

AITCセミナー 「AI技術鼎談」 ～先端ソフトウェア技術者とAI～ サマリー

2017-04-26

(国研) 産業技術総合研究所
和泉 憲明

【AITCセミナー&オープンラボ】

18:30～18:40 オープニング

18:40～19:00 講演1:

「AIとデータ循環」橋田浩一先生

19:00～19:20 講演2:

「演繹から帰納へ:情報技術のパラダイムシフト」丸山宏先生

19:20～19:35 休憩

19:35～21:00 「AI技術鼎談～先端ソフトウェア技術者とAI～」

- 100年前から見た今
 - シンギュラリティ以上の期待が...。
 - 鉄腕アトム、ドラえもん、...、ガンダム？ターミネーター？
- 自律適応サイクル(仮説検証サイクル)
 - 価値 ... 意味(仮説)の確認として、偏在性がある！
 - 学習や進化、の他に...。
 - 思いついたこと(確からしいことを)、実践してみる！
 - 多層 vs サイクル
 - 今のAIに足りないモノ？
 - これからの情報技術の方向性！
- 対称性バイアス
 - 人間だったらできる機能：メタ戦略

- AIの基盤としてのBPR
(← 社会がシフトすべき方向性？情報技術戦略？企業戦略？)
 - AIの研究開発よりも実運用
 - 構造化されたデータが容易にアクセスできるために！
 - 研究開発のコストはほとんどデータ整備
 - AIによるサービスの相手(個人)に関する良質なデータが必要
 - (次世代の)BPRは、データ循環社会のために！
 - 社会全体でのデータ循環
 - 意味構造化(標準オントロジー)されたデータの循環
 - 意味理解が必要だが、AIによるBPRの自動化は...
 - データ主体の意思によるデータの流通
 - 手に入るデータと、今後、必要なデータ
 - 個人消費がGDPの60%(リッチなパーソナルデータが必要...の根拠)
 - 企業データから個人データへ!?

- PDS にみる近未来
 - XML前夜としてのGDA (<http://i-content.org/gda/tagman.html>)
- AIには意味理解ができない: サイクルがない
 - 社会レベルで循環していれば良い!
 - パーソナルデータを本人の意思で流通させれば、技術者がより健全に活躍できる社会が到来する!?

- 企業・研究所の技術戦略
 - IBM東京基礎研究所
 - 統計数理研究所
- AI、機械学習、統計、...。
 - 汎用AI、特化型AI
 - AIとは「ITを進化させる営み」
 - 第一次AI: GC, コンパイラ、...。(出来るとは思っていなかったこと)
 - 第二次: オブジェクト指向、(セマンティック)Web
 - 第三次: 統計的機械学習、深層学習、...、機能的プログラミング
 - 総務省の方針には...

- 演繹的プログラミング
 - モデルベースの思考：設計仕様を理解する...
 - ←従前の技術者の思考
- 統計的なモデリング
 - 過学習に陥らずに、適切なモデルを選択する教養(能力)
 - Cf: 深層学習はどのように動くか？
 - (あまり)過学習しない！...なんだかわからないけど...
 - 技術者の態度としては...？
 - これを日常問題に適用すると？

- 演繹的(還元論的)思考(システム開発?)の近未来
 - Cf: 自動運転のデモ
 - 273次元のインプット⇒5ビットの出力
 - 乱数がなぜ必要...「わかる人」にはわかる理屈
 - これからのシステムの「要件定義」とは？
 - 訓練データセットを与えているだけ？
 - ロボットの二足歩行、自動車の自動運転、...
 - 強化学習によるシステム開発
 - センサー入力に対してアクチュエータ出力
 - 要件定義: 結果に対してフィードバックすること

- システム開発のパラダイムシフト
 - ものづくり
 - フロントローディング
 - AIoTの世界
 - やってみて、要求に合わなければ修正しよう！
 - アジャイル、DevOpsの先にあるもの？
- 統計的機械学習の本質的限界
 - 見たことのないデータに対して無力：内挿、外挿。
 - 機械学習はひらめかない、臨機応変な対応をしない。
 - 本質的に確率的
 - 元分布・訓練データ(ランダムなサンプリング)・学習済みモデル
 - 100%の保証は原理的にできない！
 - おそらく、日本の鉄道会社が自動運転に興味を示さない理由。

- 誰が機械学習を計画するのか？
 - 訓練データの準備、学習、推論、...。
 - Cf: ソフトウェア危機...⇒ ソフトウェア工学
 - 機械学習工学: 帰納的にシステムをつくる方法論(技術体系)
 - 要求工学
 - MLの限界と、データセットの準備
 - ツール
 - IDEとリポジトリ
 - テスト
 - リグレッション
 - 運用
 - モデルの賞味期限

【AITCセミナー&オープンラボ】

18:30～18:40 オープニング

18:40～19:00 講演1:

「AIとデータ循環」橋田浩一先生

AI技術は、入手可能なデータに依存していたのでは？

これからは、どんなデータが入手可能になるのか？

19:00～19:20 講演2:

「演繹から帰納へ：情報技術のパラダイムシフト」丸山宏先生

情報技術部門はどういう技術に牽引されて行くのか？

AIブーム ::= 情報技術 | AI | 機械学習 ?

19:20～19:35 休憩

19:35～21:00 「AI技術鼎談～先端ソフトウェア技術者とAI～」

- AIを歴史的に眺めると...。
 - 自然言語処理は、文法から研究していたのでは？
 - しかし、今のNLPは、データ駆動!?
 - AIの本流は、意味論・モデル論にあるのか？統計にシフトしたのか？
 - 今は、統計的AIの成果を注意深く、洗いざらい、使いこなす「べき」時期。
 - 人間の知性に対して、統計的な側面は半分程度では？
 - 知識、フレーム理論、オントロジー、...(知識や理解に関する方法論も、十分成果はあったが...)
 - » 人間の理解を可視化する方法論「だけ」では不十分であったことも事実。
- 100%を保証しないのはAIだけではないのでは？
 - そもそも、100%を保証できていたモノは？

「次の」AIは、どういうものなのか？

– では、「意味を理解する」とは？

- (仮説検証の)サイクルに基づく知的な活動: 人間の熟考を、機械ではなく、プロセスとして
 - サイクルのなかで、記述的に定義されて(明確になって)行くもの...?
 - » ボルツマンマシンのセマンティクスは定義できるか????
- 「人間と機械の共通言語」のあるべき論
 - 集中型でなく、分散型のAIに対して...。
 - » ※和泉注釈: 自然言語が人間と人間を協調させるための手段であって、かつ、プログラミング言語が機械と機械を動作させるための手段であるとしたら、人間と機械が正しく協調できるようになるための「言語」として「これからのAIパラダイム」を考えるべきではないか? これまでの研究・開発は、自然言語理解やプログラミングとして行われてきたし、協調AIなどもパラダイムとして存在したが、人間と機械を包括したような系として(社会インフラを含めて)情報システムが最適化されるための方法論についても学ぶべきであったか?

– ⇒ 行動経済学的な説明

- 人間は、#1:直感的に思考する & #2 努力しながら考える、という両面がある。
- ⇒ #1はDLが得意なところ & #2は(今の)AIが苦手なところ。
 - 知性とは、合理性とは?...。論理的な思考はAIの基礎か?
 - ← 今は、批判的な立場から、バイアスがかかっているのか?

「次の」AIは、どういうものなのか？

- (これからの)AIは、(意味理解を指向する)AIを目指すのか？
 - AI(というシステム)が社会に埋め込まれているべき
 - 社会レベルの最適化が重要
 - » ※和泉注釈: 丸山先生のコメント(見立て)を考えると、(100%否定はできないものの、ある程度の近未来では)今後、驚くような新規の発見(相対性理論的な、「 $P \neq NP$ の証明」な...)がないと想定されるので、現状の技術の改良にブレークスルーを期待することは難しい。だとすると、今、手に入るAIの(基礎)理論に基づくシステムが、最大限活躍できるような社会構造の構築を目指すべきではないか？
 - » ⇒ 鮮度の良いデータが社会基盤になるような、Web以上の仕組みがあれば、結果として、AIのブレークスルーになるのでは？
 - 究極の宣言型プログラミング(ゴールを示せば、AIが解く)
 - 過去は上手くできない、ブレークスルーは必要。
 - » ※和泉補足: 論理的な思考の限界(←すなわち、どういうブレークスルーが必要かという観点)は、例えば、制約充足系のシステムでさえ上手く動いていない、という現実にある？
 - ←ゴールを与える、という技術の難しさ: そもそも、何に価値があるのか、事前に全てがわからない(書けない)！
 - ←シンギュラリティが達成できない理由でもある!? ⇒ 価値は、そもそも、人があたえる「べき」もの。
 - 逆に、機械が価値を作り出すとしたら、人間は機械の活動を制御できないことになる。(価値は記述できない)
 - ←事前記述ができないこと、と、価値が与えられること

「次の」AIはどこに向かうのか？

- 直感と熟考は、対立すべきでなく、融合すべき
 - たとえば、データ同化という方法論(パラダイム?)はすでに確立している。
 - ←システムのアウトプットに関する)予測と(そのシステムのパフォーマンスの)チューニング。
 - 大局的に見れば、融合された活動だが、レベル間が全く異なるので、混同に注意！
 - » 技術者・研究者のチューニング
 - » 学習アルゴリズムを固定したチューニング
- DLでできないことが、機械でできるようになるのか？
 - 今、技術者に必要なことは...。
 - (今の時代に)最適なこと(方法論)が(今後の社会で)最適であり続けるはずがないのでは？
 - » 今、なんだかわからないことでも、次は、よくわかるようになるはず！
 - » 実践の過程で、何らかの理論・工学に洗練されてゆく！
 - 仮説検証 vs 試行錯誤
 - » パラダイムが固定できれば仮説検証。
 - » パラダイムにとらわれない思考(仮説検証)は、試行錯誤に見える。
 - 例えば、...。
 - » Alpha GO が示したモノ: ZeroSumGame は、AIにかなわない。
 - 対話・協調のインタラクション: 高次のアダプテーション
 - » そもそも、過去に革命的なブレークスルーが計算機科学にあったっけ？
 - アイディア(理論・モデル)は、昔からあった(ものしか役立っていない)よね。
 - ⇒ (先端技術者は)見えない未来を見ようとするのではなく、目の前の技術に取り組む(べき)!!
 - » もしかしたら、機械が機械をチューニングできるようになるかも？(現時点では、難しい...)

- システム同定としてのAI
 - 機械学習工学と訓練データの提供
 - 訓練データそのものが大事ではなくて、
 - 実運用で使える(実運用に耐える)データが大事。
 - そういう大事なデータ... 経済活動に「相当」するデータ
 - ビジネスとチューニング
 - » Web上のデータは、意図的に(特定の研究者・技術者が、個別の目的のために)作られたデータではなく、「実際の(生きた)」データ
 - 多くの人の日常生活に「直結」したデータ
 - 生きたデータ ← 意図的に作られたのではないデータ:
 - » 入力データと出力データを区別しなくて良いデータ

- 「データ循環のなかで、価値が発現する！」ということですか？
 - 「価値の定義」って出来ることなんですか？
 - 「仕様(ゴール)の記述」が、そもそも、難しいものが、次のAIの対象。
 - 機械学習エンジンのチューニングは、「暗黙的な価値や仕様」に対するアダプテーション
 - ← 機械学習工学としての「つきあい方」
 - ⇒ モデル思考ではなく、結果に対して議論せざるをえないほど、技術が進化してしまっている？
 - 「技術の進化と向き合うべき技術者」とは、「もう少し」過分に理解する技術者では？

- 強化学習は、リアルタイムになるのか？
 - マクロにみえれば、強化学習は、一種のサイクリックな活動である。(仮説かな...？違うなあ～)

- 人間とAI:リアルとサイバー:データがブリッジするのか？
 - 運用を設計するとは？それはエコシステムの完成なのか？
 - 「エコシステム」が目指すべきものは、「生きたデータ」の鮮度を保つ「社会的な仕組み」

後日記(和泉感想)

- 事前に頂いた質問は、次の3つに分類できました。これについて、今回の鼎談を後日記としてまとめてみました。
 - AIの可能性 ... AIの限界
 - 機械学習の方法論(アルゴリズム)に対して、訓練データや計算資源がリッチになった効果が「AIの成果」として得られている。訓練データを与えて、事後に機能(方法論)をチューニングする、という形(すなわち、演繹から帰納)にエンジニアの思考はシフトすると考えられる。
 - 技術者・研究者の教養・学習
 - 機械学習(深層学習)のツールをブラックボックス的に扱うのではなく、訓練データの特性や影響など、注意深く思考できる能力が求められている。今後も、機械学習を用いたシステム構築について、要求分析からテスト、運用までを「(上位プロセスの)設計」できる能力・人材が求められる。
 - オープンイノベーション、シンギュラリティ
 - 層的な構造を重ねても、次のブレークスルーや、シンギュラリティは、難しいと考えられる。(例えば、ソフトウェアが自分で自分を省みて修理する、のように?) 仮説検証を自律的に推進できるような「サイクル」が、(機械的)システムとして実現される必要がある。(そういう研究・開発が重要ではないか?)