsymbol{x}_T\rightarrow\boldsymbol{x}_{T-1}\rightarrow\cdots\rightarrow\boldsymbol{x}_0$ と表される．拡散過程は雑音（ノイズ）を加えていく過程であり，一般的には学習を簡便化するため，あらかじめ定めたルールに従ってサンプリングされた雑音をデータに加算していくことで実現される．一方，逆拡散過程は，雑音を除去していく過程である．拡散過程では，入力データに関係なく雑音を付与すればよかったのに対し，逆拡散過程では入力データに応じて雑音を推定して除去する必要があるため解くのは簡単ではない．そこで，パラメタライズされたモデル（具体的には，ニューラルネットワーク）を用意し，学習によって最適化を行う．学習の詳細については，次の理論的背景で説明する．

理論的背景　拡散モデルの拡散過程及び逆拡散過程について数式を交えながら詳述する．なお，拡散モデルの定式化の方法としては様々なものが提案されているが，本論文では，それらの中でも代表的

な定式化の一つであるデノイジング拡散確率モデル（DDPM: Denoising Diffusion Probabilistic Model）[8]について説明を行う．まず，拡散プロセス$\boldsymbol{x} \to \boldsymbol{z}$においては，マルコフ過程が仮定される．すなわち，あるステップの状態$\boldsymbol{x}_t$は前のステップの状態$\boldsymbol{x}_{t-1}$のみで決定される．このとき，1ステップの拡散過程は以下の式で表される．

$$q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) = \mathcal{N}\left(\boldsymbol{x}_t; \sqrt{\alpha_t}\boldsymbol{x}_{t-1}, \beta_t\mathbf{I}\right)$$

ここで，$\alpha_t = 1-\beta_t$であり，右辺は平均$\sqrt{\alpha_t}\boldsymbol{x}_{t-1}$，分散$\beta_t\mathbf{I}$の正規分布を表す．正規分布の再生性（正規分布$\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\sigma}_1^2)$とそれと独立な正規分布$\mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\sigma}_2^2)$を足し合わせると正規分布$\mathcal{N}(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2, \boldsymbol{\sigma}_1^2+\boldsymbol{\sigma}_2^2)$になる）により上式は下式に置き換えることができる．

$$q(\boldsymbol{x}_t|\boldsymbol{x}_0) = \mathcal{N}\left(\boldsymbol{x}_t; \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0, (1-\bar{\alpha}_t)\mathbf{I}\right)$$

ここで，$\bar{\alpha}_t = \prod_{i=1}^{t}\alpha_t$である．リパラメタライゼーショントリック（Reparameterization Trick）[1]により，$\boldsymbol{x}_t$はノイズ$\boldsymbol{\epsilon}\sim\mathcal{N}(\mathbf{0},\mathbf{I})$を用いて下式で表すことができる．

$$\boldsymbol{x}_t = \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}$$

この式から，拡散過程は徐々にノイズを加えていく過程であるが，$\boldsymbol{x}_t$は元のデータ$\boldsymbol{x}=\boldsymbol{x}_0$にノイズを一回加算するだけで求められることが分かる．

続いて，逆拡散過程について説明すると，1ステップの逆拡散過程は，パラメータ$\boldsymbol{\theta}$で特徴付けられたモデルを用いて以下で表される．

$$p_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t) = \mathcal{N}\left(\boldsymbol{x}_{t-1}; \boldsymbol{\mu}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t), \boldsymbol{\Sigma}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)\right)$$

計算の簡便化のため，デノイジング拡散確率モデル[8]では，$\boldsymbol{\Sigma}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)$をステップ依存の定数$\sigma_t^2\mathbf{I}$とする．より具体的には，$\sigma_t^2$は$\beta_t$または，$\tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$と置かれる．また，拡散過程で与えられたノイズ$\boldsymbol{\epsilon}$を$\boldsymbol{x}_t$と$t$から推定するノイズ推定器を$\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)$とすると，$\boldsymbol{x}_t = \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}$の関係式により，$\boldsymbol{x}_t$をデノイズすることによって得られるデータ$\tilde{\boldsymbol{x}}_0$は下式で表される．

$$\tilde{\boldsymbol{x}}_0 = \frac{1}{\sqrt{\bar{\alpha}_t}}\left(\boldsymbol{x}_t - \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)\right)$$

これにより，$\boldsymbol{\mu}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)$は，$\boldsymbol{x}_{t-1}$を$\boldsymbol{x}_t$と$\tilde{\boldsymbol{x}}_0$から求めたときの平均$\tilde{\mu}_t(\boldsymbol{x}_t, \tilde{\boldsymbol{x}}_0) = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\tilde{\boldsymbol{x}}_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\boldsymbol{x}_t$に等しいとすると，次の式が得られる．

$$\begin{aligned}
\boldsymbol{\mu}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t) &= \tilde{\mu}_t(\boldsymbol{x}_t, \tilde{\boldsymbol{x}}_0) \\
&= \tilde{\mu}_t\left(\boldsymbol{x}_t, \frac{1}{\sqrt{\bar{\alpha}_t}}\left(\boldsymbol{x}_t - \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)\right)\right) \\
&= \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)\right)
\end{aligned}$$

最終式の括弧内は$\boldsymbol{x}_t$をデノイズする処理を表し，式全体では，デノイズした結果を更に$\frac{1}{\sqrt{\alpha_t}}$でスケールした結果を表す．これらの式とリパラメタライゼーショントリックにより，逆拡散過程において$\boldsymbol{x}_{t-1}$はノイズ$\tilde{\boldsymbol{\epsilon}}\sim\mathcal{N}(\mathbf{0},\mathbf{I})$を用いて下式で求めることができる．

$$\boldsymbol{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)\right) + \sigma_t\tilde{\boldsymbol{\epsilon}}$$

つまり，逆拡散過程は，デノイズ（右辺第一項）とノイズ加算（右辺第二項）を繰り返し行うことで，徐々に$\boldsymbol{x}_0$に近づいていく過程と見ることができる*4．

最後に，拡散モデルの目的関数について説明する．変分自己符号化器やフローベースモデルと同様に，拡散モデルでも最ゆう推定が行われる．拡散モデルの各拡散・逆拡散ステップに対して，変分自己符号化器と同様の変分近似を行うと，負の対数ゆう度$-\log p_{\boldsymbol{\theta}}(\boldsymbol{x}_0)$の変分上界は以下のようになる．

$$\begin{aligned}
&-\log p_{\boldsymbol{\theta}}(\boldsymbol{x}) \geq \mathbb{E}_{q(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)}\left[-\log\frac{p_{\boldsymbol{\theta}}(\boldsymbol{x}_{0:T})}{q(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)}\right] \\
&= \mathbb{E}_q[D_{KL}(q(\boldsymbol{x}_T|\boldsymbol{x}_0)\|p_{\boldsymbol{\theta}}(\boldsymbol{x}_T)) + \\
&\sum_{t>1} D_{KL}(q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0)\|p_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)) - \log p_{\boldsymbol{\theta}}(\boldsymbol{x}_0|\boldsymbol{x}_1)]
\end{aligned}$$

ここで，第二式の第一項については，上述のようにデノイジング拡散確率モデルでは事後分布qは学習可能なパラメータを持たないと仮定しているため無視することができる．第二項，第三項については，拡散過程で与えるノイズ$\boldsymbol{\epsilon}\sim\mathcal{N}(\mathbf{0},\mathbf{I})$，ステップ$t$におけるデータ$\boldsymbol{x}_t = \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}$，及び，逆拡散過程で使われるノイズ推定器$\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)$を用いて書き換えると以下のようになる．

$$\mathbb{E}_{\boldsymbol{\epsilon}\sim\mathcal{N}(\mathbf{0},\mathbf{I})}\left[\frac{\beta_t^2}{2\sigma_t^2\alpha_t(1-\bar{\alpha}_t)}\left\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_{\boldsymbol{\theta}}\left(\sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t\right)\right\|_2^2\right]$$

＊4　本稿では，深層生成モデルを$\boldsymbol{z}$から$\boldsymbol{x}$の写像を学習するものと定義し，$\boldsymbol{z}$に対応する$\boldsymbol{x}$は一意に定まるものとして説明してきたが，上式のようにDDPMでは，逆拡散を行うたびにノイズを加算するため，$\boldsymbol{z}$に対応する$\boldsymbol{x}$が必ずしも一意に定まるとは限らない．この点を考慮したモデルとしては，Denoising Diffusion Implicit Model（DDIM）[20]があり，$\boldsymbol{z}$から決定論的に$\boldsymbol{x}$を生成できるように改良が加えられている．

デノイジング拡散確率モデル[8]では，上式における重み付け項を無視した以下の目的関数を用いて最適化が行われる．

$$\mathbb{E}_{\epsilon\sim\mathcal{N}(\mathbf{0},\mathbf{I})}\left[\left\|\epsilon-\epsilon_\theta\left(\sqrt{\bar{\alpha}_t}x_0+\sqrt{1-\bar{\alpha}_t}\epsilon,t\right)\right\|_2^2\right]$$

この変形により，デノイジング拡散確率モデルでは，各拡散・逆拡散ステップにおいて異なる重み付けがされた変分近似がされることになる．直感的な説明をすると，この重み付けはノイズの量が少ないときは目的関数を小さくし，ノイズの量が多いときは目的関数を大きくする効果があり，これにより，モデルはより難しいデノイズタスクに着目して学習することが可能になる．また，上式には最終的な出力であるx_0は表れず，逆拡散過程の部分問題である各拡散ステップのデノイズしか表れないことに留意されたい．つまり，拡散モデルでは，一般的に複雑で困難な生成過程の学習を，より簡単な部分問題を解くことで実現することが可能であり，これは拡散モデルの大きな強みの一つになっている．

特徴　拡散モデルについても品質，速度，多様性の三つの観点に着目して特徴を説明する．まず，品質に関しては，拡散モデルでモデル化するのはあくまでもデノイズであり，最終的な出力であるxの分布$p(x)$に対して近似を行う必要がないため近似誤差が生じないのに加えて，拡散モデルでは同一のニューラルネットワークを繰り返し適用することで飛躍的に表現能力を増大させることができるため，高品質なデータを生成可能である．一方で，速度に関しては，ほかの深層生成モデルが一回のフォワード計算で生成できたのに対し，拡散モデルでは繰り返し計算が必要になるため遅い．多様性に関しては，変分自己符号化器やフローベースモデルと同様に最ゆう推定に基づく最適化を行うため，学習データの多様性をカバーすることが可能である．

また，ほかの特徴についても説明すると，拡散モデルではデータサイズを維持したまま拡散・逆拡散を行う必要があるため，観測変数xと潜在変数zの次元を同じにする必要があることが弱みである．

発展　拡散モデルで使われるノイズ推定器では，データxと同じサイズのデータを入出力するニューラルネットワークを使う必要があるため，解像度の高いデータを扱おうとすると，解像度に応じて計算時間が飛躍的に増加するという問題があった．特に，拡散モデルでは繰り返し計算が必要でこの問題は深刻であった．この問題に対処するため提案されたのが，Latent Diffusion Model（LDM，別名 Stable Diffusion）[20]である．LDM では，データをまず事前学習した自己符号化器を用いて低次元の潜在変数に変換した後，低次元の潜在変数に対して拡散モデルを適用するように構成することで処理の高速化を実現している．

また，計算量が解像度に応じて増加する問題に別のアプローチで取り組んだものとして Imagen[21]がある．Imagen では，高解像度のデータを表現する拡散モデルをいきなり用いるのではなく，低解像度のデータを表現する拡散モデルを使用した後，超解像を行う拡散モデルを適用するようにすることで，学習の容易化と処理の高速化を実現している．

3 深層生成モデルの応用

深層生成モデルの応用は多岐にわたり，例えば，データ変換，データ編集，表現獲得，異常検知，欠損値補完などの生成以外のタスクへの応用や，画像以外のデータ，例えば，音声やテキストなどへの応用などがある．本章では，誌面の都合により，様々な応用の中でも画像，音声，テキストなどのデータの種類によらず有効なアイデアで，また，データ変換やデータ編集などの様々な応用タスクにおいてベースアイデアとなる条件付き生成モデルへの拡張に着目して説明を行う．

3.1　条件付き生成モデルへの拡張

2. では生成器の入力として潜在変数zのみを用いる条件なし生成モデルについて説明してきた．一般的に，潜在変数 z の中には様々な表現が複雑に絡まり合った状態で存在しており，zを直接操作することによって，出力画像xをコントロールすることは簡単ではない．このような問題に対処するために提案されたのが条件付き生成モデルである．条件付き生成モデルでは，zに加えて条件情報yを生成器の入力に加えることによって，yによって生成画像をコントロールすることを可能にする．ここで，条件情報とは，クラスラベルやテキスト，画像などのデータ生成を制御する上で有用な補助情報のことである。例えば，yとしてクラスラベル（例えば，動植物のカテゴリーなど）を用いた場合，そのクラスに合った画像を生成することが可能になる．同様に，yとしてテキストを用いた場合は，テキストから画像への変換（Text-to-Image）が可能になり，yとして画像を用いた場合は，画像から画像への変換（Image-to-Image）が可能になる．

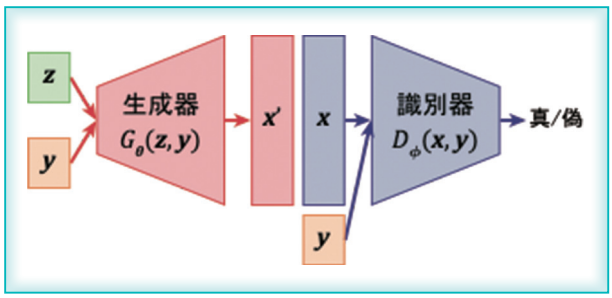


図６　条件付き敵対的生成ネットワークの構造

このように，条件付き生成モデルは画像生成の制御性を向上させる上で重要な技術となっている．以下では，誌面の都合により，深層生成モデルの中でも特に広く使われている敵対的生成ネットワーク（３．２）と拡散モデル（３．３）に着目し，条件付き生成モデルへの拡張について具体的に説明する．なお，変分自己符号化器，フローベースモデルを含むいずれの深層生成モデルも，条件付き生成モデルへの拡張において重要なことは，モデル化対象を$p(\boldsymbol{x},\boldsymbol{z})$から$p(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$に拡張することである．そのため，後述の敵対的生成ネットワークや拡散モデルに関する例と同様に，基本的には，条件情報 y をうまくモデルに組み込むことによって条件付き生成モデルに拡張できることに留意されたい．

３．２　敵対的生成ネットワークの条件付き拡張モデル

敵対的生成ネットワークの代表的な条件付き拡張モデルとしては，条件付き敵対的生成ネットワーク（cGAN: Conditional GAN）[22] と補助分類器付き敵対的生成ネットワーク（AC-GAN: Auxiliary Classifier GAN）[23] がある．本節ではこれらを順に説明する．

３．２．１　条件付き敵対的生成ネットワーク（cGAN）

条件付き敵対的生成ネットワーク[22] では，図６に示すように，生成器 G_θ と識別器 D_ϕ の入力に条件情報 $\boldsymbol{y}$ を与える．この変更に伴い，目的関数は下式に拡張される．

$$\mathbb{E}_{(x,y)\sim p_{\mathcal{D}}(x,y)}\left[\log D_\phi(\boldsymbol{x},\boldsymbol{y})\right]$$
$$+\,\mathbb{E}_{z\sim p(z),y\sim p(y)}\left[\log\left(1-D_\phi(G_\theta(\boldsymbol{z},\boldsymbol{y}),\boldsymbol{y})\right)\right]$$

この拡張により，識別器 D_ϕ は $\boldsymbol{y}$ に基づいて画像を識別，すなわち，条件情報 $\boldsymbol{y}$ を考慮した上で実画像か生成画像かの識別を行う．一方で，生成器 G_θ は $\boldsymbol{y}$ に基づいて生成した画像が $\boldsymbol{y}$ を考慮した上で実画像に識別されるように最適化される．これにより，生成器 G_θ は $\boldsymbol{y}$ を考慮した上でリアリティのある画像を生成することが可能になる．

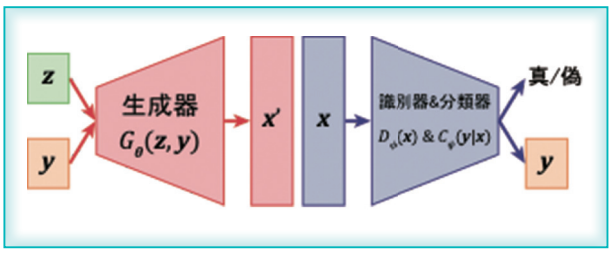


図７　補助分類器付き敵対的生成ネットワークの構造

３．２．２　補助分類器付き敵対的生成ネットワーク（AC-GAN）

補助分類器付き敵対的生成ネットワーク[23] の構造を図７に示す．この図に示すように，補助分類器付き敵対的生成ネットワークの生成器 G_θ は条件付き敵対的生成ネットワークの生成器 G_θ と同じで，入力に条件情報 $\boldsymbol{y}$ が与えられる．一方，識別器については，条件なし敵対的生成ネットワークの識別器 D_ϕ に加えて，与えられた画像 $\boldsymbol{x}$ のクラスを分類する補助分類器 $C_\psi(\boldsymbol{y}|\boldsymbol{x})$（$\boldsymbol{\psi}$ はモデルパラメータ）が合わせて使われる．この変更に伴い，目的関数は以下に拡張される．

$$\mathbb{E}_{x\sim p_{\mathcal{D}}(x)}\left[\log D_\phi(\boldsymbol{x})\right]$$
$$+\,\mathbb{E}_{z\sim p(z),y\sim p(y)}\left[\log\left(1-D_\phi(G_\theta(\boldsymbol{z},\boldsymbol{y}))\right)\right]$$
$$\mathbb{E}_{(x,y)\sim p_{\mathcal{D}}(x,y)}\left[-\log C_\psi(\boldsymbol{y}|\boldsymbol{x})\right]$$
$$\mathbb{E}_{z\sim p(z),y\sim p(y)}\left[-\log C_\psi(\boldsymbol{y}|G_\theta(\boldsymbol{z},\boldsymbol{y}))\right]$$

第一式は，生成器が $G_\theta(\boldsymbol{z})$ から $G_\theta(\boldsymbol{z},\boldsymbol{y})$ に変更されただけで，ほかの部分は条件なし敵対的生成ネットワークと同じである．第二式，第三式が新たに導入された目的関数で，まず，第二式では，実画像を用いて補助分類器 C のクラス識別が学習される．一方，第三式では，補助分類器 C_ψ は固定したまま生成器 G_θ の学習が行われる．具体的には，生成器 G_θ が $\boldsymbol{y}$ に基づき生成した画像が，補助分類器 C_ψ に適切に分類されるように生成器 G_θ の学習が行われる．

このように，条件付き敵対的生成ネットワークと補助分類器敵対的生成ネットワークでは，識別器において，$\boldsymbol{y}$ を入力で使うか出力で使うかという違いがある．

３．３　拡散モデルの条件付き拡張モデル

拡散モデルの代表的な条件付き拡張モデルとしては，分類器ガイダンス（Classifier Guidance, CG）[24] と分類器なしガイダンス（Classifier-Free Guidance, CFG）[25] の二つがある．本節ではこれらを順に説明する．

各々の詳細について説明する前に，前提知識とし

て，ノイズ推定（**2．2．4**で説明した$\epsilon_\theta(x_t,t)$）とスコア推定の関係について説明する．なお，スコアとは対数ゆう度$\log q(x)$の入力xに関する勾配$\nabla_x \log q(x)$である．ノイズ推定とスコア推定の間には下式が成り立つ[26]。

$$\epsilon_\theta(x_t,t) = -\sqrt{1-\bar{\alpha}_t}\nabla_{x_t}\log q(x_t)$$

上記関係式の直感的な解釈を説明すると，左辺は画像を良くするために除外が必要なノイズの量を意味し，一方，右辺は画像を良くするために必要な更新ステップの大きさを意味する．つまり，左辺と右辺は画像を良くするということを異なる表現をしたものと言える．また，上式の左辺と右辺をそれぞれ条件付き設定に拡張すると以下が得られる．

$$\epsilon_\theta(x_t,t,y) = -\sqrt{1-\bar{\alpha}_t}\nabla_{x_t}\log q(x_t,y)$$

3．3．1　分類器ガイダンス（CG）

分類器ガイダンス[24]では，条件なしノイズ推定$\epsilon_\theta(x_t,t)$が得られている状況下で，条件付き生成ができるようにすることを考える．この動機の下，上述の条件付きスコア推定$\nabla_{x_t}\log q(x_t|y)$をベイズの定理に基づき変形する．

$$\nabla_{x_t}\log q(x_t|y) = \nabla_{x_t}\log q(x_t) + \nabla_{x_t}\log q(y|x_t)$$

右辺の第一項に関しては，ノイズ推定とスコア推定の関係式より，条件なしノイズ推定$\epsilon_\theta(x_t,t)$から算出することができる．一方，右辺第二項の$q(y|x_t)$については，分類器ガイダンスではクラス分類器${C_\psi}'(y|x_t)$（ψはモデルパラメータ）を用いて表す．更に，この項に関しては，ガイダンススケール$s>0$を導入することでクラスに対する忠実性をコントロールできるようにする．ガイダンススケール付きの条件付きスコアを$\log q_s(x_t|y)$と置くと，上式は下式に置き換えられる．

$$\nabla_{x_t}\log q_s(x_t|y) = \nabla_{x_t}\log q(x_t) + s\cdot\nabla_{x_t}\log {C_\psi}'(y|x_t)$$

分類器ガイダンスの課題は二つあり，一つは，クラス分類器${C_\psi}'(y|x_t)$は，クリーンな画像x_0だけではなく様々なノイズレベルの画像x_tに対応できる必要があること，もう一つは，yとx_tの関連度が低い場合，${C_\psi}'(y|x_t)$のx_tに対する勾配が不安定になることである．

3．3．2　分類器なしガイダンス（CFG）

上記の分類器にまつわる二つの課題を回避するため提案されたのが，分類器なしガイダンス[25]である．分類器なしガイダンスでは，ベイズの定理に基づき，ガイダンススケール付きの条件付きスコア$\log q_s(x_t|y)$を以下のように変形する．

$$\nabla_{x_t}\log q_s(x_t|y) = \nabla_{x_t}\log q(x_t) + s\cdot\nabla_{x_t}\log q(y|x_t)$$
$$= \nabla_{x_t}\log q(x_t) + s\cdot\left(\nabla_{x_t}\log q(x_t|y) - \nabla_{x_t}\log q(x_t)\right)$$

分類器ガイダンスでは，分類器${C_\psi}'(y|x_t)$によって，生成過程を誘導していたが，分類器なしガイダンスでは，条件付きスコアと条件なしスコアの差分（第二式の第二項）に基づき生成過程を誘導している．重要なことは，第二式では分類器${C_\psi}'(y|x_t)$は使われないため，上述の分類器ガイダンスの分類器にまつわる二つの課題を回避することができることである．また，分類器なしガイダンスでは，条件付きスコア$\nabla_{x_t}\log q(x_t|y)$と条件なしスコア$\nabla_x \log q(x)$の二つのスコアを計算する必要があるが，条件なしスコアについては，条件付きスコアのyに無情報$\emptyset$を代入することで表現することができる．これにより、一つの条件付きノイズ推定$\epsilon_\theta(x_t,t,y)$を用いて両者を表現することが可能である．

4　おわりに

本論文では，深層生成モデルの基礎として，認識と生成の対比から説明を始め，続いて，代表的な深層生成モデルの仕組み，理論的背景，特徴，及び発展について説明した．また，深層生成モデルの応用では，代表的な応用の一つとして条件付き生成モデルへの拡張について紹介した．本論文では，誌面の都合で筆者の主観により研究事例をピックアップして紹介してきたが，深層生成モデルの研究は紹介したものにとどまらず，ほかにも数多くの興味深い研究があることに留意されたい．例えば，本論文では，主に画像生成に着目して説明してきたが，音声，音楽，テキストなどのほかのデータの生成に関する研究も活発に行われている．筆者らも深層生成モデルを用いた画像生成にとどまらず，音声合成・変換の研究も活発に行っている．具体的な研究事例は筆者の Web サイト[*5]で紹介しているので，興味のある方は参照されたい．

また，深層生成モデルの今後を語る上で無視できない問題は，生成 AI の普及によるメリットを享受しながら，ディープフェイクや著作権侵害などの生成 AI の悪用のリスクをいかにして軽減するかということである．これに関しては，研究面では，ディープフェイクの検出や生成画像用の電子透かしの研究が活発に行われている一方，法規制の議論・整備も進みつつあるところであり，今後も注視して

いくことが重要であると考える．

最後になるが，深層生成モデルの応用範囲は飛躍的に広がっており，読者の皆様も何らかの形で使用する機会が増えていくと予想される．そのようなときに本論文が読者の一助となることを心から願い，本論文の結びとする．

■ 文献

(1) D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” ICLR (International Conf. Learning Representations), 2014.
(2) D. J. Rezende, S. Mohamed, and D. Wierstra, “Stochastic backpropagation and approximate inference in deep generative models,” ICML (International Conf. Mach. Learning), 2014.
(3) L. Dinh, D. Krueger, and Y. Bengio, “NICE: Nonlinear independent components estimation,” ICLR Workshop, 2015.
(4) L. Dinh, J. Sohl-Dickstein, and S. Bengio, “Density estimation using real NVP,” ICLR, 2017.
(5) I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” NIPS (Neural Inf. Process. Syst.), 2014.
(6) J. Sohl-Dickstein, E. A. Weiss, N. Maheswaranathan, and S. Ganguli, “Deep unsupervised learning using nonequilibrium thermodynamics,” ICML, 2015.
(7) Y. Song and S. Ermon, “Generative modeling by estimating gradients of the data distribution,” NeurIPS (Neural Inf. Process. Syst.), 2019.
(8) J. Ho, A. Jain, and P. Abbeel, “Denoising diffusion probabilistic models,” In NeurIPS, 2020.
(9) Y. LeCun, S. Chopra, R. Hadsell, M. Ranzato, and F. J. Huang, “A tutorial on energy-based learning, “Predicting Structured Data,” MIT Press, 2006.
(10) A. van den Oord, N. Kalchbrenner, O. Vinyals, L. Espeholt, A. Graves, and K. Kavukcuoglu, “Conditional image generation with PixelCNN decoders,” In NIPS, 2016.
(11) A. van den Oord, O. Vinyals, and K. Kavukcuoglu, “Neural discrete representation learning,” In NIPS, 2017.
(12) A. Razavi, A. van den Oord, and O. Vinyals, “Generating diverse high-fidelity images with VQ-VAE-2,” In NeurIPS, 2019.
(13) D. P. Kingma and P. Dhariwal, “Glow: Generative flow with invertible 1x1 convolutions,” In NeurIPS, 2018.
(14) D. P. Kingma, T. Salimans, R. Jozefowicz, X. Chen, I. Sutskever, and M. Welling, “Improved variational inference with inverse autoregressive flow,” In NIPS, 2016.
(15) R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud, “Neural ordinary differential equations,” In NeurIPS, 2018.
(16) W. Grathwohl, R. T. Q. Chen, J. Bettencourt, I. Sutskever, and D. Duvenaud, “FFJORD: Free-form continuous dynamics for scalable reversible generative models,” In ICLR, 2019.
(17) M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein generative adversarial network,” In ICML, 2017.
(18) X. Mao, Q. Li, H. Xie, R. Y.K. Lau, Z. Wang, and S. P. Smolley, “Least squares generative adversarial networks,” In ICCV (IEEE International Conf. Comput. Vision), 2017.
(19) T. Karras, S. Laine, and T. Aila, “A style-based generator architecture for generative adversarial networks,” In CVPR (Comput. Vision and Pattern Recognition), 2019.
(20) R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Omme, “High-resolution image synthesis with latent diffusion models,” In CVPR, 2022.
(21) C. Saharia, W. Chan, S. Saxena, L. Li, J. Whang, E. Denton, S. K. S. Ghasemipour, B. K. Ayan, S. S. Mahdavi, R. G. Lopes, T. Salimans, J. Ho, D. J. Fleet, and M. Norouzi, “Photorealistic text-to-image diffusion models with deep language understanding,” In NeurIPS, 2022.
(22) M. Mirza and S. Osindero, “Conditional generative adversarial nets,” arXiv preprint arXiv:1411.1784, 2014.
(23) A. Odena, C. Olah, and J. Shlens, “Conditional image synthesis with auxiliary classifier GANs,” In ICML, 2017.
(24) P. Dhariwal and A. Nichol, “Diffusion models beat GANs on image synthesis,” In NeurIPS, 2021.
(25) J. Ho and T. Salimans, “Classifier-free diffusion guidance,” In NeurIPS Workshop, 2021.
(26) P. Vincent, “A connection between score matching and denoising autoencoders,” Neural Computation, 2011.

(2024 年 6 月 3 日受付，9 月 2 日再受付)

金子卓弘 (正員)

平 24 東大・工・機械情報卒．平 26 同大学院修士課程了．同年日本電信電話株式会社入社．令 2 東大大学院博士課程了．令 2 から NTT コミュニケーション科学基礎研究所特別研究員．画像生成，音声合成・変換を中心としたコンピュータビジョン，信号処理，機械学習の研究に従事．博士（情報理工学）．日本機械学会畠山賞，ICPR Best Student Paper Award，音声研究会研究奨励賞，東大大学院研究科長賞，電気通信普及財団テレコムシステム技術賞，日本音響学会粟屋潔学術奨励賞等を各受賞．

＊5 https://www.kecl.ntt.co.jp/people/kaneko.takuhiro/

グラフニューラルネットワークの最新動向

佐藤竜馬　Ryoma Sato　国立情報学研究所

1 はじめに

グラフニューラルネットワーク（GNN）とはグラフデータを処理するニューラルネットワークである．画像・音声・テキストデータにおいてニューラルネットワークが機械学習モデルの事実上の標準になったように，グラフデータにおいても，グラフニューラルネットワークが標準的な機械学習モデルとなっている．グラフニューラルネットワークの原型は，1990 年代から 2000 年代中頃に登場した [(1), (2)]．当時はカーネル法が主流であり，グラフニューラルネットワークは標準的な選択とはならなかったが，2017 年頃に深層学習と結び付いて急速に普及が進んだ [(3)]．その勢いは現在も止まることなく，機械学習系の国際会議 ICLR 2024 ではグラフニューラルネットワーク関連の論文が 170 本発表された．ICLR 2024 の全論文のうち 7.4 %がグラフニューラルネットワーク関連ということになる．本稿では ICLR 2024 で発表されたグラフニューラルネットワークの最新動向について解説する．

2 概要

グラフニューラルネットワークの研究トピックは大きく分けて二種類ある．第 1 は，解釈性，頑健性，高速化，汎化性能の解析など，通常のニューラルネットワークにもあるトピックである．通常のニューラルネットワークにおける結果を，グラフニューラルネットワークに拡張することが基本的な方針であるが，入力データの離散性など，グラフ特有の課題があり，それらの課題を克服することが主な焦点となる．第 2 は，同変性，過平滑化など，グラフニューラルネットワーク特有のトピックである．これらの分析は群論やスペクトルグラフ理論など，グラフ特有の理論を本質的に扱うことが多い．図 1 に ICLR 2024 で発表されたグラフニューラルネットワーク関連のトピックの分布を掲載する．以下では代表的なトピックについて詳細に解説する．

3 解釈性

ほかの深層学習手法と同様，グラフニューラルネットワークでも解釈や信頼性の重要性が叫ばれている．このトピックの嚆矢は GNNExplainer [(4)] であり，そこから 5 年ほど精力的な研究が続いている．グラフデータにおいては，頂点特徴量ベクトルのうちどの次元が重要かなど，通常のニューラルネットワークにおいても研究されている問題設定のほか，どの頂点やどの辺が重要かなど，グラフ特有の問題設定，そしてそれらを組み合わせた問題設定もある．これらの問題設定は特徴ベクトル以外に様々な要素が存在する点で，通常のニューラルネットワークの解釈性よりも一般的な問題設定であり，一段と難しい問題である．通常のニューラルネットワークの解釈性が完璧に解かれていないのと同様に，グラフニューラルネットワークでも完璧な解決はまだ遠く，様々なアプローチの模索が続いている段階である．

ICLR 2024 では，様々なグラフニューラルネットワークの解釈性手法の特性や，性能をベンチマークする GNNX-BENCH という研究がマート・コサンら [(5)] により発表された．この論文の結論はそれなりに良い解釈手法もあるが，安定性などまだまだ課題はある，というものである．この論文では反実仮想説明を詳しく取り扱っている．反実仮想説明とは，入力例をどのように変化させたら別の予測がされるかを基に解釈を行うというアプローチであり，グラフニューラルネットワークに限らず一般の機械学習モデルの解釈方法として人気である．ローンの審査において「あなたは落ちてしまったけれど，年収があと 100 万円高ければ，当社のモデルは合格と判断していました」というような説明である．グラフニューラルネットワークにおいても「この化合物は毒

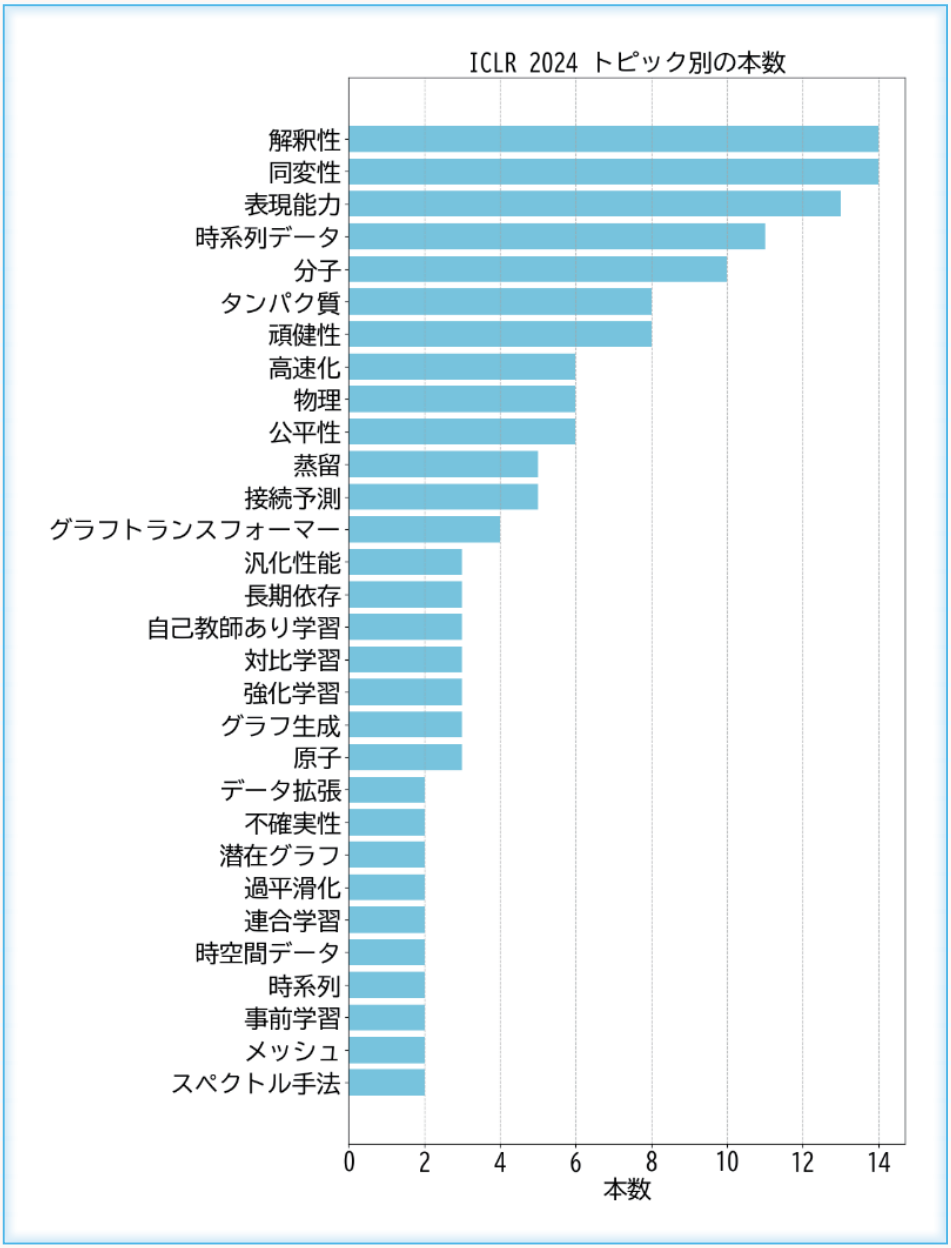


図 1　ICLR 2024 で発表されたグラフニューラルネットワーク関連のトピックの分布

があると判断されてしまったけれど，この官能基が付いていれば毒がないと予測していました」というような反実仮想説明が用いられる．そういった説明を生成する手法は幾つかあるが，現実的ではない仮想データを生成してしまうことが，この論文で指摘されている．非連結なグラフを生成してしまったり，炭素原子に 10 個の水素原子が結合していたりして，化合物としての体裁を保っていないような仮想データが生成されてしまう．「もしも○○だったらなぁ」の○○が突飛過ぎて現実的な解釈としては役に立たないというわけだ．このような問題意識はこの論文以前からあり，生成モデルを用いてもっともらしいグラフだけを使う [6], [7] といったことも提案されているが，それでも不合理なグラフを生成してしまうことはある．生成モデルも完璧ではないので，生成モデルの上で説明を探索すると，生成モデルがうまく訓練されていない領域に到達して，生成モデルがもっともらしいと勘違いしているが，実はそうではないデータが生成されてしまうのだ．特に，反実仮想の生成を目的とする場合はそのような領域に到達する傾向がある．生成モデルも日進月歩なので，生成モデル自体が成長することでこのアプローチも自動的に改善すると考えられ，そのような生成モデルを取り入れることが一つの将来の方向性だとこの論文では指摘している．このほか，連結性や化合物の価数制約などは，説明手法の側でハードに制約を設けながら説明を探索するアプローチも考えられる．

このほか，解釈性について様々なアプローチの手法が ICLR 2024 では提案された．シャオキ・ワンら [8] は決定境界にあるぎりぎりのグラフを構築することで，モデルの分類法則を解釈する方法 GNNBoundary を提案した．ペーター・ミューラーら [9] はメッセージ伝達を決

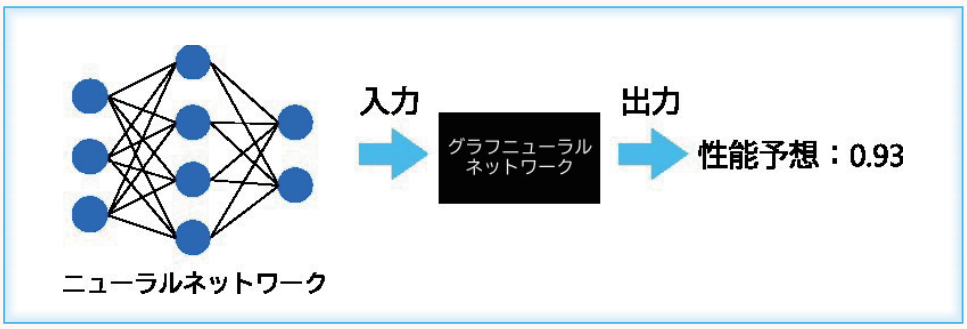


図２　メタネットワークの概要
メタネットワークはニューラルネットワークを入力として受け取り，数値や別のニューラルネットワークを出力する．ニューラルネットワークはグラフであるので，メタネットワークはグラフニューラルネットワークにより実現できる

定木に蒸留することでモデルの分類法則を解釈する方法 GraphChef を提案した．

これを使っておけば間違いないというような手法は少なくとも今のところはないので，タスクの性質やどういう説明が欲しいかといった個々の状況に応じて，適切な手法を選択する力量が求められるのが現状である．

4 同変性

モデルが同変（equivariant）であるとは，ざっくり言うと，モデルが入力データの対称性を保存して出力を決定するということである．例えばベンゼン（6 個の炭素が輪になった化合物）は対称である．ベンゼン中の一つの炭素原子は正例，ほかの五つの炭素原子は負例と分類するモデルは対称性を破っているので同変ではない．同変なモデルでは全部を正例とするか，全部を負例とするかのどちらかであることが保証される．タスクによっては，入出力が同変であることが要請される場合もあり，そのようなときには同変なモデルを使うことで明らかな間違いを防げる．また，1 個の炭素原子にだけ訓練ラベルを付けた場合にも，実質，ほかの五つの炭素原子にも訓練ラベルを付けたことと同じ効果が得られ，データ効率が向上する．ベンゼンの場合は明らかであるが，もっと複雑な場合には人手でケアしきれないので，モデルが同変であると保証されていることは有益である．グラフニューラルネットワークは同変であることが保証されている．

同変性についての研究は大きく分けると，同変性が保証されるモデルを提案したり，どういう場合に同変性が成立するかを数学的に証明する研究と，同変性をうまく活用した応用を提案する研究に分かれる．前者は関数解析などのニューラルネットワーク解析の基本的な道具のほか，群論など対称性を扱う数学的な道具が活躍するので，数学寄りの人に人気の研究分野である．ICLR 2024 では，回転や平行移動などの操作について同変である高速なモデルを提案したこと [10] が一例である．

応用研究としては，モデルパラメータを入力として受け取るメタネットワークの研究が２件発表された [11), (12)]．これらは，何百万次元から成るモデルパラメータベクトル θ を受け取り，そのモデルの性質を予測するモデルを得ることを目指す（図 2）．入力モデルの精度を予測することが応用の一例である．すなわち，モデル自体が別のモデルに入力される一段メタな問題である．多層パーセプトロンなど，多くのアーキテクチャには対称性が存在する．モデルパラメータの「○次元目」と「△次元目」を入れ替えても同じモデルを表すため，それらは同じ予測をするべき，つまり次元の入れ替えについて同変（不変）であるべきである．文献（11），(12）ではモデルパラメータベクトルを入力として受け取るタスクをグラフニューラルネットワークで解くことを提案している．グラフニューラルネットワークの出力は入力モデルのニューロンの順序によらないので，このような対称性に自動で対処できる．実験では，モデルパラメータベクトルを見るだけでそのモデルの性能をかなりの精度（順位相関係数 >= 0.9）で当てられている．

5 表現能力

モデルが解ける問題の範囲をそのモデルの「表現能力」と言う．グラフニューラルネットワークの表現能力は複雑で，どのような問題が解けるかを調べたり，解けないとされている問題をどのようにすれば解けるようになるかを考えたりということが精力的に研究されている．このトピックにおける嚆矢はグラフニューラルネットワークとワイスフェイラー・リーマン検査の等価性を指摘したクリストファー・モリスらとケユリュー・シュウらによる研究 [13), (14)] である．以降５年以上にわたり盛んに研究が続けられているトピックである．このトピックは関連する問題がグラフ理論やグラフアルゴリズム分野で古くから研究されており，それらの道具をグラ

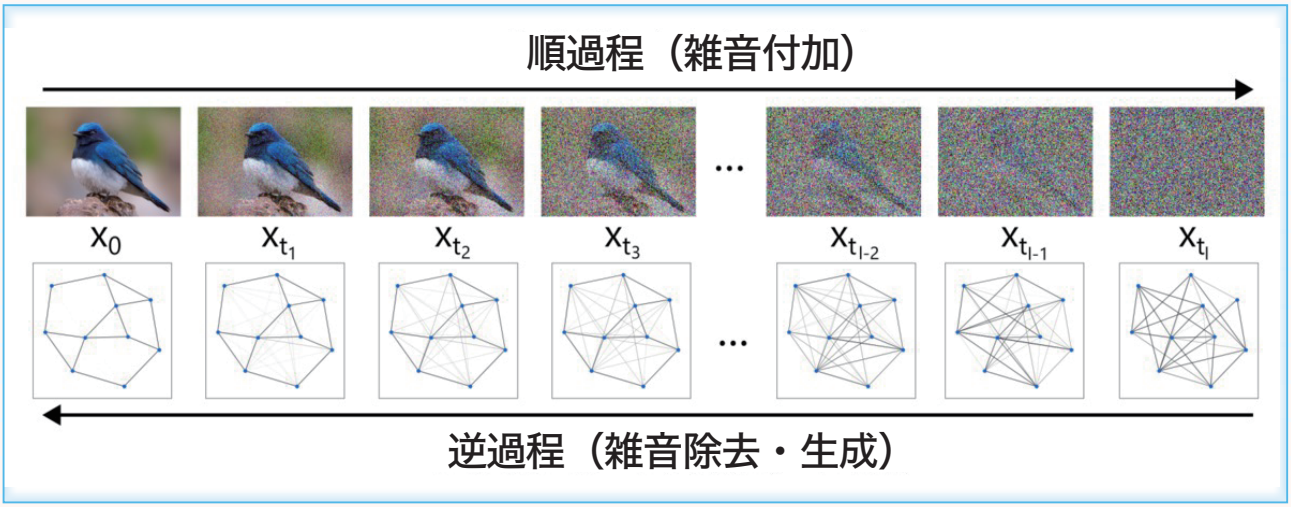


図３　拡散モデルの図示
画像に対する拡散モデルと同様に，雑音を付加する過程を逆算するモデルを訓練し，完全な雑音から雑音成分を繰り返すことでグラフデータを生成する．

フニューラルネットワークに持ってきて解析するという流れが定番である．私の以前の研究[15]も局所分散アルゴリズムという古典的な道具をグラフニューラルネットワークに持ってきて解析に用いるというものだった．最近は掘り出し物の飛び道具も尽きたので，細かな話題を拾いながら漸進しつつ，さてどうするかというような空気のように思う．

ICLR 2024では部分グラフの認識可能性に基づいた研究が目立った．文献（16）では，どういう部分グラフを見分けられるかということを基準に，グラフニューラルネットワークの表現能力を分類することを提案している．文献（13），（14）以来，今まではワイスフェイラー・リーマン検査を基に表現能力を分類することが一般的だった．ワイスフェイラー・リーマン検査はグラフニューラルネットワークと相性が良く，解析しやすいことが利点だが，大雑把な分類しかできないことと，ワイスフェイラー・リーマン検査を基準に「○○である」と言われても，応用者にはぴんと来ないという問題があった．どういう部分グラフを見分けられるか，であれば応用者も解釈しやすく，粒度も細かく分類できる．このような考え方自体は文献（17）など以前からあったが，文献（16）はこの考え方を体系化したことが貢献である．このほか，文献（18）や（19）は，ではどうすればグラフニューラルネットワークでより多くの部分グラフを見分けられるようになるか，ということを研究している．

ゼハオ・ドンら[20]は別の方向性を提案している．グラフを見分ける問題は古くから研究されており，具体的なアルゴリズムやツールも多く開発されている．ドンら[20]は，それらのツールを前処理として使って，結果をグラフニューラルネットワークに提示することで，グラフニューラルネットワークの表現能力を高めることを提案する．全てをニューラルネットワークに任せたい流派と，外部の資源を活用しながら能力を高める流派はこの分野に限らず常に綱引きの状態だが，表現能力では前者が主流だったことに対して，ドンらの研究はその反動の一例とも捉えられる．

6　時系列データ

グラフニューラルネットワークは時系列データにも適用できる．グラフニューラルネットワークを時系列データに適用することは，非常に簡単にできてしまい技術的な課題が少ないので，基礎的な研究よりも応用研究が多い．ICLR 2024では，脳波などの生体センサデータ[21],[22]や電子カルテデータ[23]など医療応用の研究が多く見られた．このほか，文献（24）では，センサの欠損値を周囲のセンサから推定する仮想センシングにグラフニューラルネットワークを応用している．地理的な接続をグラフで表し，各地点での時系列を処理する時空間データの分析にもグラフニューラルネットワークが応用されることが多い．

7　分子

分子データはグラフニューラルネットワークの応用先の花形だ．創薬や材料工学など，社会的にインパクトの大きい魅力的な応用がたくさん提案されている．創薬にしろ，材料工学にしろ，良い分子を見つけることが目標になることが多いので，生成モデルの研究が盛んである．

分子データの分野では拡散モデル（図３）が最も勢いがある．拡散モデルとグラフニューラルネットワークの

組合せ自体は文献（25）で提案されてから既に 4 年以上経過しているが，ICLR 2024 でも 文献（26），（27）など新しい拡散モデルが提案された．拡散モデルをグラフデータに適用すること自体にはそこまで技術的な障壁はないが，分子特有の対称性（同変性）を入れたり，三次元空間での原子の位置関係を考慮したり，応用に根差した帰納バイアスを入れるところがこのトピック特有の課題としてよく研究されている．

ICLR 2024 においてこの分野における特筆すべき成果は文献（28）である．この研究では 1 億以上の分子から成る超巨大なデータセットを整備し，分子データのための基盤モデルの構築を試みた．MILA など複数のグループにわたる 35 人の研究者から成る大きなプロジェクトである．この論文はどちらかというとデータセットの整備が中心であり，基盤モデルの構築はまだ道半ばといった感じである．複数タスクで訓練したモデルがほかのタスクにうまく転移できている例もあるが，最大のデータセットではうまくいっていない．ここは今後の課題である．だが，グラフデータに対する生成モデルはテキストや画像などと比べてデータ不足のために遅れがちだった問題に対して，大きく前進することはできた．この研究を足掛かりに，モデルもスケールするなどして今後 1〜2 年でこの分野は大きく進展するのではないかと思う．

8 物理

グラフニューラルネットワークを物理シミュレーションへ応用させることも最近人気である．一つのきっかけは ICLR 2021 で傑出論文賞（outstanding paper award）を受賞した文献（29）である．この研究ではグラフニューラルネットワークを使ってかなり高精度に物理シミュレーションを実現した．

ICLR 2024 でも物理シミュレーション応用が目立った．中でも，各粒子にセンサがなくても（RGBD）動画像観測から粒子シミュレータを学習できる [30]，部分的な観測だけからシミュレータを学習する [31]，メタ学習を基に未知の物理システムに汎化できる [32] など，データ駆動特有の利点を生かした研究が多く見られた．

9 おわりに

グラフニューラルネットワーク自体のコア技術はかなり成熟した一方で，様々な分野に応用が広がったために実用上の課題が多く現れた．今はそれらの解決を目指している段階であるという感触である．応用分野が広がった分，論文の数も非常に増えた．今後もこの傾向は持続すると思われる．あらゆるドメインと組み合わせられることがグラフニューラルネットワークの大きな強みであり，どのようなドメインでもグラフニューラルネットワークを活用するチャンスがある．本稿で興味を持った方は各々のドメインに適用できないか考えて頂ければ幸いである．

■文献

(1) A. Sperduti and A. Starita, “Supervised neural networks for the classification of structures,” IEEE Trans. Neural Netw., vol. 8, no. 3, pp. 714-735, May 1997.
(2) M. Gori, G. Monfardini, and F. Scarselli, “A new model for learning in graph domains,” Proc. 2005 IEEE International Joint Conf. Neural Netw., July 2005.
(3) T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” ICLR 2017.
(4) Z. Ying, D. Bourgeois, J. You, M. Zitnik, and J. Leskovec, “GNNExplainer: generating explanations for graph neural networks,” NeurIPS 2019, pp. 9240-9251, Dec. 2019.
(5) M. Kosan, S. Verma, B. Armgaan, K. Pahwa, A. K. Singh, S. Medya, and S. Ranu, “GNNX-BENCH: unravelling the utility of perturbation-based GNN explainers through in-depth benchmarking,” ICLR 2024.
(6) J. Fang, W. Liu, Y. Gao, Z. Liu, A. Zhang, X. Wang, and X. He, “Evaluating post-hoc explanations for graph neural networks via robustness analysis,” NeurIPS 2023.
(7) J. Ma, R. Guo, S. Mishra, A. Zhang, and J. Li, “CLEAR: generative counterfactual explanations on graphs,” NeurIPS 2022.
(8) X. Wang and H. W. Shen, “GNNBoundary: towards explaining graph neural networks through the lens of decision boundaries,” ICLR 2024.
(9) P. Müller, L. Faber, K. Martinkus, and R. Wattenhofer, “GraphChef: decision-tree recipes to explain graph neural networks,” ICLR 2024.
(10) E. J. Bekkers, S. Vadgama, R. Hesselink, P. A Van der Linden, and D. W. Romero, “Fast, expressive SE (n) -equivariant networks through weight-sharing in position-orientation space,” ICLR 2024.
(11) M. Kofinas, B. Knyazev, Y. Zhang, Y. Chen, G. J. Burghouts, E. Gavves, C. G. M. Snoek, and D. W. Zhang, “Graph neural networks for learning equivariant representations of neural networks,” ICLR 2024.
(12) D. Lim, H. Maron, M. T. Law, J. Lorraine, and J. Lucas, “Graph metanetworks for processing diverse neural architectures,” ICLR 2024.
(13) C. Morris, M. Ritzert, M. Fey, W. L. Hamilton, J. E. Lenssen, G. Rattan, and M. Grohe, “Weisfeiler and leman go neural: higher-order graph neural networks,” AAAI 2019, pp. 4602-4609, Jan. 2019.
(14) K. Xu, W. Hu, J. Leskovec, and S. Jegelka, “How

powerful are graph neural networks?” ICLR 2019.
(15) R. Sato, M. Yamada, and H. Kashima, “Approximation ratios of graph neural networks for combinatorial problems,” pp. 4083-4092, NeurIPS 2019..
(16) B. Zhang, J. Gai, Y. Du, Q. Ye, D. He, and L. Wang, “Beyond Weisfeiler-Lehman: A quantitative framework for GNN expressiveness,” ICLR 2024.
(17) Z. Chen, L. Chen, S. Villar, and J. Bruna, “Can graph neural networks count substructures?,” NeurIPS 2020, Feb. 2020.
(18) C. Kanatsoulis and A. Ribeiro, “Counting graph substructures with graph neural networks,” ICLR 2024.
(19) M. Lanzinger and P. Barceló, “On the power of the Weisfeiler-Leman test for graph motif parameters,” ICLR 2024.
(20) Z. Dong, M. Zhang, P. Payne, M. A. Province, C. Cruchaga, T. Zhao, F. Li, and Y. Chen, “Rethinking the power of graph canonization in graph representation learning with stability,” ICLR 2024.
(21) C. Liu, X. Zhou, Z. Zhu, L. Zhai, Z. Jia, and Y. Liu, “VBH-GNN: variational bayesian heterogeneous graph neural networks for cross-subject emotion recognition,” ICLR 2024.
(22) Y. Song, B. Liu, X. Li, N. Shi, Y. Wang, and X. Gao, “Decoding natural images from EEG for object recognition,” ICLR 2024.
(23) R. Poulain and R. Beheshti, “Graph transformers on EHRs: better representation improves downstream performance,” ICLR 2024.
(24) G. De Felice, A. Cini, D. Zambon, V. V. Gusev, and C. Alippi, “Graph-based virtual sensing from sparse and partial multivariate observations,” ICLR 2024.
(25) C. Niu, Y. Song, J. Song, S. Zhao, A. Grover, and S. Ermon, “Permutation invariant graph generation via score-based generative modeling,” AISTATS 2020.
(26) T. Le, J. Cremer, F. Noe, D. A. Clevert, and K. T. Schütt, “Navigating the design space of equivariant diffusion-based generative models for de novo 3D molecule generation,” ICLR 2024.
(27) Z. Huang, L. Yang, X. Zhou, Z. Zhang, W. Zhang, X. Zheng, J. Chen, Y. Wang, B. Cui, and W. Yang, “Protein-ligand interaction prior for binding-aware 3D molecule diffusion models,” ICLR 2024.
(28) D. Beaini, S. Huang, J. A. Cunha, Z. Li, G. Moisescu-Pareja, O. Dymov, S. Maddrell-Mander, C. McLean, F. Wenkel, L. Müller, J. H. Mohamud, A. Parviz, M. Craig, M. Koziarski, J. Lu, Z. Zhu, C. Gabellini, K. Klaser, J. Dean, C. Wognum, M. Sypetkowski, G. Rabusseau, R. Rabbany, J. Tang, C. Morris, I. Koutis, M. Ravanelli, G. Wolf, P. Tossou, H. Mary, T. Bois, A. Fitzgibbon, B. Banaszewski, C. Martin, and D. Masters, “Towards foundational models for molecular learning on large-scale multi-task datasets,” ICLR 2024.
(29) T. Pfaff, M. Fortunato, A. Sanchez-Gonzalez, and P. W. Battaglia, “Learning mesh-based simulation with graph networks,” ICLR 2021.
(30) W. F. Whitney, T. Lopez-Guevara, T. Pfaff, Y. Rubanova, T. Kipf, K. Stachenfeld, and K. R. Allen, “Learning 3D particle-based simulators from RGB-D videos,” ICLR 2024.
(31) S. Janny, M. Nadri, J. Digne, and C. Wolf, “Space and time continuous physics simulation from partial observations,” ICLR 2024.
(32) Y. Song and H. Jeong, “Towards cross domain generalization of hamiltonian representation via meta learning,” ICLR 2024.

佐藤竜馬

国立情報学研究所．2024 京大大学院情報学研究科博士課程了．博士（情報学）．現在，国立情報学研究所助教．専門分野はグラフニューラルネットワーク，最適輸送，及び情報検索・推薦システム．著書に「深層ニューラルネットワークの高速化」（技術評論社），「グラフニューラルネットワーク」，「最適輸送の理論とアルゴリズム」（講談社）がある．

サービス学と ICT

仙石正和　Masakazu Sengoku　事業創造大学院大学

1 はじめに

近年，SNS などのサービス産業が産業構造の中で大きな部分を占めるようになり，そのサービス産業における ICT（情報通信技術）の役割は一層増している．本稿では，サービスの捉え方の変遷，サービス学，サービス工学から関係する工学教育について述べる．更にサービスに関係するネットワークの基礎理論についても紹介する．

サービス産業もサービスに関する研究も発展段階であり，統一的に論ずることは難しく，本稿は，あくまでも筆者の私見であることをあらかじめお断りしておく．

2 サービスの捉え方

およそ 70 年前の日本の第一次産業，第二次産業，第三次産業の割合は，45 %，25 %，30 %であったが，現在は 3 %，23 %，74 %となっている．第一次産業は農業，林業，漁業などで，第二次産業は製造業，建設業，工業など，第三次産業はその他の産業が入るが，第三次産業を全てサービス産業と呼んでよいのか，第一次産業，第二次産業の部分にもサービスの部分はあるのではないかとの考え方もあるであろう．そこで，サービスに関する考え方の変遷について考えてみたい [(1)~(3)]．

現在，パソコン，スマートフォン，ネットワークなどから成る IT 基盤は，電気，水道などの社会基盤と並んで，必要不可欠なものになってきている．その技術である ICT は，環境，都市，農業，資源，流通，土木，医療，教育などのそれぞれの産業を抜本的に変革し，産業構造，経済構造，社会構造までをも大きく変える力を有すると認識されている．そして，多くの分野へ活用され，その一層の広がりも期待されている．このことは，ICT の汎用品化と見ることができ，その要因が，無形財の経済へ影響が大きくなってきていることにあると考えられている [(4)]．

初期の段階のサービスの捉え方（理解）は，

① 無形性：実物を見たり触ったりできない．
② 異質性・非均一性：サービスの質は時間，場所，関係する人間によって異なる．
③ 同時性：生産と消費が同時に発生する．
④ 消滅性：消費のために蓄積しておけない．

である．そして，‘もの’とサービス（goods and services）をそれぞれ有形財，無形財と捉えていた．この中で，特に無形性を有する無形財の経済への影響と ICT の活用とが深く関わることから，サービス産業と ICT は元々強く結び付いていると考えられてきた．

有形財，無形財の捉え方の二分法では，図 1 の（a）のように‘もの’そのものの機能的価値を最大化するグッズドミナントロジック（Goods Dominant Logic）が主流であったが，1980 年頃から‘もの’をサービスに含めて考えるようになり，2000 年頃に，‘もの’をサービスの一部と捉えて，図 1（b）のように，サービス全体としての経験的価値も最大化するという，サービスドミナントロジック（Service Dominant Logic）が提案された [(5)]．更に，提供物を介して，提供者と受容

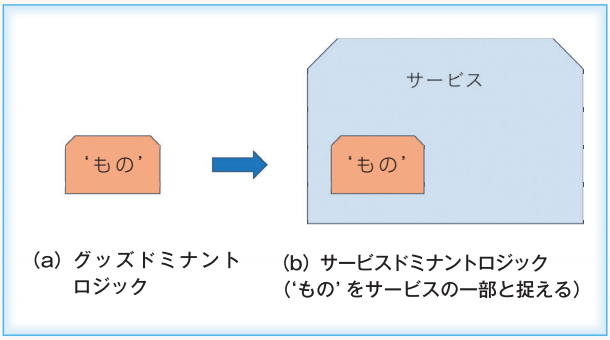


図 1　グッズドミナントロジックからサービスドミナントロジックへの移行

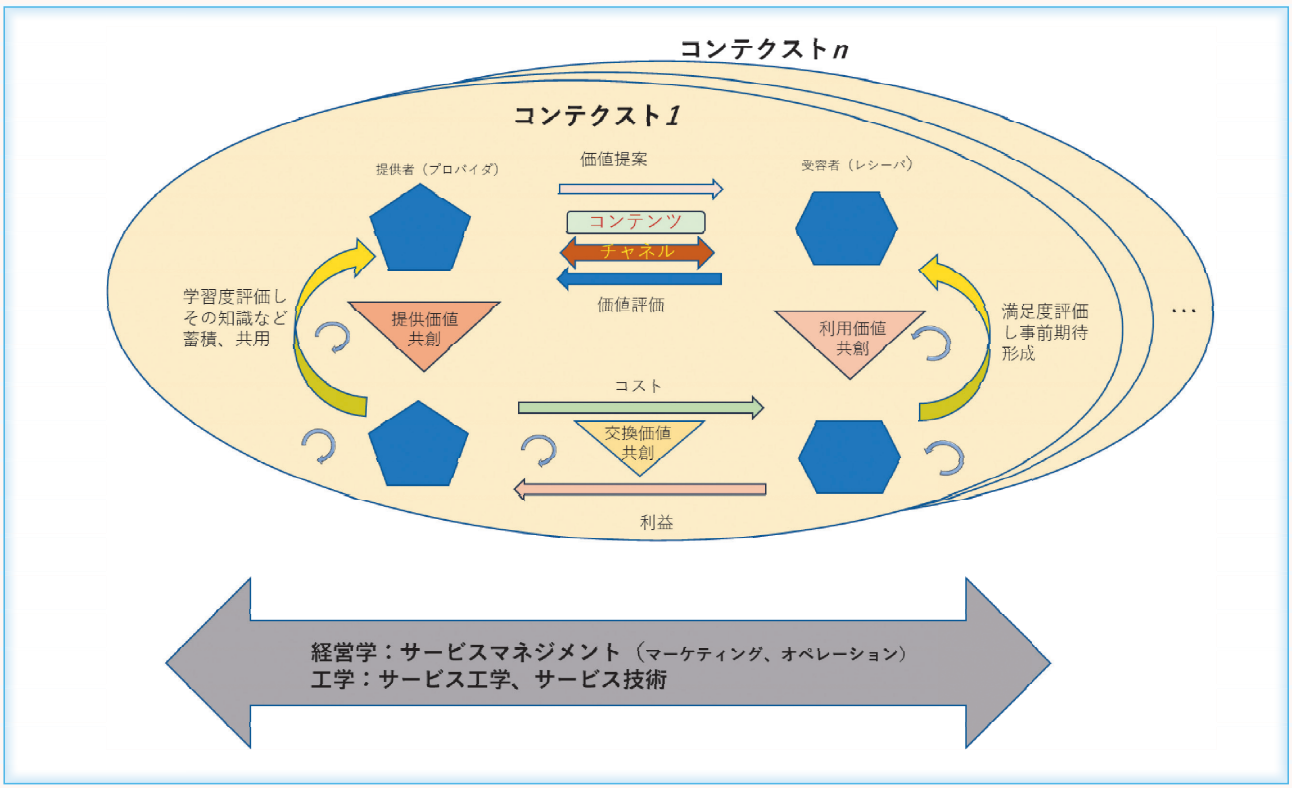


図２　サービス価値共創モデル（文献（1）を参照して作成）

者とが相互に働き掛けることの全体を指して「サービス」と呼ぶことが最近は一般化してきている[(1)]．

図２に一つのサービス価値共創モデルを示す．五角形は提供者（個人も団体も含む），六角形は受容者（個人も団体も含む）を示す．上側の五角形，六角形の間で，提供者はチャネル（輸送網，通信網など）を通してコンテンツ（有形，無形）を受容者に届ける．これは，提供者が価値を提案して，受容者はその価値を評価する．下側の五角形，六角形は，同種の提供者と受容者で，コストと利益の流れ（交換価値共創）を表している．これは提供者と受容者は相互に働き掛けながら協同して価値を創造（価値共創）することを図で小さなループで表している．（説明上，上側の五角形，六角形ではコンテンツとチャネルの機能を，下側で交換価値共創の機能を示したが，上側も下側も両機能を有している．交換価値共創を小さなループで表示した．）更に，提供者同士（提供価値共創），受容者同士（利用価値共創）の間でも相互に価値共創がなされ，この価値共創も小さなループで表している．また，五角形，六角形が団体などの場合はその中でも価値共創がされる可能性がある状況を五角形，六角形の左または右隣の小ループで表している．現状でサービスに特に関係が深い既存分野は経営学と工学であるので，図２の下部にサービスを支える経営学，工学の分野名を示している．

コンテンツが有形，無形を含むので，第一次産業や第二次産業にも，サービスは含まれることになる．このよ

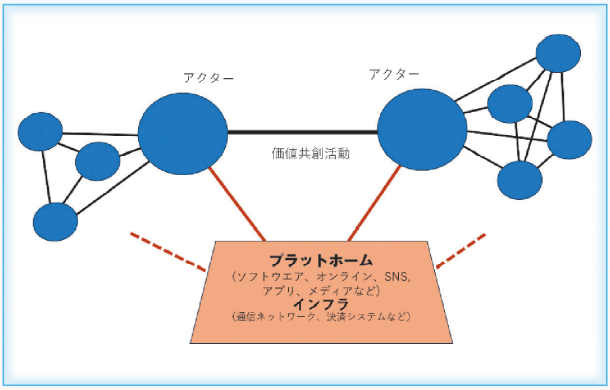


図３　サービスと通信ネットワーク

うに社会全体へのサービスの役割が大きくなってきている社会を「サービス化社会」と呼ぶようになってきている．サービスの概念は業種によって伝統的な捉え方と現代風な捉え方とが混在しているが，徐々に図２のような捉え方が定着しつつある．

サービスを図２のように捉えると，提供者と受容者を区別しておくのではなく，サービス化社会では，アクターを提供者・受容者の両方を表しているとして，図３のようにモデル化してよいであろう．

図３において，丸はアクターを表し，アクターは個人，団体（企業など）を表す．このアクターがネットワーク化され，それをインフラやプラットホームが支えている．インフラには通信ネットワーク，決済システムなどが含まれ，プラットホームには，ソフトウェアプ

ラットホーム（Windows や iOS・Android），オンラインプラットホーム（検索エンジン，SNS，アマゾン・楽天のようなショッピングサイト），アプリプラットホーム，メディアプラットホーム（YouTube ほか）などが含まれる．

サービスは，図３のようなネットワークの中で価値共創活動をするが，どのような価値を目指すのであろうか．各アクターは資源提供者でもありそれを総合的にまとめて価値を創造するが，現実にはこの提供される資源（知識やスキルも含む）をまとめる仕組みを提供する主体が企業になる．このとき，創造すべき価値をどのように考えればよいであろうか．図１のところでは‘もの’そのものの機能的価値を最大化するグッズドミナントロジックから，サービス全体としての経験的価値も最大化するというサービスドミナントロジックへの変遷を述べた．更に，‘もの’づくりから‘もの’／‘こと’づくりへ拡大する中で，経験的価値，使用価値，意味的価値，文脈価値などの価値の言葉（表現）が使われてきている．（例えば文献（6）．）これらの言葉は，サービスには人間が主役として関わっていることから来る各種表現と考えられる．最近の社会では医療，教育，観光，エンターテインメントなどへの支出が増大している．このことは，消費者の選択の対象は，より便利でより快適な生活を実現するものだけでなく，より楽しく，新鮮な経験を提供してくれるものに移ってきていることと無縁ではない．このような中では個人個人の好みで，価値が変わる．そこで，最近の価値の表現として，以下の三つが提案されている [7]．

・機能価値
・知識価値
・感情価値

ここで機能価値は例えば製品の性能などであり，知識価値は例えば経験や利用で蓄積される価値で，感情価値は例えば満足・不満足や好き・嫌いの程度の価値である．これらの価値の中で，機能価値は客観的に評価できる場合が多いが，知識価値，感情価値はアンケートなどに頼らざるを得ない．このアンケートによる評価をより客観的にすることが，サービスの学術の構築には大切であり，感情価値などを科学的に測ることの必要性と重要性は，産業界からも指摘されている [8]．

グッズドミナントロジックの時代では，‘もの’づくり側（提供者側）の倫理観が重要視されていたが，サービス化社会では，アクターは提供者と受容者の両方であることから，社会全体での倫理観，社会的責任などが問われることになる．

3 サービス学

サービスが社会の中で大きなウェイトを占めるようになってきて，サービスを学問の一つである「サービス学」として捉えて教育研究がされるようになってきている．サービス学は以下のように定義されている [2]．

「サービスとは，提供者と受容者が価値を共創する行為で，サービスは人間を含むシステムにおいて持続的かつダイナミックに生産・提供・消費される．」と定義される．このような性質を持つ「サービスに関する総合的な学問体系」をサービス学と呼ぶ．

図２で説明したようにサービスは提供者からの一方的な行為では成り立たず，受容者との「共創・相互行為」によって生み出される．また，サービスにおける提供者と受容者の共創関係や価値は，相互作用を繰り返すことによってダイナミックに変容する．このようにサービスには人間が深く関わるため，複雑で現状ではサービス学は十分に体系化がされているとは言い難い状況のように見える．物理学や化学のような自然科学のように，人間の感情などは排除した法則性に基づき体系化し，その応用手法に慣れた伝統的な工学領域では異質に感ずるかもしれない．一方で，工学分野においては，OR（オペレーションズリサーチ）や通信理論でも利用される待ち行列理論などを活用し，輸送業，ホテルや外食業での予約システムが構築されている．更に，IoT を利用して遠隔で計測されたビッグデータを蓄積，機械学習し，AI を用いることで，大規模社会システムの管理分野でサービスの仕様策定から設計，制御などが可能となってきている．このような分野はサービス工学に含まれる．また，経営学では，サービスマーケティングの研究が盛んで，サービス品質，顧客満足などの概念を使い，サービス提供事業体の経営理論などが利用されている．このように，現状でサービスに特に関係が深い既存分野は経営学と工学である．以上ように経営学，工学の分野とサービスへの関係を述べたが，サービス学にはそのほかにも多くの分野が関わっており，その体系化が望まれるところである．

4 サービス工学，サービス技術と工学教育

従来，工学とは‘もの’づくりの際に，設計，生産，消費，利用，廃棄と続く，製品の一生の体系であると考えられてきた．同様に，サービスをその設計から廃棄までの一生に役立てることを目指した体系をサービス工学と称している．サービス工学がサービスの分野の一つであるならば，当然サービス技術もその分野の一つと考え

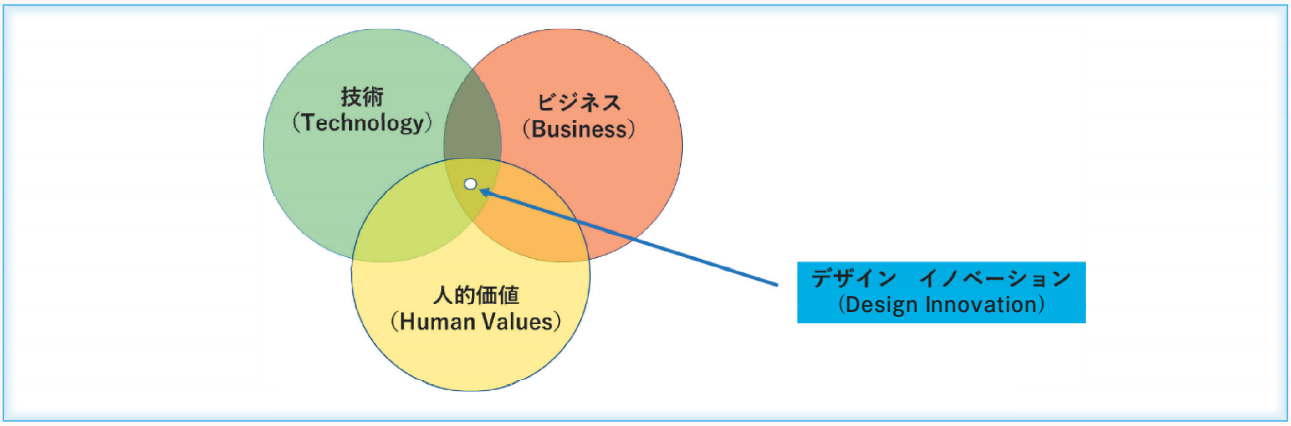


図４　スタンフォード大のデザイン教育　（文献（10），（11）を参照して作成）

なければならない．近年，各大学の工学部の教育プログラムの設計の中で，サービス化社会などの時代の変化へ向けて，工学と技術の教育をどのようにすべきかの議論がされている．この議論の中で，工学と技術について次のように定義されている [9]．

・工学（Engineering）：数学，自然科学の知識を用いて，健康と安全を守り，文化的，社会的及び環境的な考慮を行い，人類のために（for the benefit of humanity），設計，開発，イノベーションまたは解決を行う活動．
・技術（Technology）：エンジニアリング活動における実践的応用を可能にするツール，テクニック（一連の処理方法またはやり方），用具，コンポーネント，システムまたはプロセスを伴う，（それらのメカニズム，性能，機能などを含む設計，製造に用いられ，役に立つことが実証済みの）知識体系（an established body of knowledge）．

工学の定義では，人類のために，設計，開発，イノベーションまたは解決を行う活動となっているので，工学教育はこのような活動ができる人材育成を目指すことになる．技術は工学の中で使用されるツール，手段となっていることから，技術は工学に含まれ，独立した存在ではないと思われるかもしれないが，そういう解釈は妥当ではない．技術そのものは長い歴史と伝統を持つ深い領域で，最先端のサイエンスを利用したり，それが使えない場合は独自に研究，開発したりする独立した学問の一領域でもある．実際，現代の基礎教育に身に付けるべき領域としてSTEAM（Science, Technology, Engineering, Art, Mathematics）と言われるように，技術と工学は明確に区別されている．直感的には，技術は文字どおり技（わざ）であり，サイエンスで明らかにされた知識を使って，有用な事物を発明することで，工学は技術などを使って公共の安全，福祉，環境を考慮して人間に役立つ事物を構築することと考えることができる．

工学には，イノベーションも含まれる．米スタンフォード大でイノベーションを興す人材育成に「d. school」の役割が注目されているが，そのコンセプトは図４のようなデザイン教育である [10], [11]．

図４で，デザインは技術（Technology）と経営（Business）と人的価値または人間の価値（Human Values）の共通部分としており，人間がどのように感ずるかの人間性の部分を考慮することの重要性を示唆している．すなわち，機能価値のみならず知識価値，感情価値を含む価値創造を目指している．このコンセプトは正にサービス化社会に向けての‘もの’／‘こと’づくりの肝と言うべきであろう．

日本では，日本技術者教育認定機構（JABEE）が2005年にワシントン協定（WA: Washington Accord）に加盟が認められ，日本の工学，技術の教育の国際認定を行っている．この中で，学生がプログラム修了時に確実に身に付けておくべき知識・能力には，「デザイン能力」が含まれている [12]．現状で，JABEEで審査を受けるかどうかにかかわらず，どの工学系，技術系の教育機関もこのデザイン能力の重要性は認識しており，デザインという呼称を使わないものも含めて，デザイン能力の向上のための様々な工夫がなされた教育がされている．デザインについて，次のように定義されている [9]．

・デザイン（Engineering Design）：人文社会科学，数理科学，自然科学，工学基礎，工学専門，テクノロジー等の学習成果を統合し，経済的，環境的，社会的，倫理的，健康と安全，製造可能性，持続可能性などの現実的な条件の範囲内で，ニーズに合ったシステム，構成要素，手順と方法を設計，開発，イノベーションを行う，創造的で，度々反復的で，オープンエンドなプロセス．

この定義の中で，経済的，環境的，安全で製造可能

性，持続可能性などの現実的な条件下でニーズに合ったシステムを作る際に，ニーズには人間の要望，要求，満足度なども含む．更にシステムを作る際は，イノベーションを含む創造的なプロセスを繰り返すことになる．このプロセスでは，機能，知識，人間の満足度などを向上，増加させることを目指す．図４のスタンフォード大の例のように，デザイン（デザインイノベーション）は，テクノロジー，ビジネス，人間の価値（満足・不満足，好き・嫌いなどの感情）の共通部分が土台となっており，サービス化社会での共創価値である機能価値，知識価値，感情価値を含む価値につながっていることに注目すべきである．すなわち，機能価値のみならず知識価値，感情価値などの創造を目指すイノベーションに向かう能力がデザイン能力である．デザイン能力の向上のための具体的教育好例は，建築学領域の建物の建築や都市計画などで見受けられるよう思う．この領域ではデザイン教育のコンセプトに近い教育が行われ，長い歴史的なノウハウを多く有しており，現在でもその内容は色あせていない [13]．この中で，学生が住民など人々の意見を聞きながら協同作業で建物を設計し，プロトタイプから，大工さんと一緒に実際に建物を建築する共創活動は，研究室で行うプロトタイプ型教育とは本質的に異なる．このような例には，サービスの機能価値，知識価値，感情価値の例をたくさん含んでいる．

5 ネットワークの基礎理論

サービスの社会での構造が，図３のようなネットワーク形で，アクター同士の共創的な価値の生産となっているので，現状の多くの産業のサプライチェーンやバリューチェーンのような一方向の構造とは様相が異なる．このように考えると，参加者が限られた価値を取り合うゼロサムゲームではなく，参加者が共創によって全体の各種価値を増加させることができる構造である．これは異業種の企業や個人がそれぞれの技術やノウハウを共有して収益を上げるエコシステムと言える．このシステムでは，サービスの状況をディジタル技術，IoT により多数で多様なデータを取得してサイバーフィジカルシステム（CPS）でシステムの分析，AI を活用しての最適化などを行うことが必要であろう．その中で，メタバースを用いて，アクターのアバタを通じてコミュニケーションや経済活動をシミュレーションして，サービスの特性を明らかにしてサービス学の体系化に役立てる必要も出てくるだろう．ここでも ICT の役割は大きい．この役割の中で，ネットワーク基礎理論上から眺めたときの具体的な課題について，幾つか挙げる．これらは重要性とか緊急性によって列挙したものではないことに注意頂きたい．

5.1　サービタイゼーション

サービスにおいて，提供者と受容者が ICT で結ばれ，無形財はほとんど同時に生産と消費が起こることもあり，これによって生まれる知識（共創価値，または資源）は同時に共有されることになる．つまり，サービスの提供者と受容者が持っている有形財（‘もの’）や無形財（‘こと’）を含む資源の統合が ICT によって簡単になることで，製造業にサービタイゼーション（servitization）またはサービサイジング（servicizing）という変化をもたらしている．サービタイゼーションとは，生産した製品を販売し売上げを伸ばすのではなく，製品をサービスとして顧客に提供することによって売上げを伸ばすビジネスモデルのことである．サービサイジングとは，製品機能を‘もの’として販売するのに代えて，「リース」「レンタル」「シェアリング」のようにサービスの形で提供する機能販売型のビジネスモデルである．シェアリングエコノミーサービスでは，‘もの’，スキル，空間など各種‘もの’や‘こと’をシェアする．各種の‘もの’や‘こと’はそれぞれ特性があるが，時には到着特性が異なる複数の‘もの’／‘こと’を一緒にシェアしたい場合もある．この場合，その性能を解析するとき，待ち行列理論が有効であろう．（例えば，文献（14）．）この課題は，移動通信における，移動流トラヒックと通信トラヒックが混在する解析問題 [15] と類似する興味ある問題である．

5.2　ソーシャルネットワーク

サービス学におけるネットワーク論は，ソーシャルネットワークの分野とも関係しており，急速に発展している領域である．ソーシャルネットワークの研究は，人と人との社会的つながりがどのように成り立っているのか，そのつながりは人や組織にどのように影響を与えるかとの疑問に答えようという分野である．インターネットのネットワーク構造などとも関わっている．筆者らは 40 年ほど前に，情報ネットワークに関わるネットワーク工学の研究を行ってきた中で，ネットワークの点（vertex）や辺（edge）の中心らしさを測る関数の研究を行ってきた（文献（15），（16）等）．点の中心性は様々な尺度があるだろうが，次数中心性，離心中心性，近接中心性，媒介中心性，隣接中心性などがある [17]．近接隣接性を利用した応用例を以下に示す．

$N(G,l,w)$ を連結有向ネットワークとし，$G=(V,E)$ は点集合 V，辺集合 N のグラフとする．辺と点はそれぞれ正の実数の長さと重みが付されており，$l(e_k)$ と $w(v_i)$ はそれぞれ辺 e_k の長さ，点 v_i の重みとする．点 v_j と v_k の

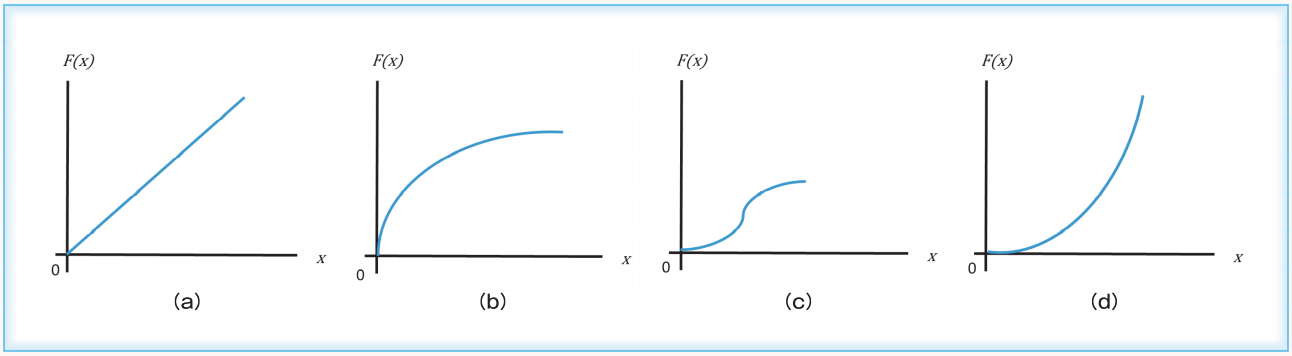


図５　コストなどを表す幾つかの増加関数

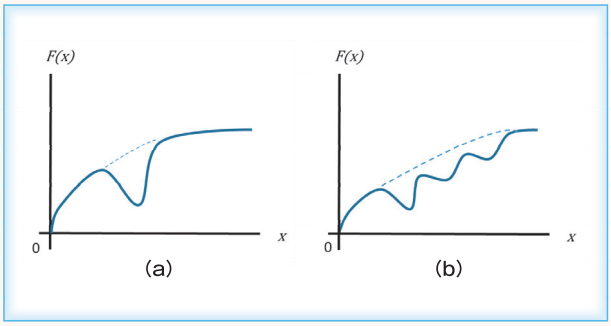


図６　コストなどを表す特殊な関数

最短経路の長さ（距離）を$d(v_j, v_k)$としたとき，点v_jの出伝送数（out-transmission number）$t_o(v_j)$，入伝送数（in-transmission number）$t_i(v_j)$はそれぞれ式(1)，(2)のように定義される．この伝送数は近接中心性の典型例である．

点の重みがどの点でも同じ場合，ある点の出伝送数は，その点からほかの全ての点に何かを送るときの総距離（スピードが同じならば，かかる総時間）を表し，入伝送数はほかの点からその点に送るときの総距離（総時間）を表す．この伝送数が小さいほど，何かを送るときに有利な点ということになる．

$$t_o(v_j) = \sum_{v_k \in V} d(v_j, v_k) w(v_k) \qquad (1)$$

$$t_i(v_j) = \sum_{v_k \in V} d(v_k, v_j) w(v_k) \qquad (2)$$

ある点から何かを送る場合，距離などによりコストなどが変化する場合がある．このコストなどを加味した伝送数を一般化出伝送数（generalized out-transmission number），一般化入伝送数（generalized in-transmission number）と言い，それぞれ式(3)の$h_o(v_j)$と式(4)の$h_i(v_j)$で表される．

$$h_o(v_j) = \sum_{v_k \in V} F_k(d(v_j, v_k) w(v_k)) \qquad (3)$$

$$h_i(v_j) = \sum_{v_k \in V} F_k(d(v_k, v_j) w(v_k)) \qquad (4)$$

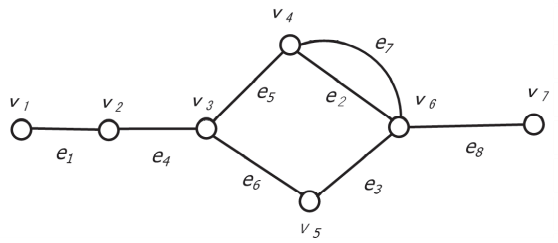
一般にコストは距離の増加関数であろうから，この関数$F(x)$の増加関数として，例えば図５(a)～(d)の例がある．文献(18)では，ネットワークの一辺に故障が生じた（等価的に長さが長く変化）場合，(a)，(b)の場合は，伝送数の変化から，故障した辺を特定できる（故障診断が可能である）ことを証明している．(a)は線系，(b)は下に凹関数（concave function）である．この関数の形は，列車の料金体系など実社会で見られる典型的なものである．(c)，(d)のような関数に関する故障診断の問題は解かれていない．

ソーシャルネットワークの研究では，「ストラクチャルホール」と「弱いつながりの強さ」が注目されている．いずれもソーシャルネットワークに関係しているが，スタンフォード大のマーク・グラノヴェッターは論文“Strength of Weak Ties”の中で，「弱いつながりの強さ」について論じている[19]～[21]．この「弱いつながりの強さ」の意味を，コスト関数$F(x)$の面から眺めると，図６の(a)，(b)のように，点から遠方にある場合にもコストが下がる場合があることになる．図の点線と実線は，全体としては増加しているが，途中で下がる場合があることを意味している．このような場合のネットワークの考察は未解決の興味ある課題である．媒介中心性などの応用例なども最近発表されている[22]．

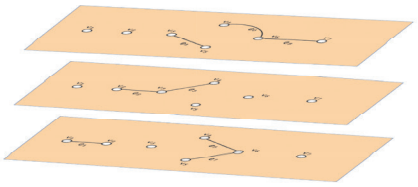
５．３　多次元ネットワーク

図７(a)のグラフ$G=(V, E)$で，辺集合が３種類に分割され$E=E_1+E_2+E_3$，$E_1=\{e_1,e_2,e_3\}$，$E_2=\{e_4,e_5\}$，$E_3=\{e_6,e_7,e_8\}$とする．例えば，各点は人を表し，辺は関係を表すとし，E_1はビジネス上の関係，E_2は趣味のスポーツの関係，E_3は趣味の絵画の関係を表すとする．このグラフGは$G_1=(V, E_1)$，$G_2=(V, E_2)$，$G_3=(V, E_3)$とすると図７(b)のように多層グラフのようになる

一般に，グラフ$G=(V, E)$で，辺集合がn種類に分割され，$E=E_1 \cup E_2 \cup \ldots \cup E_n$，$E_i \cap E_j = \phi$とする．



(a) グラフ



(b) 3 層ネットワーク

図 7　多次元ネットワーク（多層グラフ）

$G_1=(V,E_1), G_2=(V,E_2), \ldots G_n=(V,E_n)$ とすると，グラフ G は $G_1, G_2, \ldots, G_n$ から成る多層グラフになり，これを n 層グラフ，または n 次元ネットワークと呼ぶ．多次元ネットワークの次数中心性，離心中心性，近接中心性，媒介中心性，隣接中心性に関して，様々な研究がなされている [23], [24]．最近では，次元間の相関係数の研究もされている [25]．次元間の相関係数では，例えば，図 7 のグラフで，ビジネス上の人間関係と趣味上の人間関係の相関係数を計算することにより，ソーシャルネットワークの性質を調べるのに役立つ．

6　サービス学の発展に向けて

サービス学は，**2.** と **3.** で述べたように，学問として完成しているのではなく，コンセプトを構築している段階のように見える．幾つかの学問の成り立ちを眺めると，何か課題に向かったとき，まず関係しそうな様々な現象，技術を寄せ集めて，論理モデル化してコンセプトとして捉える．そして，関連したほかのコンセプトやデータを蓄積し，抽象化して理論となり，学問へと体系化され学問領域（discipline）となる．例えば，電磁気学では，最初様々な電気，磁気の現象，技術を寄せ集めて，コンセプト化から抽象化してマクスウェルの理論へと発展した．理論になることで誰でも利用できる知識として共有でき，多くの人々の議論や考察の対象になることで，知識やその応用のデータが蓄積，活用される．このようにして社会への影響が一層拡大する．このようなプロセスを経た電磁気学という学問への体系化がいかに大きな役割を果たしたかは，御存じのとおりである．サービスにおいては，コンセプト化や抽象化の段階では，個々の人間の感情などによる影響，すなわち感情価値などはできるだけ除かなければ，一般論とはならない．ところが，サービスには，**2.** で述べたように，本質的にこれらの価値が入っている．そのため，感情価値などを除く研究すなわち科学的な観測で感情価値などの計測を行い [8]，抽象化の範囲を広げ深める研究も必要であろう．または，ある程度感情価値などが入ったままで進める，ある種の学問領域への発展もあるかもしれない．本学会の会員の多くは，物理，化学，生物，数学などの体系化された基礎学問を出発点として工学を学んできた方々が多いかもしれない．近年の **4.** の工学教育の変化からも分かるように，産業構造は機能価値，知識価値，感情価値などを求める方向にあり，**5.** で述べたようにゼロサムゲームではないという魅力もある．本会会誌の巻頭言 [6], [26] からも，機能価値，知識価値，感情価値の価値創造や価値獲得への研究は学会の方向性とも合致している．本会でサービスを含む研究の活性化が楽しみである．

一般的なサービスに関する様々な研究がサービス学会 [27] でも活発にされており，サービス学の発展への期待は大きい．

7　おわりに

ICT がサービス産業に与える影響は大きい．そこで，サービスの捉え方，サービス学，サービス工学から工学教育の状況を述べ，更にサービスに関係したネットワークの基礎理論の幾つかを紹介した．

大学でのカリキュラムは学問領域に基づくカリキュラム（discipline-based curriculum）が伝統的であったが，社会の変化のハイスピード化から，課題に基づくカリキュラム（issue-based curriculum）も取り入れられつつある．筆者は大学で教育研究に携わってきて個人的に感じたことであるが，初学年では前者を重点に，高学年，大学院では後者を研究課題と一体としたカリキュラムが良いように思う．その理由は，初学年から，体系化されていない個別の課題とその解決策を学ぶとその課題の多さや複雑さ，個別の解決策の多様さに，気後れを感じてしまう可能性があることと，体系化された学問の大きな威力を最初に感じてもらうことが大切と思うからである．

■ 文献

(1) 村上輝康，松井拓己，価値共創のサービスイノベーション実践論，生産性出版，東京，2021.
(2) 日本学術会議，報告「大学教育の分野別質保証のための教育課程編成上の参照基準，サービス学分野」，日本学術会議経営学委員会・総合工学委員会合同サービス学分科会，2018.
(3) 日本学術会議，提言「サステナブルで個人が主体的に活躍できる社会を構築するサービス学」，日本学術会議経営学委員会・総合工学委員会合同サービス学分科会，2020.
(4) 日本学術会議，見解「情報通信分野を中心に据えた産業化追求型（価値獲得型）研究開発プロジェクトの推進」，日本学術会議電気電子工学委員会・通信・電子システム分科会，2023.
(5) S. L. Vargo and R. F. Lusch, “Evolving to a new dominant logic for marketing,” J. Marketing, vol. 68, no. 1, pp. 1-17, Jan. 2004.
(6) 水落隆司，“巻頭言 研究に意味的価値を，”信学誌，vol. 107, no. 4, April 2024.
(7) 戸谷圭子，水野　誠，新井康平，大浦啓輔，大西立顕，石井　晃，加登　豊，小沢佳奈，金融サービスにおける企業・従業員・顧客の共創価値測定尺度の開発 終了報告書，科学技術振興機構社会技術研究開発センター，東京，2016.
(8) 経済同友会，“人間及び人間社会の本質的欲求と企業経営―非連続な環境変化と継続的価値創造―，”第 18 回企業白書，2022.
(9) 仙石正和，“基礎研究を続ける大切さ，”信学誌，vol. 100, no. 6, pp. 431-439, 2017.
(10) J. Plummer, “Educating engineers and scientists for the 21st century,” JUNBA 2013, Jan. 2013.
(11) 仙石正和，“工学教育の変遷と工学研究の広がり，”信学 FR 誌，vol. 9, no. 1, pp. 5-13, July 2015.
(12) JABEE, https://jabee.org/
(13) 仙石正和，小山　純，龍山智榮，米田政明，丸山武男，茂地　徹，長谷川　淳，升方勝己，西村伸也，工学力のデザイン，丸善，東京，2007.
(14) 高木英明，サービスサイエンスことはじめ：数理モデルとデータ分析によるイノベーション，筑波大学出版会，つくば市，2014.
(15) M. Sengoku, Graph Theory and Mobile Communications (Advanced Series in Electrical and Computer Engineering, 23), World Scientific Pub Co Inc, 2023.
(16) 仙石正和，“ネットワークにおけるロケーション問題，”信学誌，vol. 71, no. 6, pp. 568-572, June 1988.
(17) 田村　裕，中野敬介，仙石正和，ネットワーク工学，電子情報通信学会（編），コロナ社，東京，2020.
(18) M. Sengoku, S. Shinoda, and W-K. Chen, “A fault diagnosis in a nonlinear location network,” J. the Flanklin Institute, vol. 323, no. 3, pp. 395-405, 1987.
(19) M. S. Granovetter, “The strength of weak ties,” American J. Sociology, vol. 78, no. 6, pp. 1360-1380, May 1973.
(20) R. S. Burt, Structural holes: the social structure of competition, Harvard University Press, 1992. (安田　雪訳，競争の社会的構造－構造的空隙の理論，新曜社，2006.)
(21) 入山章栄，世界標準の経営理論，ダイヤモンド社，東京，2019.
(22) 小林　司，成末義哲，森川博之，“金融機関の事業性入出金データに基づくサプライチェーンネットワークの構築と中心性分析，”2023 信学総大，D-20-24, March 2023.
(23) M. Coscia, G. Rossetti, D. Pennacchioli, D. Ceccarelli, and F. Giannotti, “You know because I know: A multidimensional network approach to human resources problem,” 2013 IEEE/ACM International Conf. Advances in Social Netw. Analysis and Mining (ASONAM2013), Aug. 2013.
(24) 仙石正和，“(特別講演) 多次元移動通信から多次元ネットワークへ，”信学技報 ICTSSL2023-7, pp. 35-40, May 2023.
(25) M. Sengoku, K. Nakano, H. Tamura, and A. Ohtsuka, “On correlation between dimensions based on neighbors in multidimensional networks,” Proc. ITC-CSCC, July 2024.
(26) 森川博之，“巻頭言 価値創造と価値獲得，”信学誌，vol. 107, no. 1, Jan. 2024.
(27) サービス学会，http://ja.serviceology.org/index.html
(28) 仙石正和，“(特別講演) サービス技術と ICT，”信学技報，ICTSSL2020-6, pp. 31-36, May 2020.

仙石正和（名誉員：フェロー）

1967 新潟大・工卒．1972 北大・大学院博士課程了（工博）．北大助手．1978 新潟大助教授，その後，教授，工学部長を経て，2008 新潟大理事（研究担当）・副学長．2014 新潟大名誉教授．2014－2022 事業創造大学院大学長・教授．2022 事業創造大学院大名誉学長・名誉教授．電子情報通信学会基礎・境界ソサイエティ会長（理事），通信ソサイエティ副会長，英文論文誌 A 編集委員長，日本学術会議連携会員など．電子情報通信学会 1992, 1996, 1997, 1998 論文賞，2004 業績賞，2008 功績賞，1995IEEE 国際会議 Best Paper Award 各受賞．“Wireless Networks”（ACM, URSI） の Editor, The Journal of Communications and Networks の Editor, Journal of Circuits, Systems, and Computers（JCSC）の Regional Editor など歴任．IEEE Life Fellow.