

第2回: チューリングとチューリングマシン

チューリングマシンで有名な、イギリスの数学者で現代計算機科学の父と呼ばれるアラン・チューリング (Alan Mathison Turing) を紹介する。チューリング・マシンは、現在のコンピュータの基本的なモデルであるとともに、言語処理の基本的な構造も提供している。関連して、マルコフ連鎖 (Markov Chain) についても紹介する。チューリング・マシンは、抽象的な計算モデルで、無限の記録テープとそれを読み書きするヘッドからなる。チューリングは、理論的に証明できることは手続的に書き下すことができると考えた (有本 2000: 25)。



Alan Turing 1912–1954

1 チューリング・マシン

チューリングは言語学者ではない。しかし、チューリングの思想は言語学に有益である。直接的にはあまり言われないことであるが、チューリング・マシンとマルコフ・チェインは、隣接する記号の読み取りという点では同じである。この2つの形式化はリニアな言語音声を読める点において酷似している。

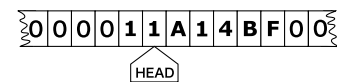


図 1: チューリングマシンの原理図

チューリングマシン

アラン・チューリングが 1943 年に、今日のコンピュータの基礎概念といわれるチューリングマシンという概念を英国で発表しました。図 1 のようなもので、テープの上に書き込み読み出しのできるヘッドを置き、テープの上では一駒ごとに情報があり、機械側に取り込んだり、書き込んだりして、機械の内部状態を変え、その上で、ヘッドをテープ上で一駒だけ動かすことができるという簡単なものです。この装置で、テープに無限の長さがあれば、コンピュータが行うような演算の総ては可能であるというものです。そういわれれば、そんな気がしますが、全くの原理としての先見性であって、実際に作られたわけではありません。コンピュータが実際に作られたのは、それから約 10 年後です。磁気記録媒体としても、ワイヤ型の録音機は 19 世紀からあったそうですが、磁気テープによる録音機は、その 2 年後 1936 年にドイツで発明されましたが、連合国側が知ったのは戦後のことだそうです。チューリングのこのマシンによって、アルゴリズムという概念が確立されたのだそうです。現在のコンピュータも突き詰めれば、このチューリングのマシンの原理に従っているといえるというのが定説です。

「コンピュータと脳」深町一彦より

- 問 1 日常生活で図に似た構造のものを探し、それとチューリングマシンはどんな点で似ているだろうか。
- 問 2 言語をチューリングマシンでモデル化するならば、「ヘッド」「テープ」「セル」はどんな言語の要素と対応するかを話し合え。
- 問 3 また、「ヘッド」の「移動」はどんな機能あるいは行為なのかを話し合え。
- 問 4 言語を抽象化するとはどういうことかを考えなさい。
- 問 5 チューリングマシンの形式的定義 (数理モデル) を調べ、それがどのように表現されているのかを考えよ。

2 マルコフ連鎖

アンドレイ・マルコフ (1856–1922) はロシアの数学者で、確率論において多くの業績を残す (セントアンドリュース大学: 数学者の歴史参照)。



Andrei Markov 1856–1922

次の文を読んで、マルコフ連鎖とチューリングマシンとの関係について考えよ。

マルコフ過程とマルコフモデル

ある文字（あるいは単語）の出現確率が、直前の m 個の記号によって決定されるような確率過程を、「マルコフ過程 (Markov Process)」という。その中でも $m = 1$ の場合、いわゆる直前の場合を特に、単純マルコフ過程と呼ぶ。このマルコフ過程にしたがい、記号が出現すると仮定する確率モデルのことを、「マルコフモデル (Markov Model)」という。 $m = 1$ の、ある時 t の記号の出現確率は、直前の記号のみに依存する。時刻 t における記号を x_t と書くと、マルコフ過程は、 $p(x_t | x_{t-1}, x_{t-2}, x_{t-3}, \dots, x_{t-m})$ となる。直前の記号のみに依存する単純マルコフ過程は、 $p(x_t | x_{t-1})$ となる。マルコフモデルは、Manning and Schütze (1999) の第9章がすべてそれに当てられている。

問6 マルコフモデルとチューリングマシンがどんな点で似ているか、話し合いなさい。また、空也像（六波羅蜜寺）や Annunciation of cortona との共通点について考えよ（図2、3）。



図2: 口から玉のような単位となって語が順番に出てくることは間違いない。ただし、頭で考えたことが、リアルタイムにどのように語が選択され、口から出るのかわからない。空也上人像, 康勝作, 13世紀, 木造彩色, 117.6cm, 六波羅蜜寺, 京都, 日本, パブリック・ドメイン

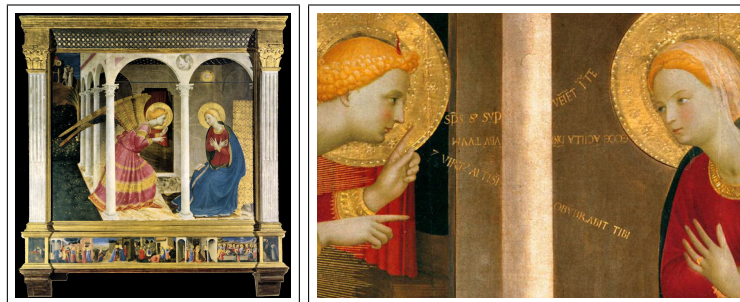


図3: Fra Angelico: Annunciation of cortona (1433-4)

問7 言語（単語、文、段落など）をマルコフモデルで計算するためにはどのような情報を用いて計算すれば良いかを話し合いなさい。

問8 n -gram モデル（シャノンの情報理論）も隣接の確率を計算するモデルである。 n -gram モデルを調べ、どのようなモデルで、どのようなところで利用されているか、調べてみなさい。

問9 隣接する単語はすべて均等な確率で隣接しているのではないことがわかった。では、高い確率で関係している任意の2語（もしくは2つの句）をルール化することはできるだろうか。

3 Markov Chain による文章の生成

Mark V. Shaney とは次のようなものである。

Mark V. Shaney is a fake Usenet user whose postings were generated by using Markov chain techniques. The name is a play on the words “Markov chain.” Many readers were fooled into thinking that the quirky, sometimes uncannily topical posts were written by a real person.

上記の仕様を再現してみる。単語分割には **MeCab** を使い、それを **shaney** に渡す。

リスト 1: 「坊っちゃん」をマルコフチェーンで表現する。

```
1 % cat bochan.txt | mecab -O wakati | sed -e 's/。 /。 \n/g' | shaney | sed -z 's/\n//g;\n/s/。 /。 \n/g;\n/s/」 /。 \n/g;\n/s/「//g' | tail -10 |nl |lv
```

1. 赤シャツは知らん顔をして都合さえつけば、しらを切る積りで図太く構えている。
2. だからあいつを一番へこます為にならない。
3. 抜き身を自分で床の間へかけて、取って、五十名の寄宿生が、乗ぜられる事があるから、余計な口を歪めて、善悪の考もなかった。
4. 何で、御負けにその机が部屋の中は……煽動している。
5. 野芹川の堤へ出懸た。
6. 清の墓は小日向の養源寺の座敷に寐てしまった。
7. それじゃ御出しになる気か知らないが、御嬢さんも仕方がないから大丈夫だ。
8. おれはあるまいかと聞いてみると、非常な勢で尋ねた位だと云うから、そうだ。
9. 学校の方で漢学の先生を愚弄する様な声を出す山嵐を雇って、蒟蒻版の様に後ろからは受け留めにくい。
10. おれは笑われる様なものだから致し方がないと云うのが自慢になるなら、開いてもらわない方がいいだろうと、左に大きな懸物を読んでいる

問 10 shaney はどのような出力をするのかを考えなさい。

問 11 上の出力を見て、これらの文と普通の日本語との違いについて議論せよ。

問 12 言語のモデルはどうすれば記述できるかを議論せよ。

参考文献

- 有本卓 (2000) 『ロボットにかける夢: 知と身体の調和をめざして』, 岩波科学ライブラリー, 岩波書店, 東京, Japan. アラン・チューリングに関わるページ p.52-63.
- Manning, Christopher D. and Hinrich Schütze (1999) *Foundation of statistical natural language processing*, Cambridge, Massachusetts: The MIT press.