・・・・5・・・・

人工無能のつくりかた -文章生成の初歩の初歩-

60 回生 Jyakky

さわやかな初夏の風が・・・いや、違ったか。ええと、本日は灘校文化祭へようこそ。わが部は特殊すぎる事情によって五月二日も半分ほど準備に費やしている可能性が非常に濃厚ですが、ご容赦ください (ぉ

さて。編集担当の Faey さんが部長である僕の言うことを聞いてくれればこの文章は真ん中よりちょっと後ろの目立たなさそうなところに溶け込んでいるはずなのですが、いかがでしょうか。

♯ いや、別に読んでほしくないだとか書きたくなかったとか、そういうんじゃないですよ？全然。

本当は、「暗号入門」を書こうと考えていたのですが、締め切りが一週間後まで迫った段階でまだ「考えて」いる段階だったので、ちょうどその頃単純な人工無能の開発を暇つぶし程度に行っていたこともあって、人工無能について書くことにしました。国語の成績が年々下降し、とうとう通知表を夕焼け色に染める程になってしまった僕の文章なので、全体的に読みにくいこと甚だしいかと思われますが、そういう運命だと諦めていただけるとうれしいです。

ところで。この部誌というものが対象とする読者の年齢層がどの程度なのかさっぱりわからないので、どういったことを書けばいいのか混乱しているわけですが、やはりいきなりプログラムのソースコードを載せるわけにもいかないので、簡単な解説だけにとどめておくことにします。「簡単」かどうかはともかくとして。

人工無能を知らないあなたへ

♯ もしあなたが人工無能についてそれなりに知っている (or 知りたくもない) ならこの章は読み飛ばしてもいいです。

> **人工無能**
> 主にチャット等で発言の中のキーワードに反応して適当な対応を返すプログラムのこと。従来の人工知能研究とは異なり、会話における『表象の現象だけ』を考えて会話をシミュレートしようとするアプローチを取る。
> 古くは Eliza から最近では「どこでもいっしょ」のトロに至るまで、様々な人工無能がある。
>
> (はてなダイアリー キーワードより)

ええと、まず「人工知能」はわかりますよね。まぁ細かい定義までは僕も知らないのでそのあたりは Google 先生にでも聞いてもらうとして (おい)、とりあえずイメージとしては「コンピュータのくせに人間みたいに考えたりする生意気な奴」という感じで

しょうか[1]。いや、とりあえず僕はそう思っているのですが、ならば人工「無能」とはなんぞや。「能が無い」などというレッテルなど貼ってしまって外交問題に発展したりはしないのか。しないだろうけど。

・・・脱線しすぎました。とにかく、人工知能は人間のようにいろいろ考え (るように研究されてい) ますが、人工無能はなにも考えちゃいません[2]。というか、初歩的な人工無能は、覚えている言葉をランダムに組み合わせて話すことしかできません。人工知能は言葉を理解して受け答えしようとしますが、人工無能は正直そんなことあきらめています。「それっぽい」受け答えができればいいや、というコンセプトです。そのため、人工知能とは違い、「それっぽさ」を追求する研究が多く行われています。

♯ ずいぶん大雑把な書き方してしまったな・・・まぁいいか。

ごくごく初歩の文章生成

上記のように、人工無能というのは「それっぽさ」が命なので、「それっぽい」文章を自動で生成するアルゴリズムが要となるわけです。なので、以下ではそのあたりのことを少しばかり書いて行くことにします。

♯ まぁたいしたことできないのだけど。

文章を生成する、と聞いて最初に思いつくのは、「とりあえず元の文を用意しておいて、単語を適当に置き換える」というものだと思います。この場合、例えば、

元の文「A って B よね」から、
「ネコ って かわいい よね」
「夏 って 暑い よね」
「宿題 って だるい よね」
「俺 って かっこいい よね」
etc.

という感じの文章が生成できます。

♯ いや、文章の内容は気にしない方針で (ぉ

とにかく、こうすればあとはたくさんの元の文と単語を保存しておけば、基本的に文法的には完璧な文章が自動で生成できます。「しくみも簡単だし文法上も無問題！まるで魔法だ！」とか「もうこれで完成でいいんじゃないの？」とかいう声が聞こえてきそうですが、違います。断じて違います。我々の目指すところはもっとはるかなる高みにあるのです。そうだ、そうに違いない。

♯ クールダウンクールダウン。

・・・何が問題なのかというと、「型にはまった応答しかできない」ということ。「型どおりの返答ができる」というのは裏を返せばつまりはそういうことです。このままではちっとも面白くありません。それどころか退屈です。文章生成計画、早くも挫折か。

あるいは、「文法とかどうでもいいから単語並べちゃえ」というアプローチもあります。これなら、文章が一定の型にとらわれるなどということはまずありません。例えば

---

[1] 実際はちょっと違うんですが。　て。
[2] まぁ「考える」がどういうことか、は置いておい

「来たは数学でもあった返信普通と急いで」
「記号も一般的は違いかな分からないいや良くない」
etc.

という感じです。悲惨です。当然です。文法を無視してしまったら意味が成立するわけないのは火を見るより明らかです。話しかけたらこんな返事が返ってきたのでは、驚きを通り越して悲しみまで覚えます[3]。文章生成計画、早くも挫折か。

マルコフ連鎖で手軽な作文

マルコフ連鎖 (Markov Chain)
ある変数を順次、発生させるときに、現世代のパラメタの値 (のセット) のみをもとに次世代のパラメタの値を発生させる方法

(はてなダイアリー キーワードより)

・・・これだけでは何のことやらさっぱりなので、具体的にどのように文章を生成するのか説明することにします。

♯ そこ、「ページ稼ぎ」とか言わない！

・まず、文章を用意します。

例) 今日も明日も学校があるなんて信じられない。

♯ 繰り返すけれど、文章の内容は気にしない方針 d(略)

・次に、文章を形態素[4]に分割します。

例) 今日　も　明日　も　学校　が　ある　なんて　信じ　られ　ない　。

・形態素を連続する二つごとに並べます。

例)

(先頭)	今日
今日	も
も	明日
明日	も
も	学校
(省略)	(省略)
ない	。
。	(末尾)

・二つのうち前半が同じものをまとめます。

---

[3] まぁこっちのほうが面白い、という人もいるかもしれませんが。

[4] 言語の中で意味を持つ最小単位のこと。

例) 例文が短いのでいまいちわかりにくいのですが、

(先頭)	今日
今日	も
も	明日 学校
明日	も
(省略)	(省略)
ない	。
。	(末尾)

・これを多量の文章で行うと、「ある形態素の次にどのような形態素が続くか」の傾向が見えるようになってきます (「学習」みたいなものです)。

例) 使った文章は割愛しますが、

(先頭)	今日 今日 今日 もし だいたい そもそも いや
今日	も は は
(以下)	(省略)

・文頭からはじめて、次に続く形態素の中からランダムに一つ選んで文章を伸ばしていきます。

例) 今日　は　やはり　学校　が　、　しかし　実際　に　行く　。

これで完成です。簡単ですよね。ただ、これではぜんぜん意味のある文章にならないので (まぁもともと意味なんか考えてないけれど)、ここからもう少し改良することにします。

手軽な改良計画

1.二重マルコフ連鎖

マルコフ連鎖+「二重」。だいたい意味は分かると思いますが、単純マルコフ連鎖が直前の一形態素から次を決定するのに対し、手前の二形態素から次を決定する方法です。単純マルコフ連鎖はどうにも日本語と相性が悪いようなのですが、このようにすればその問題もかなり解消されます。同じように、三重にすることもできますが、あまりさかのぼり過ぎるとより多くの例文が必要になったり、保存するデータの量が増えてしまうという問題点もあります。

♯ 実装もかなり面倒になるし (ぉ

2.双方向マルコフ連鎖

マルコフ連鎖+「双方向」。これまでは文章の先頭→末尾への解析しかしませんでしたが、末尾→先頭への解析も行うようにします。このようにすれば、文章の先頭に限らず、任意の単語から前と後ろに同時に文章を延ばしていくことができます。

♯ ということは、主語を種にすればきちんとした文章が作れるということかもしれない。

3.品詞で形態素を区別する

そのままです。形態素を文字の並び+品詞で区別することによってより精度の高い文章生成が行えるようになります。

とりあえず、現在すでに実装されているのはこの三つです。他にもやらなければならないことや考えていることはいくらかあるのですが、それは次章に。

今後の指針のようなもの

さて。この章の内容は基本的に未確認情報です。実際に実装したわけではありません。ただ、これを読んだ方で、もしも人工無能を作ってみたいという人がいれば、参考になるかもしれないと思ったので書くことにしました。

♯ あと自分用メモとして。

いや、別に暗殺された天才科学者の手記とかそういうたいそうなものではないのですが。

1.前後のつながりをもっと大事に

「双方向」とはいっても、結局文章を伸ばす時には一方向のつながりしか考えていないのですが (A B → C のときの B ← C を考えていないということ)、ここで逆方向のつながりも考えるとより自然な文章が生成できるかもしれません。

2.そもそもマルコフから離れる

今のアルゴリズムは、文章の形を自然にすることにこだわりすぎて、意味をほとんど考えていないので、もしかすると新しい方針が必要かもしれない。例えば、品詞だけでマルコフ連鎖を作って、それを仮想的な「文法」として、そこに単語を当てはめていく。同時に単語と単語のつながりをうまく定義して単語のネットワークのようなものを形成して、とかも考えているのですが (というか、仮想文法を作るところまではすぐできるかもしれない)、単語同士の関連性をどのように表すかがさっぱりわからないので挫折しています (ぉ

おわりに

長かった・・・、いや、実際にはそれほど長くなかったのですが、なにぶん無理やり書いたのでどうも長く感じたわけで。それでも後半はだいぶ勢いに乗れたかな、という気はします。

さて。こんなつまらない悪文を最後まで読んでくれたあなたのことです。きっとあふれんばかりの才能の持ち主なのでしょう。いや、きっとそうです。僕がそうだといえば

そうなのです。

灘校パソコン部は、そんな才能あふれる奇特なあなたをお待ちしています。もしあなたが灘校生なら、学年は問いません。旧校舎四階、地学教室横の薄汚い小部屋まで見学に来てください。それと、あまり乗り気でないあなた。長い人生、そのくらいの寄り道も悪くないものです。ここらで、パソコン部を道連れにするのもあるいは楽しいかもしれませんよ。

♯ 要は見学に来てくれ、ということなんだけど (ぉ

ともかく、もう夜も遅い (AM04:09) ので、ここらでペンを置くことにします。提出期日を五日もオーバーしているのに快く編集を引き受けてくれた Faey さん、ありがとうございました。

♯ まぁ「快く」かどうかは聞いてないから知らないのだけれど。(というか多分違う)

それでは、来年こそは「暗号入門」が書ける事を祈って。