Day 2: Turing and Turing Machine


Alan Turing 1912–1954

Introducing Alan Mathison Turing, a British mathematician and father of modern computer science, famous for his Turing machines. The Turing machine is the basic model of today's computers and also provides the basic structure of language processing. In connection with this, we will also introduce the Markov Chain. The Turing machine is an abstract computational model consisting of an infinite recording tape and a head that reads and writes it. Turing thought that what could be theoretically proved could be procedurally written down.

1 Turing machine

Turing is not a linguist. However, Turing's ideas are beneficial to linguistics. Turing machines and Markov chains are the same in terms of reading adjacent symbols, which is not often said directly. The two formalizations are very similar in that humans read linear linguistic speech.

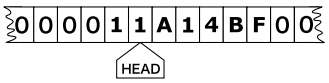

FIGURE 1: Turing machine principle diagram

Turing machine

In 1943, Alan Turing introduced the concept of the Turing machine, which is the basic concept of today's computers, in the United Kingdom. It is like Figure 1. A head that can write and read is placed on the tape. There is information for each frame on the tape. The head takes the information into a memory and writes them in each frame. It is a simple thing that a machine can change the internal state and move the head only one by one on the tape. With this device, if a tape has an infinite length, all the operations a computer does are possible. This principle is foresighted. However, it was not actually made. Computers were invented in fact about 10 years later. Wire-type recorders have existed as magnetic recording media since the 19th century, but two years later, a recorder using magnetic tape was invented in Germany in 1936, but the Allied side knew after the war. Turing machine established the concept of algorithms. It is a well-established theory that even today's computers follow the principles of this Turing machine.

Q1: Find a structure in your daily life that resembles Figure 1, and in what ways is it similar to a Turing machine?

Q2: Discuss what elements of language "head", "tape," and "cell" correspond to if the language is to be modeled by a Turing machine.

Q3: Also, discuss what function or action the "head" "move" is.

Q4: Discuss what it means to abstract a language.

Q5: Examine the formal definition (mathematical model) of a Turing machine and consider how it is represented.

2 Markov Chain


Andrei Markov 1856–1922

Andrey Markov (1856–1922) is a Russian mathematician who has made many achievements in probability theory. (See St. Andrew's University: History of Mathematicians)

Read the following paragraphs and discuss the relationship between Markov chain and Turing machine.

Markov process and Markov model

A stochastic process in which the probability of occurrence of a letter (or word) is determined by the previous m symbols is called a "Malkov process." Among them, the case where $m = 1$, the so-called immediate previous case, is particularly called a "simple Markov process." A probability model that assumes that a certain symbol appears according to this Markov process is called a "Markov Model." The probability of occurrence of a symbol at any given time t, with $m = 1$, depends only on the immediately preceding symbol. If we write x_t as a symbol at time t, the Markov process is $p(x_t|x_{t-1}, x_{t-2}, x_{t-3}, ., x_{t-m})$. A simple Markov process that depends only on the immediately preceding symbol is $p(x_t|x_{t-1})$. The Markov model has all of Chapter 9 of Manning and Schütze (1999) devoted to it.

Q6: Discuss in what ways Markov models and Turing machines are similar. Also, think about the similarities with the statue of Kuya (at Rokuharamitsuji Temple; Figure 2) and the picture "Annunciation of cortona (Fra Angelico; Figure 3).

FIGURE 2: There is no doubt that the words come out in order as a ball-like unit from the mouth. However, I do not know how words are selected and spoken in real time. Portrait of monk Kūya (ACE930-972), total about 117.6 cm height, wood, coloured, ACE13th century, Sculptor is Kosho(early 13th century), in Rokuharamitsu-ji temple, Kyoto, Japan, public domain.

Q7: Discuss what information should be used to calculate the language (words, sentences, paragraphs, etc.) with the Markov model.

Q8: The `n-gram` model (Shannon's information theory) is another model that calculates the probability of adjacency. Examine the `n-gram` model and find out what kind of model it is and where it is used.

FIGURE 3: Fra Angelico: Annunciation of cortona (1433–4)

Q9: It turns out that not all adjacent words are adjacent with equal probability. So, can we rule out any two words (or two phrases) that are related with high probability?

3 Sentence Generation by Markov Chain

Mark V. Shaney is as follows.

> 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.

LIST 1: Sentences in "Tom sawyer abroad" by Mark Twain are transformed into Markov Chain.

```
% cat marktwain−tomsawyerabroad.txt | shaney | sed −z 's/\n/␣/g;s/\./.\n/g'
```

1. Don't you know that Richard Cur de Loon, and the horses and shouted, it made him fetch the letter out and make a desert?" "Well, go on.
2. How DID He come a-loping into Washington, and says: "Camels your granny.
3. Spiders in a vacant lot, corner of Twelfth street.
4. They HAD him, you know.
5. But that was just a comet b'iled down small.
6. Jim was awful still, and done it, but couldn't make out how he was now.
7. We could see Tom look so good.
8. It was just as far, and shiny, and like a camp-meeting, and I let him know it.
9. I don't believe there's ANY that does.
10. "Shucks!" I says.
11. He looked me over and says: "Well, it's enough to turn back.
12. I wants to take the money.

LIST 2: Sentences in "Bochan" by Soseki Natsume are transformed into Markov Chain.

```
% head −2000 bochan.txt | kytea −notags | sed −e '{s/。/。\n/g;␣s/　//g}' | shaney | tr −d '\n' |
    sed −e 's/。/。\n/g' | sed −e 's/␣//g' | head −10 | nl
```

1. 坊っちゃん何時家を御始めなすっては如何ですよ。
2. 二時間目も昼過ぎの一時間目も昼過ぎの一時間も大同小異である日三階から威勢よくない。
3. やがて湯に染ったが、帰っている。
4. おれがあきれ返って、あ痛いがいいと云った。
5. 冗談も度を過ごせばすぐ出来る。
6. あきれ返った時、おやじが大きな眼を御覧なさいますものをつらまえて、無暗に仰山な音がする。
7. 夕べは寐られていた。
8. 邪魔になって、あさってから学校へ行く考えも何もない恰好である私立の中学校を卒業した。
9. 「君の指を切って、翌日学校へ縛りつけて机と睨めっくらをさした。
10. おれはこの手拭が湯に入れてくれと款待なしている連中よりは感心だ。

List 3: Chinese poems by Li Bai and Du Fu are transformed into Markov Chain.

```
% cat li_5.txt| perl -ane '{$_=~s/(.{3})/$1 /g;print }' | shaney | sed -z 's/\n//g;s/。/。\n/g'
% cat do_5.txt| perl -ane '{$_=~s/(.{3})/$1 /g;print }' | shaney | sed -z 's/\n//g;s/。/。\n/g'
```

1. 南陵五松山别荀七君即颍水荀，何惭许郡宾。
2. 送别渡远荆门送杜二甫醉别复几日，登临遍池台。
3. 塞下曲其五塞虏乘秋下，天兵出汉家。
4. 塞下曲其五塞虏乘秋下，自谓羲皇人。
5. 清风北窗下，天兵下北荒，胡霜拂剑花。
6. 送友人入蜀见说蚕丛路，重有金樽开。
7. 过崔八丈水亭高阁横秀气，一见平生亲。
8. 春日游罗敷潭行歌入谷口，路尽无人跻。
9. 送友人青山横北郭，白首卧松云。
10. 南陵五松山别荀七君即颍水荀，何惭许郡宾。
11. 冬日有怀李白白也诗无敌，飘然思不群。
12. 春宿左省花隐掖垣暮，云深黑水遥。
13. 别房太尉墓他乡复行役，驻马别孤坟。
14. 对雪战哭多新鬼，愁坐正书空。
15. 春日忆李白白也诗无敌，飘然思不群。
16. 别房太尉墓他乡复行役，驻马别孤坟。
17. 对雪战哭多新鬼，愁吟独老翁。
18. 发潭州夜醉长沙酒，晓行湘水春。
19. 岸花飞有底急，老去愿春迟。
20. 日暮牛羊下来久，各已闭柴门。

Q10: Discus what kind of output the `shaney` will produce.

Q11: Look at the output above and discuss the differences between these sentences and normal language.

Q12: Discuss how a model of a language can be described.

References

Manning, Christopher D. and Hinrich Schütze (1999) *Foundation of statistical natural language processing*, Cambridge, Massachusetts: The MIT press.