

| 質問番号 | 質問 | 回答 |
|------|--|---|
| 1 | Hallucinationの問題は、「回答の確からしさ」や自信度(confidence)のような指標を定義して数値的に表示することは技術的に難しいのでしょうか？ | それぞれの単語を予測する際の尤度の合計がその文の確らしさを表すと言えるので、特別なことをしなくても確らしさの計算をしていることとなります。ただし、回答としては正しいとはいえないけれど単語の予測確率がとても高いといった状況がよく発生するため、実際には専用の良い指標を考える必要があると考えられ、実際にそういった指標の研究もおこなわれています。 |
| 2 | このモデルで、なぜプログラミングが生成できるのでしょうか？ | GitHubからもデータを取得していてそこにあるプログラムのコードも文章と同等に扱って学習をしています。 |
| 3 | 貴重なお話をありがとうございました。言語モデルで生成された文章を人海戦術に頼らずに自動で評価する方法が開発される可能性はあるのでしょうか。同じ強化学習であるAlphaGoは、勝つことだけを評価基準とするため、自分の対局をフィードバックすることで、神様みたいな碁を打つようになりました。同様のパラダイムシフトが起こり得るのか気になります。 | ゲームの世界はいわゆる閉じた世界であり、また、勝ち・負けが明確に定義できるため自己学習がうまくできるという背景があります。一方、言語モデルの学習では、言語表現は例外なども多く時間と共に無限に増えていく、或いは、は変わっていくといった現象を有するので、入力が無限通りあり得るといえます。また、良い文章・悪い文章の判断は人間に依存して決まるもので、数値的に明確に定義することは難しいため、自己学習の問題として定式化するのが困難です。ですので、現状は完全に自動で良くなっていく学習の枠組みを作るのは難しいと考えられます。ただ、こういった研究自体は進められていくと思うので将来的に何か画期的な方法が考案される可能性はあるかもしれません。 |
| 4 | ありがとうございます。立て続けにすみません、尤度が高いのに間違っている、というのはどういう原因なのでしょう？ | 言語は離散的な記号が入力となっていることもあり、連続的な入力を基本的に持たないので、出力確率は比較的ピーキーになりやすいです。そのため、予測した単語が文脈的にあまり正しいとはいえない場合でも予測確率的には100%に近いといった現象が言語を対象とした機械学習をする場面ではよく起こります。 |
| 5 | 企業で利用する例が出てきていますが、ChatGPTに自分たちの持つ文章のデータを追加して学習させて自分たちの分野に役立つように修正していると思えば良いのでしょうか。そうなるなどの程度のデータを持っていれば自分たちの分野ごとにチューニングできるのでしょうか。 | ChatGPTは外から追加で学習することはできません。もし自社データなどにチューニングしたい場合は、ChatGPTに近い言語モデルがいくつか公開されているので、そういったものを使って実施することになると思います。 |
| 6 | パラメータを増やすと性能が上がるという話がありましたが、過学習は大丈夫なのでしょうか？ | 詳細な説明はいろいろありますが、ざっくりな回答としては、問題ないと考えられています。直感的にいうと、必要なパラメータ数は問題の難しさに依存すると考えることができ、人間があつかう言葉を全て正しく理解できるようなモデルを作ることを考えれば必要なパラメータ数は非常に多く必要と言え、現状のパラメータ数が過学習が発生するほど多いわけではない、という可能性すらあるかもしれません。また別の側面として、ニューラルネットワークを使っている関係で、パラメータを増やすことで、局所解にトラップされる可能性を減らすことができ、結果としてより適切な学習ができるということもあります。このあたりは、昨今のニューラルネットワークが流行る以前の凸問題の機械学習での理論と合わないことはあります。これは現在の言語モデルなどは真の意味で準最適解に収束するまで学習しているわけではないので、理論的な話と合わないとも言えます。 |