

USING GENETIC ALGORITHM TO GENERATE CHINESE POETRY

LANG5001
Chang,
Zhangyu

WHAT CAN WE GET FROM A COMPUTER?

- Computer Science research papers?
- Philosophical essays?
- Funny short stories?
- Poetry?

WHAT CAN WE GET FROM A COMPUTER?

- “I Spent an Interesting Evening Recently with a Grain of Salt.”
- *“Perhaps only the allegory of simulation is unendurable--more cruel than Artaud's Theatre of Cruelty, which was the first to practice deterrence, abstraction, disconnection, deterritorialisation, etc.; and if it were our own past. We are witnessing the end of the negative form. But nothing separates one pole from the very swing of voting "rights" to electoral...”*

WHAT CAN WE GET FROM A COMPUTER?

- “One day Joe Bear was hungry. He asked his friend Irving Bird where some honey was. Irving told him there was a beehive in the oak tree. Joe threatened to hit Irving if he didn't tell him where some honey was. The End.”
- “Joe Bear was hungry. He asked Irving Bird where some honey was. Irving refused to tell him, so Joe offered to bring him a worm if he'd tell him where some honey was. Irving agreed. But Joe didn't know where any worms were, so he asked Irving, who refused to say. So Joe offered to bring him a worm if he'd tell him where a worm was. Irving agreed. But Joe didn't know where any worms were, so he asked Irving, who refused to say. So Joe offered to bring him a worm if he'd tell him where a worm was...”

WHAT CAN WE GET FROM A COMPUTER?

- Synthesized texts have weird logic
- Synthesized texts are exceptionally obscure

WHAT ABOUT POETRY?

- They have rigid format
- Rhyme requirement
- Not necessarily understandable

WHAT ABOUT POETRY?

- Treat creating poetry as a combinatorial optimization problem.
- Infinite monkey theorem: A monkey hitting keys at random on a typewriter keyboard for an infinite amount of time will almost surely type a given text, such as the complete works of William Shakespeare.
- Proof: http://en.wikipedia.org/wiki/Infinite_monkey_theorem

GAP: THE PROBABILITY IS TOO LOW



GAP: THE PROBABILITY IS TOO LOW

- The number of possible combinations in one line of text is more than the number of atoms in the whole solar system.

50^{50}

GENETIC ALGORITHM

- The idea comes from “evolution” in Biology
- Treat the texts as DNAs (genes)
- Measure the fitness of each text (How?)
- Let the text with higher fitness to get more copy and eliminate those who has low fitness (natural selection)
- Change the remaining text a little bit (mutation)



HOW TO MEASURE THE FITNESS?

- Information theory
- Less probability indicates more information
- “I have a father.”
- “I have a father from Sagittarius.”



HOW TO MEASURE THE FITNESS?

Ah Ah Ah Ah Ah

Ah Ah Ah Ah Ah

Ah Ah Ah Ah Ah

Ah Ah Ah Ah Ah

Ah Eh Ah Oh Oh

Eh Eh Oh Oh Ah

Oh Ah Eh Oh Ah

Eh Ah Oh Oh Ah

Ah Ah Ah Ah Ah

Ah Ah Ah Ah Ah

Ah Ah Ah Ah Ah

Ah Ah Ah Ah **Oh!**

WHAT WE EXPECT?

- More natural and readable text
- By analyzing the parameters of the program, we may get a better understanding of how to write poem.

THANK YOU

EXAMPLES(1)

一个人来到田纳西

- 毫无疑问
- 我做的馅饼
- 是全天下
- 最好吃的

Come to Tennessee Alone

- It is no doubt that
- My pizza
- Is the best
- In the world

EXAMPLES(2)

我终于在一棵树下发现

- 一只蚂蚁
- 另一只蚂蚁
- 一群蚂蚁
- 可能还有更多的蚂蚁

Finally, under a tree I found

- An ant
- Another ant
- A swarm of ants
- Maybe there are more ants