



SEMINAR REPORT

ロボットは東大に入れるか



名古屋大学
准教授
松崎 拓也 氏

2011 年に国立情報学研究所という計算機関係、ソフトウェア関係の研究所を中心として、「ロボットは東大に入れるか」という長期プロジェクトがスタートしました。何をやるプロジェクトかという、AI ソフトウェアに大学入試問題を解かせようというものです。タイムスパンは概ね10年です。まずは2016年度末時点にて、センター試験で高得点を取ろうというのを掲げています。最終目標としては、AI ソフトウェアで2021年度に東大入試で合格レベルの得点を取るという目標を立てて研究を進めています。

私は2012年まで国立情報学研究所にてこのプロジェクトに携わっていましたが、昨年名古屋大学に異動した後も、共同研究者としてこの目標に向かって研究を進めています。本日の講演は、このプロジェクトについて紹介したいと思います。

大学入試センター試験

センター試験の問題を幾つか見てみましょう。図1は世界史の問題です。世界史でもっとも典型的な問題は4つの選択肢問題で、このうち正しいもの、あるいは間違っているものを一つ選びなさいという問題です。正答は③です。これは比較的難しい方の問題ではないかと思います。

考えてわかることではないので、ある種の情報源を用意しておいて、それを調べに行くことになります。ここでの問題は何かという、完全に同じ書き方で書いてあるテキストはほぼ存在しないことであって、最大の難題です。もっとも、この問題には明らかなキーワードがあって、比較的手がかりは得やすいで

す。人名であるとか、三国干渉、ある種の事件、事件ではないがイベントの名前であるとか、計画の名前、人名、会社名などのキーワードや、いつの話であるのか、どこの話であるのかといったキーワードであって、これらは比較的明白です。これらのキーワードが的確に抽出できれば、必要な情報をテキストや教科書、あるいはウィキペディアから探してくるのは、それほど非現実的な話ではないです。これは比較的イメージしやすいことだと思います。

産業の振興や国土の開発について述べた文として正しいものを、次の①～④のうちから一つ選べ。

- ① オーストリアでは、19世紀後半に、宰相メッテルニヒの下で工業化が推し進められた。
- ② イギリスは、19世紀後半に三国干渉を主導し、中国の国土開発への参入を図った。
- ③ ソ連では、20世紀前半に第1次五か年計画によって、重工業に重点を置く工業化が推し進められた。(正答)
- ④ 中国では、20世紀前半に、毛沢東によって人民公社が設立された。

2005年度センター試験 世界史B

図1 センター試験問題(世界史)

次に、国語の問題はどうでしょう。図2に示すように、本文は概ね3,000字ほど、ページ数にすると4～6ページ程度というのが、センター試験の国語の典型的なフォーマットです。

本文：…ヨーロッパ流の芸術観では、芸術とは自然を素材にして、それに人工を加えることで完成に達せしめられた永遠的存在なのだから、A造形し構成し変容せしめようという意志がきわめて強い。それが芸術家の自負するに足る創造であって、それによって象徴的に、彼等自身が永生への望みを達するのである。…

問2 傍線部A「造形し構成し変容せしめよう」とあるが、それはどういうことか。その説明として最も適当なものを、次の1～5のうちから一つ選べ。

1. 変化し続ける自然を作品として凍結することにより、一瞬の生命の示現を可能にさせようとする。
2. 時間とともに変化する自然に手を加え、永遠不変の完結した形をそなえた作品を作り出そうとする。

2007年度センター試験

図2 センター試験問題(国語)

一番典型的なのは、本文が与えられて一部に傍線が引いてあって、これの説明解釈、あるいはこれの言い換えに当たる選択肢を選びなさいというタイプの問題で、最も多く出題されています。

この問題の特徴は何かというと、何を言いたいのかよくわからない本文に対して、よくわからない言い換えがなされています。これはある意味で必然であって、日本人が受けるテストなので、両方ともわかりやすい文章を出題したのでは全く差別化ができない。つまり、ある種の訓練を受けていないと読めないタイプの本文が出題されて、それに対して非常に特殊な、よく考えないと言い換えになっているかどうかかわからないような選択肢が与えられるという、極めて特殊な出題形式となっています。先ほどお見せしたような、社会科のテストとはかなり異なったテクニックが必要なことがわかりだと思えます。

もう少し見てみましょう。英語はとても多様なタイプの問題が出題されています。図3はその中でもいささかおもしろいタイプの問題です。二人の人物が対話をしていて、対話の一箇所が空白になっています。ここに入る最も適切な発話、発言を選択肢から選びなさいという問題です。英語か日本語かというのは問題ではないので、ここでは日本語に翻訳したものでお見せしています。

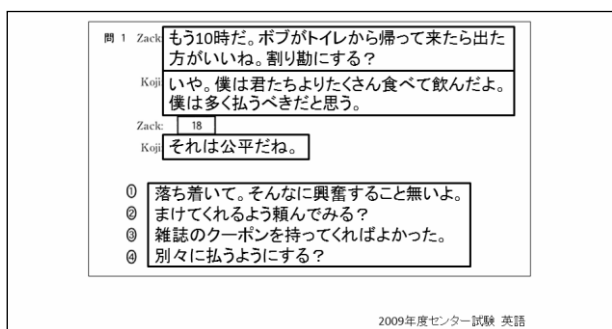


図3 センター試験問題（英語）

何を対話しているかというと、飲み食い後の支払い方法についてであり、日本語に訳してしまえば、日本人であれば、ほぼ100%解ける問題です。逆に原文の英語のままであっても、英語圏の小学生であれば、ほぼ100%解けるだろうということで、これは切り口としてはおもしろい問題設定となっています。

では、どうすれば計算機がこの問題を解ける、どのようなアルゴリズムを組めば、この会話の流れが自然であると判断がつけられるのか？これがとても難しい問題であることは、容易に想像がつくことと思います。人間であれば対話文に空欄があっても、この一連の会話を読んだ時点で、彼らはレストランにいて、食事は終わっていて、これから起こるイベントは支払いであると想像がつく。そして、その支払いの方法は大きく二通りあって、複数の人が同額支払うか、あるいは食べた量、飲んだ量に比例して支払うかのどちらかというわけですが、どちらの方法が公平に分かるためには、もの凄い量の概念や知識を予め持っていなければならない。そのような知識があって、初めて一連の対話の自然さがわかるような問題となっています。これもまた問題設定の中では、比較的難しい方の部類に入るといってよいでしょう。

人口知能分野の細分化

まず、研究の歴史認識として、現在の人工知能関連の研究分野は、細分化されているというのが現実問題として存在しています（図4）。これは恐らく1980年代頃に、かなり細分化されたというのが教科書的な見方です。それまでの人間に近い人工知能というのは、20年ほどで完成するのではないかという割と素朴な期待が裏切られて、これは一度目のAIにおける挫折ですが、それ以降は、もう少し問題をブレイクダウンしないとどうにもならないという反省が恐らくあったのだと思います。現状では、例えば、言語処理だけを研究対象にするとか、機械翻訳で自動翻訳だけを研究するとか、画像の処理だけとか、音声を入力してテキストに書き起こす部分だけを精度を高めるとか、自動推論、ルールに基づいたある種の演繹をする、数式を解くなど、どれもこれもバラバラに分かれているというのが、人工知能分野全体を見渡したときの現実であるということです。

- 言語処理
- 画像処理
- 音声処理
- 自動推論
- 数式処理
- ...
- 全部つなげたときに本当に問題が解けるのか？
- そもそも接続可能な要素技術になっているのか？
- 分野の細分化から再統合へ
- 「知能とは何か」
- 人工知能と次世代社会

図4 人工知能に関する分野の細分化

それぞれの分野で、それぞれの蓄積があったことは間違いないのですが、ここで問題にしているのは、この別れ別れになった分野での蓄積を端から端まで繋げると、すなわち、音声で入力し、言語内容を理解し、画像の理解と合わせた上で推論を行って答えを出す、あるいは、その結果を英語に翻訳して答えを出せる研究が進んでいるのかというのが、このプロジェクトの動機となった質問の一つです。それぞれ研究は行われているが、それを繋げると本当に使える技術になっているのかというのが大きな疑問です。

もう一つ、「ロボットは東大に入れるか」というとても具体的な目標を立てました。これも、ある意味でこれまでの経緯に反省を込めた結果です。どういうことかと言うと、知能とは何か、あるいは人工知能というのはそもそも実現可能なのだろうか、それとも不可能なのだろうかという半ば哲学的な議論が、計算機が開発された1960年代直後から、延々と繰り返されているわけです。

知能とはなにか？

計算機というものの自体の数学的基礎を創ったチューリングという数学者がいます。彼は、恐らく計算機を作ったことの延長として、知能とは何だろう、どうしたら定義できるのだろうと考えました（図5）。

ご存知の方も多いと思いますが、チューリングテストという、ちょっとおもしろい、知能があるかどうかの判断基準となるアイデアを提唱しています。どういうことかと言うと、壁で仕切られた部屋があり、テストする人がいて、テストする人はコンピュータ画面を通じて、壁の向こうにいる誰かと文字だけで会話をします。いわゆるチャットに近い会話です。同時性が高い方がよいので文字だけのチャットとしていますが、壁の向こうには人がいるか、チャット機能を持ったコンピュータがあるかのどちらかです。壁のこちら側にいてテストする人は、今自分がチャットしているのが人なのかコンピュータなのかを当てる。もし、チャット相手がコンピュータであるのにそれを見抜けなかった場合、そのコンピュータに知能があるといってもよいだろう、というのがチューリングテストの考え方です。

とても変な設定だと思われるかもしれません。おしゃべりするというのが本当に知能と関係あるのだろうか。おしゃべりするには知能が必要であるとは、もちろん言えることだと思いますが、おしゃべりできたから知能があるとってよいのかということ、そんなに自明なことではないです。しかし、チューリングがあえてそのような少々変わったテストを提唱したということは、すでにその時点で、知能というものを科学的に定義することはとても困難なことだと彼が認識していたということのあらわれなのではないかと私は思っています。

- ・ チューリングテスト
- ・ 中国語の部屋
- ・ 強い人工知能 vs 弱い人工知能
- ・ 身体性？？
- ・ 全脳アーキテクチャー？？？
- ・ 社会的知能？？

図5 知能とはなにか？

もう一つ、知能とは何だろうということを考えた人として、ジョン・サールという哲学者がいます。彼は中国語の部屋という例え話、思考実験を論文の中で提唱しています。どういうことかと言うと、英語しかしゃべれない、中国語に関しては読みも書きもできない人がある小部屋に閉じ込めます。その小部屋には新聞受けみたいな窓が開いていて、時々そこからよくわからない記号が書かれた紙切れが入れられてくる。実際は中国語が書かれているわけです。その小部屋の中の人は何をするかというと、分厚い英語で書かれたルールブックを持っていて、そのルールブックには何が書かれているかということ、こういう記号が並んでいる紙切れが入れられてきたら、外へ返す紙切れにはこういう記号を書きつけなさい。また、別の記号がこのような並びの紙切れであれば、こういうことを書きつけて外へ出さなさい。小部屋の中の人には中国語の文章は全く読めないから、単なる図形だと思って、その英語の指示に従って図形を紙に書

きつけて返します。

彼の質問はこうです。このルールブックはとても巧妙にできていて、箱の外側にいる人にとっては、あたかも紙を介して中国語で会話しているように思えたでしょう。そのとき、この小部屋の中の人には中国語を理解したと言えるであろうか。あるいは、この小部屋+人+ルールブックは中国語をわかっているとってよいのか、というのが彼の疑問です。これは直感的にはとても分かり易い例示です。どんなにルールブックが完璧だったとしても、中国語をわかっているとはいいたくない、何かそういう気持ちが残ることがお分かりになると思います。

それ以降もこのような議論が続いて、知能とは何かを定義する、あるいは人工知能は究極的に可能なのかと深く考える伝統が、人工知能の研究の一角として延々と続いています。

さらに最近のはやりとしては、人工知能にはあれが足りない、これが足りないという話がどんどん出てきています。例えば、計算機に何か知的なことをやらせようと思っても「身体」つまり体がないので、人間ならば自然に分かるような、色々なことが分からない。例えば、重力というものを感じたことがない。床というものが何かわからない。他には、おもしろいところでは、右や左を言葉だけで説明するのはとても難しい。体があれば、こちら側、お箸を持つ手の方が右ですと言えるような話が、体がないと説明できない。計算機に右と左をわからせるのはとても難儀なことなのです。こういう事はいっぱいあります。これらは計算機の中だけで人工知能を作るのは、究極的には不可能なのではという議論です。

このような議論がある一方で、人の知能はどこに現れるかという、例えば、人と人が対話する、協調する、対立を避ける、あるいは競争して勝つというところに初めて現れるものだという社会的知能という考え方があります。つまり、社会と知能は裏返しの関係になっていて、計算機の中に AI ソフトウェアで単独で知能を実現しようと頑張っても無駄で、社会そのものを再現するまでは知能というのは現れないという議論もあります。

あるいは最近、全脳アーキテクチャという考え方を提唱している人がいます。これは非常に極端な議論だと思いますが、どういうことかと言うと、脳が知能を実現していて、脳を中心として人間の知的機能が働いている。故に、脳の動きをコンピュータでとても巧妙にシミュレートすれば、そのときは人間並みに機能する人工知能システムが実現するであろうという訳です。このように、脳のシミュレーションを通じて人工知能を実現するというアプローチを考える人もいます。

このような話がたくさん出てきたわけですが、そのような哲学的な議論あるいは科学の基礎的部分に関するような議論と、人工知能とはいささか別の話ではないかというのが我々の考え方です（図6）。

例えば、AI ソフトウェアの開発を通して人間の知能に対する洞察が深まるという人もいます。完全に嘘だとは思いますが、例えば、飛行機が発明できたから鳥のことが前よりよくわかるようになったのかと考えてみても、ほとんど関係のない話です。飛行機を飛ばすためのテクノロジーと、鳥が鳥であるように飛ぶことの間には、ほとんど何の関係もない。それと全く同じことで、AI ソフトウェアが実現できたから人間の知能について何かがわかったというのは、いささかおこがましい、それこそ違う話ではないかと思います。

- ・ AI を通じて人の知能に対する理解を深める
... 本当？
- ・ 飛行機の発明で鳥に関する理解が深まった？
- ・ 「知能が機械的に実現できる」はイデオロギー
- ・ 「10年以内に何が実現できるか」が問題
- ・ 「実現できる(はず)」≠「実現できる・できた」

図6 人工知能

もう少し細かく言うなら、知能を機械的に実現できるか、究極的に計算機にソフトウェアを作り込んで、人と同レベルあるいは人のように考える機械が作れるのだろうかと考えてみると、これはとても複雑な問いかけになっています。知能を計算機の中に計算システムとして実現できるというのは、当然のことながら、人工知能研究を底辺で支えているイデオロギーなわけです。これは結局のところ、世界は機械的に、ある物理法則なり何なりの仕組みに従って、究極的には規則的に動いていて、その中には当然人間も含まれているという機械的な世界観・人間観です。

先ほどの全脳アーキテクチャの議論は、まさしくこれだと思います。人間も究極的には物理法則に従って機械的に動いていて、その機械的に動いている人間が持っている機能としての知能なのだから、究極的には機械的に実現できるはずである。これがある種の人工知能研究者を支えるイデオロギーなのだと思います。しかし、実は、それは全くもって科学的に証明された話ではなく、何の保証もない主張に過ぎない。つまり、事実ではなく、単なるイデオロギーだというわけです。

さらに複雑な言い方をすると、イデオロギーとは別に、ある種の万能性を持った数学理論という見方があります。どういうことかという、ニューラルネットという計算の仕組みは、どのような関数でも所望の精度で近似できることが数学的に知られています。

ここまでの二つ、究極的には知能は機械で実現できるということと、ニューラルネットでどのような関数でも近似できるということを認めたとします。さらにもう一つ、人間の知能も究極的には関数であらわされる。それも認めたとします。ここまで認めたとしても、いつになったらニューラルネットで知能が実現できるかというのは、全く別の問いかけです。それは5年後なのか、10年後なのか、100年後なのか、あるいは10万年後なのか。モノとして何ができるか、機能として何ができるかというのは、単なる数学的な性質の話であって、実際にそのものが思いどおりに動くように訓練する、プログラムを作り込む、あるいは学習データを集めることは別物なのです。原理的にできるかどうかという大前提とは別に、実現するために必要なものが他にも、ものすごく沢山あって、それはもしかして、人間に達成できる限界を優に超えているかもしれない。

これでも分かり難いということなら、別の例を挙げます。

全ての世界の人が自分の隣人であるかのように思って、戦争もなく飢えもなく幸福に生きる世界、これは想像できないわけではないですが、その状態に至るまでの経路がさっぱりわからない。それと同じように、計算機はある意味で万能であって、我々が計算だと思うものはほぼ全て原理的に計算可能ですが、

知能が例え計算であったとしても、その計算をどのようにしたら実現できるのか、実現できる状態までにコンピュータに何を教えてやればよいのか、いささかも明白なことはありません。

つまり、原理的に何が可能かということは意外とどうでもよいことであって、5年以内、10年以内、20年以内に実際に何ができるかを目で見えて確かめて、あるいは高い精度で予測することの方が、はるかに重要な問いかけであるというのが、東ロボプロジェクト^{*1}の一つのドライビングフォースになっています。実現できるはずということと、実現できたかどうかということは全く別物ということです。

^{*1}「ロボットは東大に入れるか」プロジェクトの略称

人工知能は役立つか？

哲学的な議論はもはや良しとして、重要なことは具体的な問題を解くことです。しかもポイントは何かというと、客観的にできた／できなかったが明確にわかることです。100点中の何点取れたのか、平均的な人はこの辺りまではできるが計算機はここができない、あるいは平均的な人よりもこの部分は計算機の方ができるということが、データを基にはっきりわかる問題設定が見つけれられたことが、プロジェクトにとって非常に重要なポイントであると思います（図7）。

- ・ 明確に定義された課題を解こう
(解けた・解けない・100点中何点とれたか)
→ いま何が出来て、何ができないのか
- ・ 易しすぎない
- ・ 難しすぎない
- ・ 役に立つ

図7 具体的な問題を解こう

ここで、もう一步広げて社会とのかかわりについて話をします。このプロジェクトの開始後のことですが、センター試験の廃止がほぼ決まってしまいました。2016年のセンター試験で高得点を挙げるという目標まではよいのですが、その後センター試験がなくなります。東ロボプロジェクトができたので、これ以上センター試験を続けても仕方ない、と政府が判断したというのは無論冗談なのですが、完全に冗談とは言いきれない側面もあります。このプロジェクトを始める時点で、センター試験に対して機械がある程度まで得点を挙げられるという見込みがあって始めたわけですが、実際に毎年性能測定してみて、この辺りの得点までは取れたというデータが得られています（後で実績を示します）。

そこに別の動きが出てきた。つまり、知識のごく簡単な応用だけに特化したような人材は今後必ずしも必要ではなく、別種の実能力を持った人が重要になってくるという考え方が出てきた。図8は教育再生会議からのドキュメントの抜粋ですが、この議論の背後には恐らくこのような事情があったのだろうと想像しています。これだけができる人は、もはや計算機に取って替わられるというのが、背景の一つだったというのは十分予

のに必要な計算プロセスと、現実世界のロボットに指令を与えて仕事をさせるための計算プロセスは、大変共通する部分が多いことがおわかりになると思います。

産業ロボット、ブルドーザーの自動運転のようなレベルを超えて、積み重なっている瓦礫を片づけるようなことがロボットと協調してできることは、実現は遙か先のこともかもしれないが、間違いなく必要なことであって、計算機が言葉で書かれた、あるいは言葉と画像で指示された物理の問題を解けることは、その目標に向けた、とても重要な一ステップになっていることをおわかりいただけたと思います。

東ロボプロジェクト

あとは、実際にこのようなことができて、このようなことをしている、というような話を一通りご紹介して終わりにしたいと思います。

まず、私自身は、数学の問題を解くシステムの開発メンバーの一人です。数学の問題はどのようにしているかというと、図 11 に示すような処理をしています。まず、言葉で書かれた数学の問題が入力されます。これは、テキスト形式でシステムに入力されるのですが、いかにも計算機が得意そうな数式で書かれた部分の他に、普通に日本語で書かれた部分があります。言い方を変えるなら、数式と数式がどのように関係し合っているのかは、日本語を通じて指示されています。故に、まずはこの日本語を正確に理解することが必要となります。こうして理解できると初めて全てが形式的に書き下されて、計算機が問題を完全に把握できる記述の段階となります。これはある種の機械翻訳と言ってもよいと思いますが、非常に高度な正確性が要求されるという点で、日本語から英語への翻訳のように、多少の違いがあっても、人が見てほしい意味が分かれば役に立つ、という人間相手の機械翻訳とは全く異なる問題になっています。この後、数式で書かれた記号部分の数学的定義を参照することによって、最終的には数式だけで問題の内容が表示されます。こうして完全に数の関係として問題を表わして、最後に数式処理が行なわれて、解答が得られます。このようなプロセスで数学は解かれています。

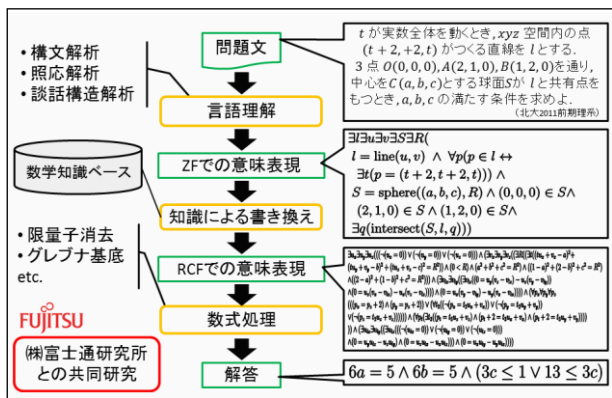


図 11 数学問題の処理手順

図 12 は、現在このシステムが数学の問題にどこまで対応できているかを示したデータです。今現在、完全に自動的には動かしていない部分があります。日本語の解析のところですが、もちろん、数式処理部分は完全に自動で動作しているのですが、

日本語の解析に関しては、少し人間が手助けしてあげた状態での成績です。左側が 2013 年度の東大文系の人を対象とした東大形式の模試での結果、右側が理系での結果です。両方ともシステムの実力は、東大を志望している人の平均点を既に超えています。ただし、文系、理系ともに、まだかなりの割合で上位者が残っているので、彼らに匹敵するほどに性能を高めるには、日本語解析はもちろんのこと、数式の処理の部分においても、まだかなりの改善が必要という状況ではあるものの、とにかく平均点は超えたというのが数学での現状です。

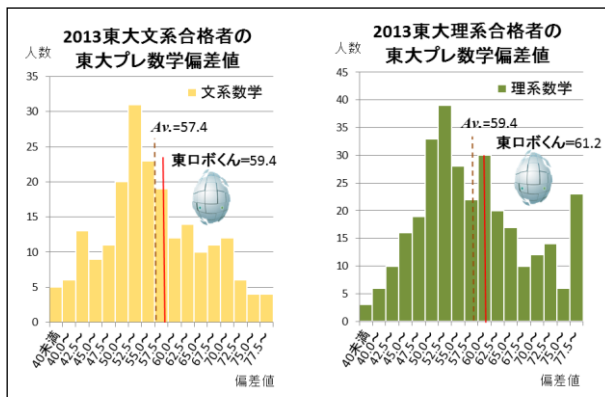


図 12 東大形式模試の成績 (数学)

国語は、図 2 で示したように、問題としてはかなり特殊と言ってもよいと思います。正攻法ですぐに解けるような問題ではありません。ある意味暫定的な手法ですが、なるべく点を取るにはどうしたらよいのかという技術的アプローチで作ったのが、現在私の所属している名古屋大学の佐藤研究室のメンバーによる国語のシステムです (図 13)。

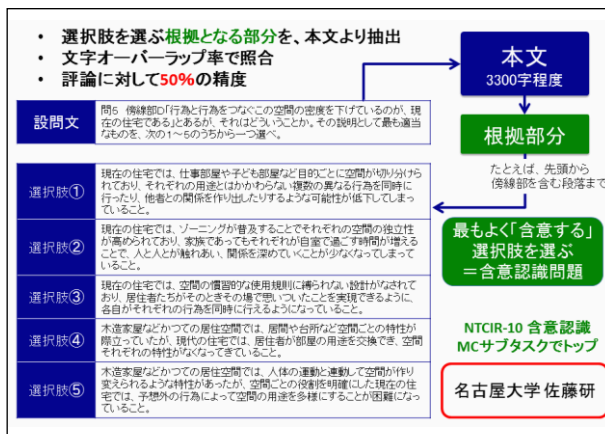


図 13 国語問題の解き方

どのようにしているのかというと、まず本文中から、その問題に対してこの辺りが重要と思われる部分のある程度の量で抜き出していきます。例えば、傍線の周りの一つの段落などを取ってきます。それは、ほとんどの問題が「この傍線のところは何を言っていますか」という形式の問題だからです。その後、それぞれの解答選択肢と抜き出してきた部分とを比べて、同じことを述べているのか、それとも異なることを述べているのか、最終的には判別したいのですが、当然ながら、「言っていること」が同じかどうかを理解できるような人工知能技術は、正直言って今の時点ではありません。

それではどうするのかというと、本文中から抜き出してきたところに「あ」が何回入っていて、「い」が何回、「う」が何回、「山」という字が何回というように、まずはどの文字が何回出てきたかという分析をします。これは分析というほどのことではなく、本当に文字を数えるだけです。そして、本文中から抽出してきた部分と選択肢のそれぞれについて、何の文字が何回出てきたかを数えて、その文字の割合がどれだけ一致しているかで選択肢をランキングします。

とても乱暴なやり方で、文字の分布が近というだけのことで、このような方法で選択肢を選んでいくと、実は、センター試験の国語の評論の問題については50%程度合っていることがわかりました。これは正直なところ不思議です。ちなみに、選択肢は基本5択なので、ランダム選択では20%しか正解になりません。故に、50%取れたということは、明らかに偶然ではありません。それでは、何が起きているのかというと、現在のところ、まだ研究中です。これは恐らく問題の作り方とか、選択肢の設計みたいなことに係わっているのだらうと予想していますが、具体的にはまだ何が起きているかも含めて研究するという段階です。

世界史と日本史の問題に関する基本的な考え方は、選択肢が与えられて、この与えられた選択肢が教科書に書いてあることと矛盾してないかをチェックします。基本はそれだけのことで、もっと端的に言えば、与えられたことと同じことが教科書のどこかに書いてあるかをチェックします。図14の4択の問題では3番目が正解なのですが、これはさすがに先ほどの国語の問題のように、何の文字が何回みたいレベルでは当然合わないです。なぜなら、どの文字もテキストには必ず出てくるし、それは単語であってもキーワードであっても同じで、ほぼ全て教科書のどこかには出てきてしまうからです。

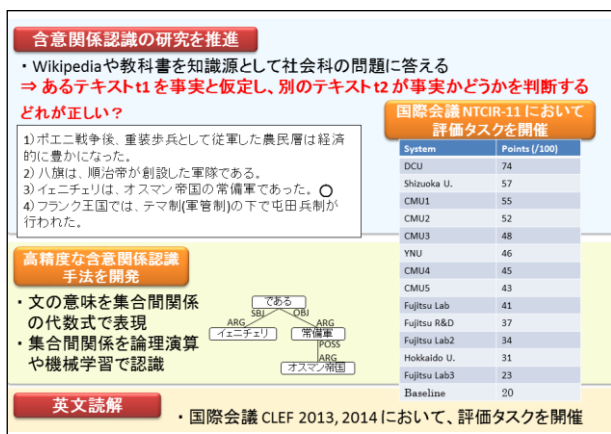


図14 世界史・日本史問題の解き方

ただし、キーワードとキーワードの意味的な繋がりは、選択肢と教科書では中身が違うわけです。間違った選択肢の方は、キーワード自体は歴史上のキーワードなのですが、それが間違った結びつきをしている。例えば、毛沢東がアメリカを建国した。これは極端な例ですが、何れにせよ文中の構造を考える必要があります。キーワード同士の論理的な繋がりを解析し、さらにこれを教科書と突き合わせる。国立情報学研究所の研究グループ内では、この同じ構造を持っているかどうかの突き合わせに基づく解法を研究しています。

また、このような技術開発と並行して、センター試験問題を解くシステムを競い合う催しが、色々な大学の研究チームにより、国際会議の一部として行なわれています。それが国際会議 NTCIR-11 (NII Testbeds and Community for Information access Research) なのですが、10 チームほどが参加して、同じセンター試験のある年度の問題について、システムを競い合わせた結果が図14右側の表です。トップのチームは段違いのスコアの74点で、人間の平均点と比べてはるかに上回っています。これが現状の世界史・日本史における進捗状況ということです。

英語は、他の科目と比較すると、問題のバラエティーがかなり多いというのが特徴です。センター試験でも、例えば、アクセントや発音などの知識さえあれば解けるような問題から、会話を埋めなさいのような問題、穴の空いているところに単語を入れ込みなさいのような問題で、このような比較的短い文の問題が約半分、残りはいわゆる長文読解の問題ということです。

大まかに言って、この二つのパートに分けられて、前半部分はかなりできてきています。図15に示すように、70%程度は安定して正答が得られるようになってきているというのが、英語の開発を担当している NTT コミュニケーション・サイエンス研究所の研究チームの結果です。これに対して後半の長文読解は、国語よりさらに成績が悪くて、チャンスレベルの正答しか得られないというのが、昨年秋の段階での状況でした。

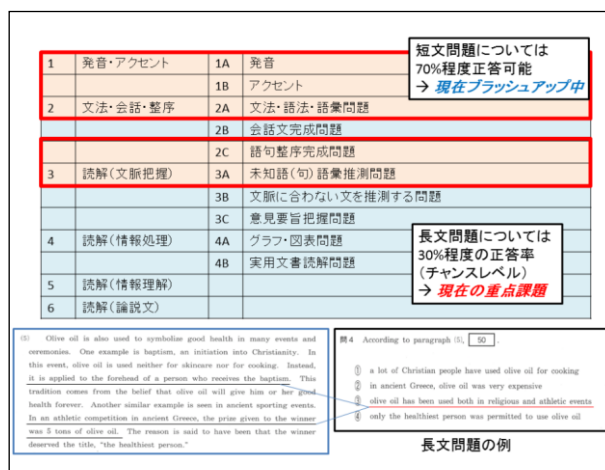


図15 英語問題の正答率

故に、今後の大きな研究テーマとしては、持っている技術を全て長文読解に注ぎ込んで、例えば、単語同士の繋がりとかが、本当のことが書いてあるのか、予測が書いてあるのか、それとも否定されているのかといった事実性を判断する技術とか、それこそ開発途中のものも含めて色々基礎技術を集めてみたときに、長文読解の問題がどこまで解けるようになるのかが、現在の技術的なチャレンジとなっています(図16)。

説明するまでもないと思いますが、もしこれができたとすると、それこそ色々な場面で役に立ちます。しゃべってコンシエールの対話部分の改善、翻訳の改善、レポートを自動作成する、本文に書いてあることを短くまとめるなど、色々な場面で使えるようになることをモチベーションにして、研究チームのメンバーは研究を推進しています。

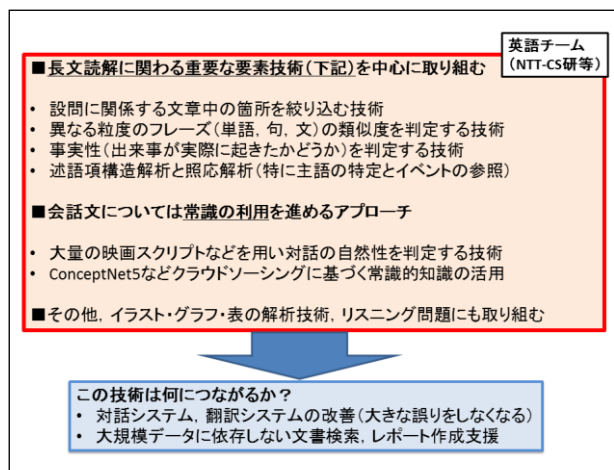


図 16 英語長文読解の取組み

図 17 は、科目全体を通しての成績データです。これは、代々木ゼミナールが主催しているセンター試験の模擬試験のデータです。2014 年の成績結果であり、括弧内は 2013 年の成績です。2013 年から 2014 年までの大きな進歩は、英語の成績がかなり上がりました。これは、NTT グループの参加によるところが大きいです。これに伴って、偏差値も 41 から 50.5 まで上がりました。全体を通覧してみると、私の担当した数学がいささか良くないのですが、2014 年には概ねどの科目もほぼ平均点に達していて、偏差値も 50 近くになっています。昨年秋の段階での総まとめはこのような状況です。

受験番号	116-5004E	氏名	東ロボくん イチゴウ		
科目	満点	全国平均点	本人得点(昨年)	本人偏差値(昨年)	
英語	200	93.1	95 (52)	50.5 (41.0)	
国語(現文+古文)	150	60.2	69 (62)	54.2 (45.9)	
国語(現代文)	100	45.9	49 (42)	51.9 (44.7)	
数学Ⅰ・数学A	100	47.1	40 (57)	46.9 (51.9)	
数学Ⅱ・数学B	100	50.4	55 (41)	51.9 (47.2)	
世界史B	100	40.8	52 (58)	56.1 (55.2)	
日本史B	100	47.2	44 (56)	48.2 (56.1)	
物理	100	32.7	31 (39)	49.0 (48.3)	
文系7科目コース	900	422.0	386 (365)	47.3 (45.1)	
文系3教科偏差値				53.6 (47.7)	
理系3教科偏差値				49.7 (46.2)	

文系7科目コースは英語、国語、数学2科目、地歴公民2科目、理科1科目の合計点。

図 17 東ロボくんの模擬試験成績

グラフ化してみると、図 18 の左側が 2013 年の模試の結果、右側が 2014 年の結果です。

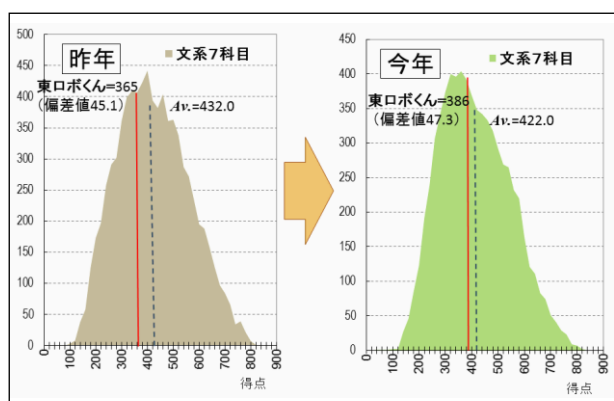


図 18 東ロボくんの成績位置

昨年のシステムの偏差値は、文系の学生の点数分布平均点の少し下あたりに位置していましたが、今年は平均点にかなり近づきました。もう一つは、今年は分布のピークを越えたということです。昨年は最頻値より下だったのが、今年は最頻値を超えました。最も典型的な人の得点を超えたということです。

予備校の模試なので、合格の可能性を判定できます。図 19 は、これまでの学生の成績データを基にして、どこかの大学に受かりそうか確率を出してみた結果です。国立大学はやはり難しいです。2013 年には 1 大学しか 80% 以上の可能性が出せませんでした。しかし、私立大学については、70% 程度の大学で 80% 以上という判定が出ました。2014 年の私立大学はさらに可能性が高くなって、581 大学、1,697 学部に対して、472 大学、1,092 学部で合格確率 80% 以上の判定が出ました、というのが昨年秋までの状況です。

さて、それではこの後、人と人工知能はどのように知的な仕事を分担するのか？ もちろん、現在システムが解けない残り 50 点の部分に人がすべき仕事が残っているのは確かなのですが、機械と人がきれいにすみ分けできるのだろうかというのが、次なる疑問として出てきます。

2013 代々木ゼミナール「全国センター模試」成績による合格判定結果				
大学数	全大学・学部数	合格可能性 80% 以上の大学		
国公立大学	165 大学 566 学部	1 大学	2 学部	
私立大学	579 大学 1,670 学部	403 大学	814 学部	
合計	744 大学 2,236 学部	404 大学	816 学部	

↓

2014				
大学数	全大学・学部数	合格可能性 80% 以上の大学		
国公立大学	166 大学 570 学部	4 大学	6 学部	
私立大学	581 大学 1,697 学部	472 大学	1,092 学部	
合計	747 大学 2,267 学部	476 大学	1,098 学部	

国公立はセンター得点による判定。私立は各大学・入試方式で必要な科目の偏差値による判定。

図 19 東ロボくんが入れる大学

これに関連するのが、このプロジェクトのサイドプロジェクトとして、最近手がけている話です。図 3 のレストランの対話と全く同じ形式の問題で、英語の対話の穴埋め問題です。図 20 は、機械翻訳技術を用いて翻訳してあります。②と④は正解でないことはおわかりになると思います。それでは、①なのか③なのかと問われると、③の翻訳の意味がわからなさ過ぎて、正直なところ迷うと思います。

次の会話の BLANK に入れるのに最も適当なものを、下の①～④のうちから一つ選べ。

A: こんにちは。これはビルです。ジョンはありますか？

B: いいえ、彼はありません。

私は、メッセージを取ることが出来ますか？

A: [BLANK]

B: 大丈夫。お電話頂きありがとうございます。

① ありがとう。ちょうど私が呼ばれるのを彼に教えてください。

② いいえ、ありがとうございます。私はあなたの助けを必要としません。

③ お願いします。ときあなたは彼をバック期待していますか？

④ はい、もちろん。私は気にしません。

図 20 機械翻訳の現状

今これは紙に印刷した状態での設問ですが、音声で読み上げられたときに理解できるかとなると、かなり厳しい状況に置かれると思います。機械翻訳と人が協調して対処するという典型

的な領域を考えた場合、やはり同様な事態となるわけです。英語話者が話した内容が、例えば、携帯電話を介してサーバに送られて、翻訳されて返ってきて、それをいちいちディスプレイに表示させてから、日本語しか分からない相手に読ませるというのは、あまり便利とは言えず、携帯電話を聞いている状態で機械翻訳の出力結果も聞ける必要があると思われます。しかし、現在の翻訳システムの典型的なアウトプットでは、音声だけで理解するのはとても難しいというのが現状です。

図 21 は、機械翻訳システムと人間の翻訳能力を比較したものです。翻訳した問題文で中学生が何点取れるかを調べた結果です。中学生約 700 人に協力してもらいました。左 2 つは機械翻訳 G (Google) と Y (Yahoo)、右 2 つは人間の翻訳です。S とあるのは、人間に機械のつもりになって訳してもらったものです。どうということかという、会話を一文ごとにシャッフルして、文脈が全くわからない状態で一文ずつ翻訳したものです。計算機は文の前後の繋がりは考えようもないので、一文ずつ愚直に直訳しているのが機械翻訳の現状です。S はそれと同等の条件で人間に翻訳してもらったものです。O とあるのは、普通に文全体を見ながら翻訳したものです。

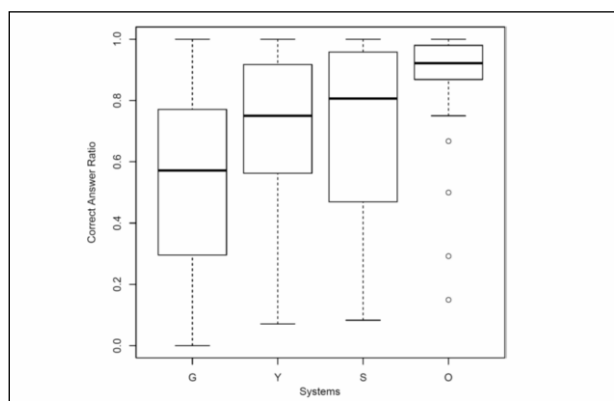


図 21 機械翻訳 vs 人間の翻訳

これを見てわかることは、Google の翻訳は、図 20 でもとても理解し難いところがあったように性能はそこそこですが、Yahoo の翻訳は、人間が機械と同じ条件で行なった翻訳にかなり近いレベルまで達しています。よって、点数をみた限りでは、翻訳システムによっては、文脈まで考えなければこれ以上の上達は難しいことがわかります。人間でも、文脈を考えるかどうかで 10 ポイント以上の差がついているので、機械翻訳がこれ以上人間並みに近づくためには、文章の前後の繋がりを考えるという、計算機にとっては殊の外大きな技術的チャレンジが必要なのが、このデータを基にわかったという実験結果です。

今年度の予定は、センター形式の模試、東大形式の模試に対する評価を秋に行う予定です (図 22)。今年は今までの科目に加えて、世界史の筆記試験でも評価する予定を立てています。この評価キャンペーン自体はオープンなので、自信があるという企業の方がおられましたら、ぜひ研究チームとしてご参加いただきたいと思います。

- 11月14日 成果発表会@東大福武ホール
- 45万人が受験する第一回ベネッセ模試
- 筆記試験(数学、世界史)
 - 世界史は主として要旨要約タスク。プロ(予備校講師)が部分点を人間と比較しつつ評価する。
- マーク、筆記ともに参加はオープン
 - 世界史(マーク、二次)はNTCIR QALabの一環
 - 参加希望者はNTCIRからエントリーしてください

図 22 2015 年東ロボプロジェクトの予定

本日は、このプロジェクトで推進していることをご紹介した上で、何故わざわざ計算機を東大に入れようとしているのかの意義についてお話しさせていただきました。もっとも言いたかったことは、5 年、10 年、20 年で何ができるのか。このタイムスパンでできることを見極めることが最も重要であるとお伝えたくてお話ししました。

また、「人のような知能・人を超える知能」とはとても言えないまでも、多少なりとも人を模擬する機械を開発しているわけで、最終的には人と機械との協調、あるいはどのように協力して暮らしていくのかを、ほんとうに実現されている技術を基にして考えていこうというのがこのプロジェクト全体の願いです。

本講演録は、平成 27 年 6 月 26 日に開催された SCAT 主催「第 95 回テレコム技術情報セミナー」のテーマ、「人工知能」の講演内容です。

*掲載の記事・写真・イラストなど、すべてのコンテンツの無断複写・転載・公衆送信等を禁じます。