学生フォーラム

第79回　八槇博史先生インタビュー「セキュリティの視点から考える，AIの強みと盲点」

今回は，東京電機大学環境情報学部の八槇博史准教授にインタビューを行った．八槇先生はこれまでにネットワーク上でのトラスト問題（「取引時に相手をどのように信頼するのか」という問題）やマルチエージェントシステムに関する研究を行っており，現在ではさらにネットワーク，セキュリティ，人工知能などの分野を学際的に研究されている．

本インタビューでは，ネットワークやセキュリティと人工知能との関係を中心に，現在行っているマルウェア*1にAIを組み込む研究について考えを話していただいた．また，AIとセキュリティという“境界分野”における，「セキュリティ分野の研究者とAI分野の研究者における認識のギャップ」についても話していただいた．

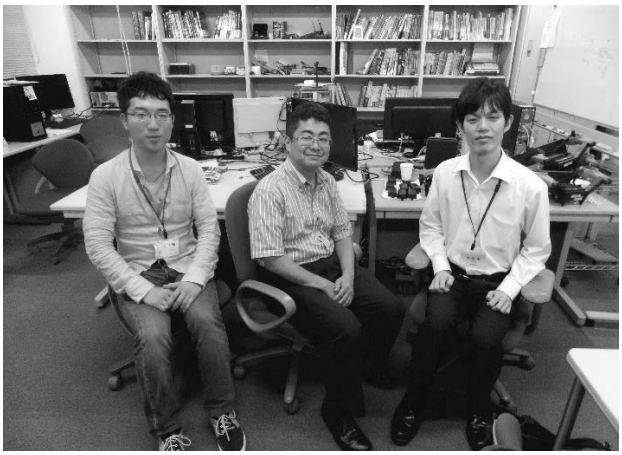

図1　八槇博史先生（中央）と並んで

――先生の原点であるネットワーク上での「トラスト問題」と，その解決策について教えて下さい．

インターネット上の店で物を買うのはある意味博打です．インターネットというオープンなネットワークで，互いの素性がわからないところで相手を信用して取引を行います．特に，初めて買い物をするときにどのように相手を信頼するのかが難しく，どうすれば安全に取引が成立するのかを京都大学で助手として働いていたときに研究していました．

取引をしていたら何かしらの信頼は必要になります．それを支援することで買う側も売る側もより安心して取引することができるようになります．相手を信用するために双方が何かしらの情報をもち，そのもっている情報を交換します．交換した情報をもとに相手を評価・信用します（オートメーティッドトラストメーション）．しかし，何でもかんでも交換するのはあまり良くありません．例えば町の中で突然「私は警察だ，免許証を見せろ」と言われて，あなたは見せることができますか？……一目で警官かどうか判断できなければ，相手が本当に警察かどうかを確認するため，まずは「警察手帳を見せて下さい」という話になるでしょう．実際に警察手帳を見せてもらえれば，一応は警察だと信じることができます．このように，相手を信用するにはある程度の情報開示が必要です．しかし，開示する情報は最小限にしたい．そこで，開示する情報は最小限にしつつ安全に取引するためのプロトコルとして，コンピュータが自動的にネゴシエーションするエージェントを実現する研究をしていました．

エージェントというのは，人間の代わりにAIがいろいろやってくれるというものです．しかし，人間がAIにどこまで物事を任せられるのかという問題は，たぶん1990年代と今では答えが違うと思います．例えば自動運転の車に関していえば，90年代だと「そんなものに人間を乗せるなんて考えられない」といわれていました．しかし，今では自動運転の車は実現できそうだから法整備をしなければならないという話になります．そう考えると，時代変化に対応する固定的な“答え”はありません．意外と，人間ができることをすべてAIに任せられるような世界が実現するかもしれませんし，そうではないかもしれません．そこは見定める必要があります．

AIの研究者としては，「すべて任せられる時代が来る」ともてはやすのも良いかもしれませんが，冷静に考えるとそうとは言い切れない部分もあるのではないでしょうか．

――徐々にAIに任せる流れになっている気がしますが，AIに物事を任せることに危機感はないのでしょうか．

AIに物事を任せることに関しては，昔ほどの危機感はなくなっており，任せてしまう人は増えていると思います．しかし，昔よりも安全になったとは言い切れません．どちらかというと，ユーザ側の抵抗がなくなってきたのではないかと思います．

こうした中，ある意味「相手が信用できるか」は永遠の課題として残っています．この点を解決するには，オートメーティッドトラストメーションだけでは不十分で，その背景にある社会的制度もトラストには必要になりま

*1 コンピュータウイルスなど，悪意をもったソフトウェアの総称．

す．例えば，クレジットカードの場合，詐欺にあったときにその支払いを免除してくれる保険の仕組みや警察制度などへの信頼がすべて集まってトラストというのがあります．それを捉えるには昔のやり方はナイーブだったという反省があります．

――その反省は今，どのように生かされているのでしょうか．

AI に関する研究としては，LIFT プロジェクトというのを佐々木良一先生（東京電機大学）と共同で行っています．特定の組織を狙って入ってくる標的型攻撃では，標的に特化した専用のマルウェアがつくられて送られてきます．そのようなマルウェアが暴れだしたときに，「何か変なことが起きているけれど，今何をされているかわからない」という状況が想定できます．こうしたとき，ログやパケットの様子から，攻撃を受けているのではないかと疑うことができます．しかし，どのような攻撃であるかを特定することは難しく，「もしこの攻撃だとしたらこのあたりを調べる」というようにして，問題が起こっていないか，複数のシステムを順番にチェックしていかなければなりません．これを自動化するような AI の開発を LIFT プロジェクトで行っています．用いている AI 技術としては，かなり古典的なプロダクションシステムで，かつてのエキスパートシステムのエンジンを用いています．

実際に何をやっているかといえば，例えば従業員のパソコンにある種のマルウェアが見つかった場合，認証サーバが攻撃されているかもしれないということで認証サーバに変なログがないかチェックします．攻撃が成功している可能性があったならば，次は「ファイルサーバが攻撃されているのではないか」もしくは「データベースが攻撃されているのではないか」というように，想定される可能性を順番に調査します．このような場合に必要となる思考や知識を取り込むために，かつてのエキスパートシステムの知見や故障診断の知見がまるまる使えるのではないかという仮説を立てて始めたのが，LIFT プロジェクトです．

――具体的には何を実現できるのでしょうか．

実は，似たようなものに SIEM[*2] というのがあります．しかし，これを使いこなすには専門家の知見が必要になります．そこを機械でサポートすることにより，人間の負荷を下げ，調査範囲を広げたりするのが LIFT プロジェクトです．また，受けた攻撃を分析する「後追い」の防御策だけではなく，将来起こり得る攻撃を予測して事前に防御策を立てたりもします．

攻撃者はランダムに攻撃しているわけではなく，何かを狙って攻撃しているはずです．そこで，「攻撃者が狙っているものは何か」を考えつつ，将来起こり得る攻撃を予測していきます．具体的には，あり得そうな攻撃のセットを事前に生成しておいて，実際に予想した攻撃が発生しても検出できるようにするために，GA を使ってマルウェアの攻撃を自動生成する手法を研究しています．Genetic Algorithm（遺伝的アルゴリズム）でもよいかもしれませんが，まずはパラメータ調整といったアルゴリズムの範囲でどこまでできるのかを検証しています．

――なぜ GA でマルウェアの攻撃を自動生成できるのでしょうか．

攻撃者はマルウェアを作成する際，一からプログラミングしているわけではありません．既存の部品の組合せと，パラメータやプログラミングの並び，暗号化のアルゴリズムを少しだけ変えてマルウェアを生成しています．何億という種類のマルウェアが存在しているといわれていますが，実際には各種の組合せで爆発しているのです．組合せで数が爆発しているのであれば，GA の枠組みで追いかけられても不思議はありません．そう考えると，GA で爆発させても同じではないかと考えました．

GA を爆発させるような環境も 10 年かけてつくっていたので，それをうまく組み合わせてマルウェアにおける攻撃の先回りを始めています．

――AI が組み込まれたマルウェアは，具体的にどのようなことができるのでしょうか．

AI が組み込まれたマルウェアでできることはたくさんあると考えています．今はまだ，攻撃者が外からマルウェアを操作しています．言い換えると，マニピュレータが外から伸びてきている感じです．このマニピュレータが AI に変貌した場合，その瞬間からマルウェアの検出が困難になります．

マニピュレータが外にあるため，コントロールとマルウェアの間でメッセージのやり取り（C&C サーバとの通信）を行う必要があります．多くの場合，このメッセージを検知することでマルウェアを検出しています．しかし，マニピュレータが AI になると，マルウェアは自分自身をコントロールすることで，メッセージが必要なくなり，検出が数段難しくなります．ほんの一例にすぎませんが，少なくとも AI を積んだマルウェアは一定のレベルに達すれば攻撃者にとっての有用性が高いことがわかります．

――なぜ，あえて攻撃側に AI を組み込もうと思ったのでしょうか．

まず，セキュリティにおいて AI で何ができるのか考えました．守る側から発想すると，観測エージェントがいて，それを監視するという感じになります．既知の攻撃に対してはこれでも良いかもしれませんが，将来起こ

---

*2 サーバやネットワークといった複数の機器から各種ログを収集・集約し，相関分析することで攻撃を見つけ出す製品．

り得る攻撃を推測するためのアプローチとしては，これで何ができるのかイメージはつかめません．そこで，あえて攻撃側から発想してみたのがきっかけでした．攻撃側からどうするかを考えるほうがAIとしても考えやすく，まずは“悪役”として進化するAIをつくるのが第一の研究目標です．

プランニングで考えてみると，「多種多様な攻撃を順番に試し，最終的にデータベースの情報を盗んできて下さい」というのは計画しやすいです．反対に，「そのような攻撃からシステムを守って下さい」と言われても，攻撃は多種多様にあるため，どういう計画を立てればよいのかよくわかりません．そのため，いきなり守るための製品をつくろうとしてもポイントが絞れず，「攻撃の種類によらず網羅的に防げばよいのではないか」となってしまいます．攻撃側が何をしてくるのかを的確に推測するため，攻撃側のAIを用いて考えるほうが今のところは考えやすいのです．

――今後の時代変化を考えたうえで，どのような攻撃が発生すると考えますか．

ロボットや自動運転の車といった，自律的に動く機械が人間と交じり合う社会が実現しつつある，というのが人工知能学会的な見方だと思います．ところで，そのロボットが攻撃を受けたらどうなると思いますか？……中身がコンピュータである以上，攻撃を受ける可能性は大いにあります．

最近開かれたBlackHatというハッカーの祭典では，自動車や飛行機のクラッキング*3が発表されていました．攻撃者はそのようなクラッキングが楽しくなってきている頃だと思います．攻撃を受けるからといって今さらやめることはできないので，どうするのかは考えたほうが良いでしょう．

――今後の社会的な変化を考えるうえで，AIとセキュリティに最適な関係はあるのでしょうか．

AIとセキュリティの最適な関係はまだわかりませんが，少なくともセキュリティ分野の人が考えている「これは危険，これは安全」という基準はおそらく厳しすぎます．それをすべて言っていると何もできなくなってしまいます．できなくなるからといって，全くのノーガード戦法で良いかというと，もちろんそういうわけにもいきません．そこで，このバランスを調整する役割としてAIが役立つと信じています．しかし，「セキュリティとAI」といっても，国内での受けは今一つで，なかなか同志が集められずどうするべきか悩んでいます．

――AIとセキュリティの研究は，注目されているように思いますが，それでも受けが良くないのは何か特別な理由があるのでしょうか．

AI分野で発表されているセキュリティ関係の技術というのは，セキュリティ研究者のセンスで見るとまだまだ甘く，現実感のない話が多いです．反対にセキュリティ方面で発表されるAI関係の技術を見ると，AIとは到底呼べないようなレベルの話が多くあります．両者のかみ合わせはこれから高度になってくると思います．重要性を認知しているところは結構あり，安全保障関係の学会では，「セキュリティとAI」という境界分野は重要であるという論文がここ数年で出てきています．

セキュリティ分野でのAIとしては，傾向を抽出して警告を出すタイプの技術があります．しかし，論理や推論といった古典的なAIとして考えられてきた部分は，まだ活用されていません．どちらかというとビッグデータが先行しており，ネットワークのトラヒックログをディープラーニングするといった話のほうが多くあります．

逆に，AI分野でのセキュリティは，モデル化をしすぎているように感じます．マルチエージェントが盛り上がったときにあった話ですが，攻撃トラヒックを観測するエージェントがネットワーク上にたくさんあり，それらが相互に情報交換して攻撃を検知するという研究がかなりありました．しかし，こういった論文に書かれているネットワーク構成は現実的ではなく，観測できない場所にエージェントが置かれていたりしています．

エージェントの情報交換方式の設計がベースにあって，それに合わせてネットワーク構成が設計されている論文は多くあります．しかし，実際のネットワーク構成にAIが対応する形になっている論文はあまり見られません．そのため，AI方面で発表されるセキュリティ関係の論文は，やや宙に浮いている感じがします．ワークショップなどは立ち上がっていますが，AIとセキュリティという境界分野としては，きちんと成立している感じはしません．

――これからAIとセキュリティの研究を始める人に求めることは何でしょうか．

やることは大量にありますが，切り口がきちんと見つかっていないため，一緒に考えてくれる人を求めています．とはいえ，セキュリティは犯罪と隣り合わせの側面があるので，扱う人には高い倫理観が求められます．

例えば，あるソフトウェアのぜい弱性を見つけたケースで考えると，その情報をソフトウェアの開発者にフィードバックすれば世界は安全になります．しかし，悪人に売れば世界は危険になります．開発者に渡した場合の報酬と比較すると，悪人に渡したときの報酬のほうが2～3桁ほど額が高い場合があります．そのときにぜい弱性を発見した人は，開発者に渡せるかどうかがセ

*3 悪意をもって，他者のコンピュータやシステムに不正にアクセスする行為．また，改ざんや破壊を行う行為のこと．

キュリティ業界においては求められます．

犯罪的なことをするつもりは全くありませんが，手口がわかっていないと防ぐこともできません．そのため，セキュリティの研究をするのに倫理的な問題というのは常にあるため，今から研究を始める学生に対しても高い倫理観を求めます．

――インタビューを終えて

機械学習・深層学習を用いて攻撃を検知するような防御側に AI を応用した研究は進んでいるように感じるが，攻撃側に AI を応用した研究という非常に目新しい話を伺うことができた．今はまだ防御側が弱いために攻撃者も高度な攻撃を用いる必要がないが，もしかすると，すでに高度な攻撃をする準備が攻撃者にはできているかもしれない．歴史的な背景から見ても，防御側が強くなった途端に高度な攻撃が行われる可能性は高い．そのような攻撃が行われだしたときに防御側として何ができるのか，また何をしなければいけないのか．攻撃と防御は“いたちごっこ”だといわれ続けているが，この関係に終止符を打つためにも，AI が重要な役割を担うことを認識させられるインタビューであった．

〔久山　真宏（東京電機大学），林　侑輝（千葉大学）〕