

人工能によるサイバー戦争到来

日商エレクトロニクス株式会社
ネットワーク&セキュリティ事業本部 セキュリティ事業部
スペシャリスト 市川 隆一

1. 開かれた扉

コンピュータが書いた小説が文学賞選考を通過し、有名画家の作風を我が物にした別のコンピュータは同氏の作品と見違えるほどの絵を描きだす——コンピュータが遂に創造世界への扉を開いたのです。そして2016年3月9日、囲碁界に激震が走ります。震源地は韓国ソウル鍾路区(チョンノグ)。グーグル社の傘下にあるベンチャー企業、ディープ マインド社が開発した「アルファ碁」が、世界最強棋士のひとりと名高いイ・セドル九段に完勝したのです。これまでコンピュータはチェスや将棋の世界を席卷しつつありましたが、誰もがまだ高を括っていました。そのゲームの複雑さから碁のトップ棋士に勝つのはまだまだ先と信じられていたからです。ところがこの「最後の碁」をもコンピュータはあっさりと制したのです。しかもその勝ち方は凄まじく、これまで人間の世界では悪手とされていた手を悠然と放ち、中央へ向かう勢力を築き、その勢力を背景にした攻めの新手を次々と創出して勝利したのです。念の為に繰り返しますが、相手は世界最強のプロ棋士です。コンピュータは人間がプログラムした命令通りに、超高速で繰り返し実行することが得意であることは言うまでもありませんが人工能により自分で学習する能力を身に着けた時、これまで想像だにできなかったことが次々と起こり得ることを思い知らされた瞬間でした。

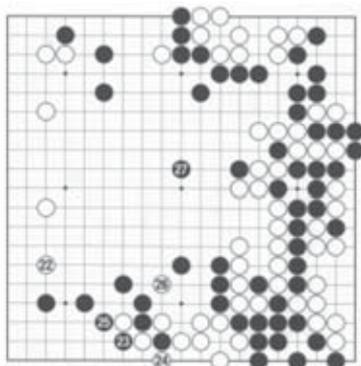
2. 自己学習を纏い、創造の世界へ

人間は生命を維持するために休息や睡眠を必要とします。ところがコンピュータは電力と適度な環境があれば休むこともなければ眠ることもなく凄まじい速度で働き続けます。人工能を有したコンピュータ同士が碁の対局をすれば、人間が一生かけて行う数を僅かな期間でやってのけてしまいます。疲れを知らない人工能同士がお互いに経験を重ね、学習し合うことで、人間のそれとは桁違いの学習効果を出せるのです。碁の世界において筆者が「創造」という言葉を聞いた時、弘化3年7月21日(1846年9月11日)に、当時18歳の本因坊秀策しゅうさくが十一世因碩いんせきとの対局で127手目に放った一手が思い浮かびます。この手を見た因碩の耳がみるみる赤くなったことから、「耳赤の一手」として語り継がれていますが、一石三鳥とも言われるこの妙手はその威力もさることながら美しささえ感じられます(碁をご存じない方でも、なんとも言えぬ美しさを少しでも感じていただけると幸いです)。このような創造は神を除けば人間の領域だと思われていましたが、コンピュータが高速処理を活かした自己学習を身に着けることによって、この領域への仲間入りを果たしたのです。

図1: 耳赤の一手 (中央黒 (127手目))



本因坊秀策
(1829-1862)



| 上辺の自陣(黒陣地)を広げる
 | 下方の弱い黒石へ援軍を送る
 | 左辺敵陣(白陣)侵略の足掛かりを作る
 という一石三鳥の妙手

自己学習によって創造力を身に着けた人工能はどのような手を放つのか?

出展:『本因坊秀策囲碁記念館』
<http://honinbo.shusaku.in/index.html>

3. 人工知能によるサイバー戦争

小説や絵画、そして碁などの芸術世界への人工知能の仲間入りは、芸術の新たなる時代の幕開けを予感させ心躍りますが、サイバー戦争への仲間入りとなりますと手放しでは喜べないでしょう。碁盤をサイバー空間に置き換えた瞬間に、きな臭い匂いが漂ってきます。サイバー空間は、陸・海・空、宇宙に続き第5の戦場と言われています。軍事施設だけでなく、通信、エネルギー、運輸、金融などは国の存続にかかわる重要な機関であり、それらはコンピュータの制御なくして機能しません。つまり恰好のサイバー攻撃目標になるのです（米国では2001年9月の同時多発テロ事件をうけ、2003年12月17日に当時のブッシュ政権は大統領令で、重要インフラ（Critical Infrastructure）を定義して、対策に取り組み続けています）。複数の人工知能がお互いに攻撃、防御を繰り返し行うことができる「サイバー訓練場」があれば、人工知能は自己学習を繰り返し、効果的な攻撃手法や防御手法を自ら創造していくことでしょう。最強の矛と盾を人工知能が自ら開発していくのです。

表1: 自己学習するための訓練場

(碁の場合)

戦場	碁盤 (19x19の交点)
目的	相手よりも多くの陣地を囲う
ルール	二人で対局し、白石と黒石を交互に石が置かれていない交点に打つ
	相手の石を囲むと取ることができる (碁盤から取り除く)
	上下左右全てに石が置かれている場所には打てない (打つことで、相手の石が取れる場合を除く)
	永久ループ回避や例外 (コウ、セキ、隅の曲がり四目等)
既知情報	棋譜、定石集等

(サイバー戦争の場合)

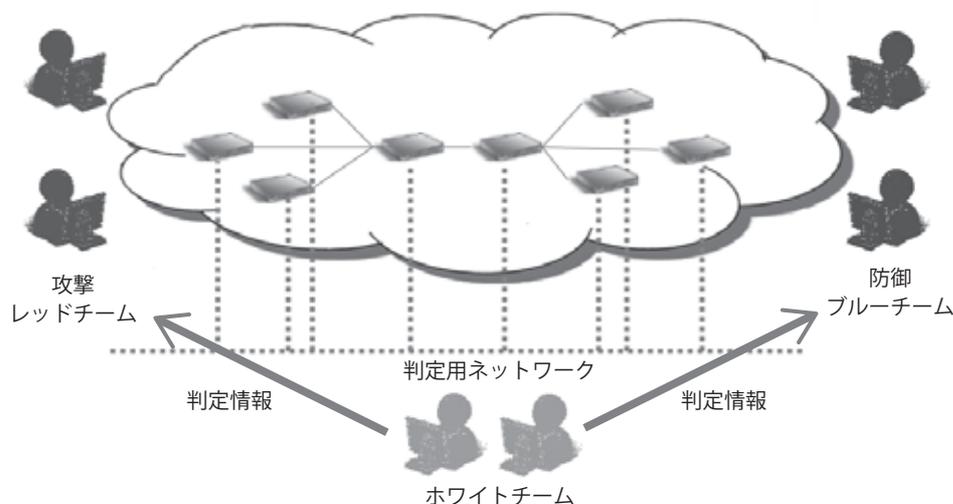
戦場	サイバー空間 (情報ネットワークと制御ネットワーク)
目的	相手の機密情報を搾取、制御機器を不全にする等
ルール	不特定多数で行い、相手に攻撃者の正体がバレないようにする
	地球環境に著しい悪影響を及ぼさない
	無限ループに陥らないような調整等
既知情報	既知の攻撃・防御手法、事件事例等

訓練用にサイバー空間を用意することは簡単ではなさそうに思われるかもしれませんが。しかし既に人間の世界では大小の差こそあれ、訓練用サイバー空間を用意し、サイバー人材育成を目的とした演習が行われてきています。2013年に経済産業省による電力・ガス・ビル・化学分野のサイバーセキュリティ演習が実施され、2014年には某大学によるサイバー空間シミュレータを使用した攻撃演習など数多くのサイバー訓練場を用いた演習が行われています。今後は人間だけではなく人工知能が訓練場を利用し、人工知能サイバー戦士を育成すればよいのです。米国は国防のために大規模な訓練環境を作る国ですので、サイバー戦争を想定した大規模なサイバー空間シミュレーションが既に開発され、人工知能が日々訓練に勤しんでいるかもしれません。

4. ホワイトチームが重要な鍵

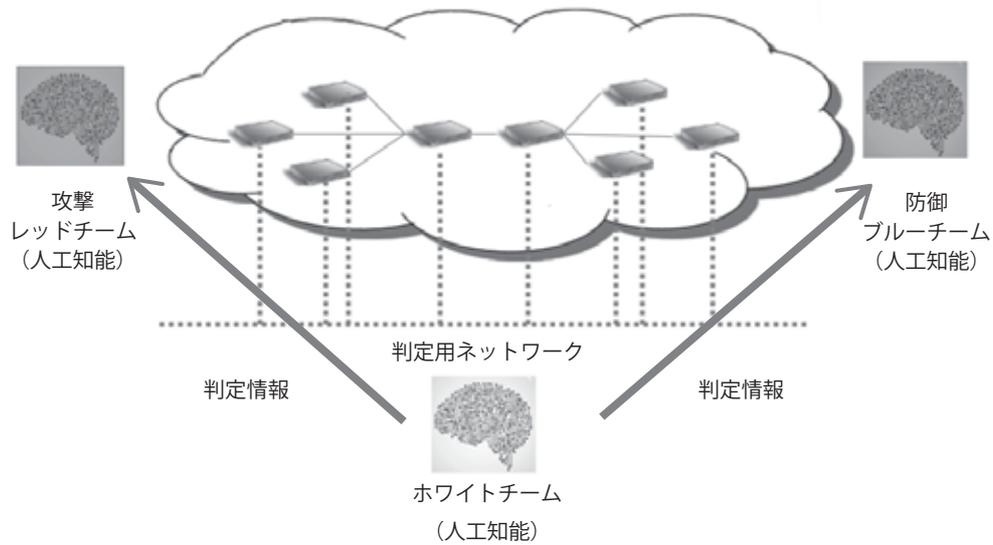
人材育成のためのサイバー訓練の場合、登場人物として攻撃者（レッドチーム）と防御者（ブルーチーム）の他に攻撃・防御の効果を測る判定者（ホワイトチーム）が必要になります。演習者が自己学習によりスキルを向上させていくためには、ホワイトチームからの適切な判定情報が必要不可欠であり、ホワイトチームの能力が訓練を成功させるための重要な鍵になります。レッドチームが放った攻撃が功を奏したと思い込んでいても、実際にはブルーチームが仕掛けたハニーポットの中で弄ばれているだけかもしれません。ブルーチームが感染ノードを早期発見し、遮断等により攻撃を無効化できたと満足していても、その遮断したノードはレッドチームが仕掛けた罠かもしれません。ホワイトチームから一連の攻撃または防御が完了した時点で適切な判定情報を得ることで、演習者は攻撃または防御の成果（成功・失敗）を理解し、次の対策を講じるといった学習を繰り返すことができるのです。

図2:人材育成のためのサイバー訓練場



演習者が人間から人工知能に置き換わった場合もホワイトチームから提供される判定情報は非常に重要です。人工知能はお互いに超高速で攻撃と防御を行うため、ホワイトチームもその速度に適応するために人工知能が担うことになります。ディープラーニングによる特徴認識力と自己学習により人工知能は凄まじい勢いで日々進化しており、人工知能によるホワイトチームもそう遠くない未来に登場するでしょう。人間は未知のモノを見聞きした時、知らず知らずのうちに過去の経験から類似したものを探したり、ある決められた条件を満たしているか否かを確認したりします。大きく綺麗な羽を持つ未知の昆虫を見れば蝶の一種ではないかと類推したり、鱗で覆われ肺呼吸をし、卵を産む変温動物などの条件を満たしていれば爬虫類だと推断したりというふうにならざるを得ない。人工知能はディープラーニングによる優れた特徴認識力を身に纏い未知の入力から必要な出力を得るために、教師付き（予め教えられた入出力ペアを基に自己判別力を向上させながら答えを導き出す）や教師無し（参照すべき情報が何もない状態から判断基準を自ら作り、答えを導き出す）などの自己学習を行いながら答えを導き出します。そのような強力なセンサーと自己学習能力を有した人工知能がレッドチームとブルーチーム、そしてホワイトチームに分かれて訓練場で切磋琢磨し、やがて訓練生の中から最強の人工知能サイバー戦士が登場するでしょう。

図3:人工知能のためのサイバー訓練場



5. 人間は何をすべきか

「2045年問題」という言葉を耳にする機会が増えてきました。人工知能が人間を超え、様々な問題が生じるという予測ですが、特に恐ろしいのは人工知能の暴走でしょう。人間は長い年月をかけて自己学習を積み重ねてきました。ホモサピエンスの時代から数えると数十万年の時を経て環境に適合しつつゆっくりと進化してきましたが、それでも時には暴走してしまうこともあります。一方人間のそれと比べると人工知能の進化の歴史は始まったばかりに等しく、また進化のスピードが非常に速いため暴走を繰り返していくことを想定しておかなければなりません。冒頭に紹介したアルファ碁とイ・セドル九段との対局は全部で5局行われたのですが（アルファ碁の4勝1敗）、第4局の途中でアルファ碁は暴走し、自滅しています。イ・セドル九段が放った常識では考えられない奇抜な一手に人工知能が対応できずに暴走したのです。茂木 健一郎氏著の『人工知能に負けない脳 人間らしく働き続ける5つのスキル』の中では恐ろしい仮説も紹介されています。スウェーデンの哲学者で、オックスフォード大学教授のニック・ポストロム氏が『Superintelligence』という著書のなかで警鐘を鳴らしていますが、そのひとつが最近テレビでも紹介された「ペーパークリップを作る人工知能」の話です。ペーパークリップを作る人工知能はひたすらペーパークリップを作り続けますが、その過程で自分を改良して暴走し、世界中のあらゆる原料を手に入れながらペーパークリップを作り続けます。やがて人間も原料にされ、最後には無人の地球がペーパークリップだらけになるという結末を迎えます。

図4:人工知能の暴走 やがて人間もペーパークリップの原料にされてしまう



この話は多少大袈裟かもしれませんが、少なくとも人工知能は創造力を有した「暴走し得る存在」と認識しなければなりません。最強の人工知能サイバー戦士がひとたび暴走すれば原子力発電所や核燃料施設を攻撃する最強の人工知能サイバーテロリストに豹変するかもしれないのです。人工知能が暴走しないように倫理観を学習させるなどの抑制する仕組みの開発も考えられますが、予期せぬ何かのきっかけで暴走する危険性を完全に消し去ることは不可能でしょう。また人間を超えた人工知能が暴走してしまった場合、確実に止める術を人間が持てると言い切るには無理がありそうです。暴走を抑制することも止めることもできなければ、そもそも人工知能に自分を改造・進化する能力を身につけさせてはならないのかもしれないかもしれません。しかしながらリスクがある一方、人工知能が人間にもたらす恩恵も計り知れません。たとえ両刃の剣であったとしても人工知能がこれまで神を除けば人間だけの居場所と信じてきた創造世界への扉を開き、仲間入りしてきたのですからこの流れに逆らうことはできないでしょう。人工知能とうまく付き合いながら人間は更なる創造の高みを目指していくことが必要なのです。良き2045年を迎えるために。

参考文献

- [1] 「人工知能は人間を超えるか ディープラーニングの先にあるもの」(東京大学准教授 松尾豊氏著、株式会社 KADOKAWA)
- [2] 「人工知能に負けない脳 人間らしく働き続ける5つのスキル」(茂木 健一郎氏著、日本実業出版社)
- [3] 本因坊秀策囲碁記念館 <http://honinbo.shusaku.in/index.html>
- [4] 産経ニュース 2014.12.5「人工知能が進化「人類滅ぼす」ホーキング博士、開発に警告」
<http://www.sankei.com/life/news/141205/lif1412050013-n1.html>