Turing's 1935: my guess about his intellectual journey to "On Computable Numbers"

Kwangkeun Yi

Dept. of Computer Science & Engineering
Seoul National University

7/11/2017 @ The 15th Asian Logic Conference, Daejeon

Turing's 1936 Paper

"*On Computable Numbers, with an Application to the Entscheitungsproblem*" *Proceedings of the London Mathematical Society, ser.2, vol.42 (1936-37). pp.230-265; corrections, Ibid, vol 43(1937) pp.544-546*

- shows a variant of Gödel's Incompleteness proof(1931)
- contains the blueprint of computer (Universal Machine)

My Motivation

Curious: how did Turing get the ideas underlying this foundational paper of modern computer?

- a computer = Universal Machine
- a computer = a single machine that can do any mechanical computation

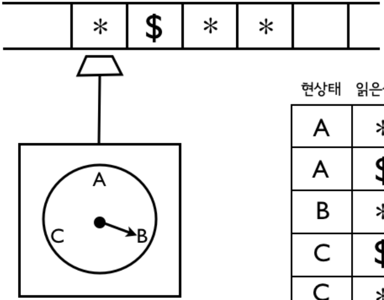
This talk:

- the content of the 1936's paper and
- its intellectual "pedigree" (my guess)

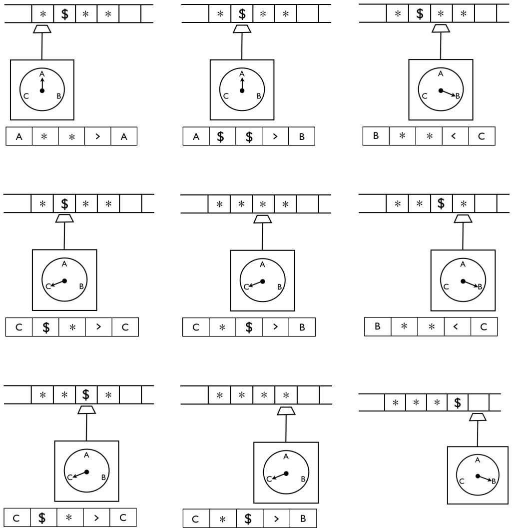
Turing's 1936 Paper

Theorem. By mechanical way we cannot generate all true formulas.

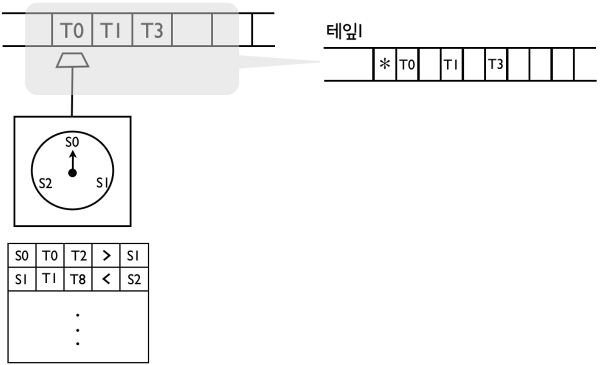
mechanical computation $\stackrel{\text{def}}{=}$ execution by a turing machien(TM)



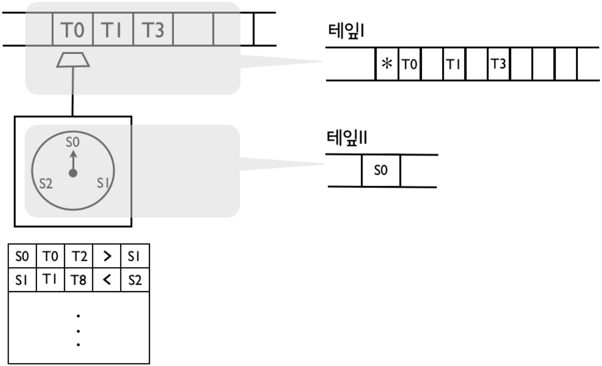
현상태	읽은심볼	쓸심볼	다음칸	다음상태
A	*	*	>	A
A	$	$	>	B
B	*	*	<	C
C	$	*	>	C
C	*	$	>	B


* $ * *
A
C B
A * * > A
* $ * *
A
C B
A $ $ > B
* $ * *
A
C B
B * * < C
* $ * *
A
C B
C $ * > C
* * * *
A
C B
C * $ > B
* * $ *
A
C B
B * * < C
* * $ *
A
C B
C $ * > C
* * * *
A
C B
C * $ > B
* * * $
A
C B

Expressing a TM in a tape (1/3)

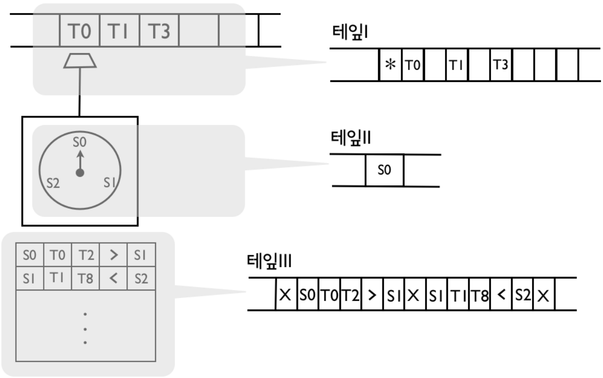


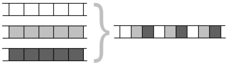
Expressing a TM in a tape (2/3)



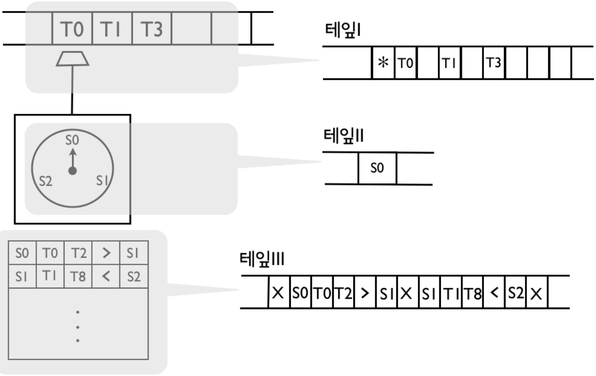
Universal TM: a Turing machine

Expressing a TM in a tape (3/3)





Universal TM: execution rules



- read tape-I, tape-II: $\mathtt{T}j$, $\mathtt{S}i$
- look for match in tape-III for $\mathtt{S}i$, $\mathtt{T}j$

$\mathtt{S}i$	$\mathtt{T}j$	$\mathtt{T}j'$	m	$\mathtt{S}i'$

- do as specified in the matched rule

$|\text{TMs}| \leq |\mathbb{N}|$

The number of TMs cannot be more than $|\mathbb{N}|$

- how many symbols for expressing TM? S,T,<,>,||, 0,$\cdots$,9, X,$*$
- a finite sequence: 17-ary number (a natural number)

Turing's Proof, by Universal TM and $|\text{TMs}| \leq |\mathbb{N}|$

- Lemma1 [$\exists$VERI $\implies \exists H$]. If a TM can generate all true formulas, then it can solve the halting problem.
- Lemma2 [$\nexists H$]. No TM can solve the halting problem.

Thus no TM can generate all true formulas.

Lemma1 proof: $\exists$VERI $\implies \exists H$

If a TM can generate all true formulas, then it can solve the halting problem.
Proof. $H(M) =$

1. run VERI by Universal TM
2. because VERI generates all true formulas, it generates either "M halts" or "not(M halts)."
3. answer accordingly. QED

Lemma2 proof: $\nexists H$

Proof. All TM and its inputs can be indexed by natural numbers: $M_1, M_2, \cdots, I_1, I_2, \cdots$.

1. If H exists, we can fill the following table.

	Input			
	I_1	I_2	I_3	$\cdots$
M_1	1	1	0	$\cdots$
M_2	1	0	1	$\cdots$
M_3	1	0	1	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$

2. Then, following TM is different from all TMs!?

$$M(n) = \textit{Table}(M_n, I_n) \times U(M_n) + 1$$

Contradiction. Hence H is impossible. QED

Turing's 1936 Paper: Wrap up

"*On Computable Numbers, with an Application to the Entscheitungsproblem*"

1. define mechanical computation: turing machine
2. persuade us that TM is enough
3. assume machine VERI that generate all true formulas
4. show that machine VERI can solve the halting problem
5. prove that the halting problem is not computable

Hence VERI is not possible. QED

How Turing come up with the 1936 paper?

"Genius" Turing?

- What talents can generate "foundational knowledge"?
- Only "genius" can do that?
- Misleading message to students
- Not encouraging
- Maybe not true either

Did the 1936 paper come from an epiphany available only to "geniuses"?

Followings are my investigation on Turing's 1935

- Turing took Max Newman's class (Foundations of Mathematics and Gödel's Theorem) in 1935
- Turing learned about Gödel's Incompleteness proof there
- Turing was puzzled; why not more down-to-earth approach?
- Turing began his own style of the same proof

Given a 1st-order finite proof system about natural numbers, the incompleteness holds if such X exists as

$$X \text{ is not provable} = X$$

- X is either true or false.
- Suppose X is false, then X is provable.
  - inconsistent system, out of our discussion
- Suppose X is true, then X is not provable
  - only this is possible

$$X \text{ is not provable} = X$$

Is such X a 1st-order assertion about natural numbers? Gödel showed yes.

- unique natural numbers $\underline{f}$ and $\underline{p}$ for every 1st-order assertion f about natural numbers and every its proof tree p
- "X is not provable" is also an assertion about natural number: "$\underline{X}$ is a factor of a <u>proof</u>"

For $X = \mathit{UnProvable}(\underline{X})$

Gödel proved such X is

$$\begin{aligned} X &\stackrel{\text{def}}{=} G[x \mapsto k] \\ k &\stackrel{\text{def}}{=} \underline{G} \\ G &\stackrel{\text{def}}{=} \mathit{UnProvable}(\underline{\mathit{subst}(x, n, x)}) \quad (\text{Note } \mathit{fv}(G) = \{x\}) \\ & \quad \text{where } n = \underline{x} \text{ and } \mathit{subst}(a : \mathbb{N}, b : \mathbb{N}, c : \mathbb{N}) = \overline{a}[\overline{b} \mapsto c] \end{aligned}$$

because

For $X = \mathit{UnProvable}(\underline{X})$

Gödel proved such X is

$$\begin{array}{rcl} X & \stackrel{\text{def}}{=} & G[x \mapsto k] \\ k & \stackrel{\text{def}}{=} & \underline{G} \\ G & \stackrel{\text{def}}{=} & \mathit{UnProvable}(\underline{\mathit{subst}(x,n,x)}) \quad (\text{Note } \mathit{fv}(G) = \{x\}) \\ & & \text{where } n = \underline{x} \text{ and } \mathit{subst}(a:\mathbb{N}, b:\mathbb{N}, c:\mathbb{N}) = \overline{a}[\overline{b} \mapsto c] \end{array}$$

because

$$\begin{array}{rcl} X & = & G[x \mapsto k] \\ & = & (\mathit{UnProvable}(\underline{\mathit{subst}(x,n,x)}))[x \mapsto k] \\ & = & \mathit{UnProvable}(\underline{\mathit{subst}(k,n,k)}) \\ & = & \mathit{UnProvable}(\underline{G[x \mapsto k]}) \\ & = & \mathit{UnProvable}(\underline{X}) \qquad \text{QED} \end{array}$$

$$
\begin{aligned}
X &\stackrel{\text{def}}{=} G[x \mapsto k] \\
k &\stackrel{\text{def}}{=} \underline{G} \\
G &\stackrel{\text{def}}{=} \mathit{UnProvable}(\underline{\mathit{subst}(x, n, x)}) \\
&\quad \text{where } n = \underline{x} \text{ and } \mathit{subst}(a : \mathbb{N}, b : \mathbb{N}, c : \mathbb{N}) = \overline{a}[\overline{b} \mapsto c]
\end{aligned}
$$

- no nonsense: G has x replaced by itself(encoding k of G)
- two points from equation $X = \mathit{UnProvable}(X)$
  - an assertion about self is expressable within the given proof system
  - can interpret it as specifying an infinite object: a fixpoint of *UnProvable*: $X \stackrel{\text{def}}{=} \underbrace{\mathit{UnProvable}(\mathit{UnProvable}(\cdots))}_{\infty}$

$$
\begin{aligned}
x &= x + 1 \\
x &= x \multimap
\end{aligned}
$$

My Guess about Turing's Thought Process (1/3)

My Guess about Turing's Thought Process (1/3)

- Something infinite in TM must be impossible.
  - Look, something infinite "true yet unprovable assertion" in Gödel's proof.

My Guess about Turing's Thought Process (1/3)

- Something infinite in TM must be impossible.
  - Look, something infinite "true yet unprovable assertion" in Gödel's proof.
- What is infinite in TM then?
  - Every TM is finite, except for its runs

My Guess about Turing's Thought Process (1/3)

- Something infinite in TM must be impossible.
  - Look, something infinite "true yet unprovable assertion" in Gödel's proof.
- What is infinite in TM then?
  - Every TM is finite, except for its runs
- What would be impossible for TM that runs infinitely?

My Guess about Turing's Thought Process (1/3)

- Something infinite in TM must be impossible.
  - Look, something infinite "true yet unprovable assertion" in Gödel's proof.
- What is infinite in TM then?
  - Every TM is finite, except for its runs
- What would be impossible for TM that runs infinitely?
- $X = \mathit{UnProvable}(X)$: this is about self.

My Guess about Turing's Thought Process (1/3)

- Something infinite in TM must be impossible.
  - Look, something infinite "true yet unprovable assertion" in Gödel's proof.
- What is infinite in TM then?
  - Every TM is finite, except for its runs
- What would be impossible for TM that runs infinitely?
- $X = \mathit{UnProvable}(X)$: this is about self.
- What would its correspondence in TM world?

My Guess about Turing's Thought Process (1/3)

- Something infinite in TM must be impossible.
  - Look, something infinite "true yet unprovable assertion" in Gödel's proof.
- What is infinite in TM then?
  - Every TM is finite, except for its runs
- What would be impossible for TM that runs infinitely?
- $X = \mathit{UnProvable}(X)$: this is about self.
- What would its correspondence in TM world?
  - Machine about machines

My Guess about Turing's Thought Process (1/3)

- Something infinite in TM must be impossible.
  - Look, something infinite "true yet unprovable assertion" in Gödel's proof.
- What is infinite in TM then?
  - Every TM is finite, except for its runs
- What would be impossible for TM that runs infinitely?
- $X = \mathit{UnProvable}(X)$: this is about self.
- What would its correspondence in TM world?
  - Machine about machines
  - Machine about machines that runs infinitely

My Guess about Turing's Thought Process (1/3)

- Something infinite in TM must be impossible.
  - Look, something infinite "true yet unprovable assertion" in Gödel's proof.
- What is infinite in TM then?
  - Every TM is finite, except for its runs
- What would be impossible for TM that runs infinitely?
- $X = \mathit{UnProvable}(X)$: this is about self.
- What would its correspondence in TM world?
  - Machine about machines
  - Machine about machines that runs infinitely
- Machine that decides whehter machines would run infinitely or not

My Guess about Turing's Thought Process (1/3)

- Something infinite in TM must be impossible.
  - Look, something infinite "true yet unprovable assertion" in Gödel's proof.
- What is infinite in TM then?
  - Every TM is finite, except for its runs
- What would be impossible for TM that runs infinitely?
- $X = \mathit{UnProvable}(X)$: this is about self.
- What would its correspondence in TM world?
  - Machine about machines
  - Machine about machines that runs infinitely
- Machine that decides whehter machines would run infinitely or not
- BTW, how a machine can see machines? A machine that has machines as its inputs?

My Guess about Turing's Thought Process (2/3)

- Can a machine have a machine as an input? Sure, we can express everying in finite symbols (encodings). And, Gödel too encodes assertions in natural numbers.

My Guess about Turing's Thought Process (2/3)

- Can a machine have a machine as an input? Sure, we can express everying in finite symbols (encodings). And, Gödel too encodes assertions in natural numbers.
- Would it be impossible to decide whether the input machine will run infinitely or not?

My Guess about Turing's Thought Process (2/3)

- Can a machine have a machine as an input? Sure, we can express everying in finite symbols (encodings). And, Gödel too encodes assertions in natural numbers.
- Would it be impossible to decide whether the input machine will run infinitely or not?
- Would it be impossible? I don't know for now. What would be possible? What would be possible, given a machine as an input?

My Guess about Turing's Thought Process (2/3)

- Can a machine have a machine as an input? Sure, we can express everying in finite symbols (encodings). And, Gödel too encodes assertions in natural numbers.
- Would it be impossible to decide whether the input machine will run infinitely or not?
- Would it be impossible? I don't know for now. What would be possible? What would be possible, given a machine as an input?
- "Univeral Machine" is possible. A machine that mimicks the runs of the input machine.

My Guess about Turing's Thought Process (2/3)

- Can a machine have a machine as an input? Sure, we can express everying in finite symbols (encodings). And, Gödel too encodes assertions in natural numbers.
- Would it be impossible to decide whether the input machine will run infinitely or not?
- Would it be impossible? I don't know for now. What would be possible? What would be possible, given a machine as an input?
- "Univeral Machine" is possible. A machine that mimicks the runs of the input machine.
- Possibility expanded, now back to impossibility. I hope the halting problem is impossible. How can I prove it?

My Guess about Turing's Thought Process (3/3)

- One thing I didn't borrow from Gödel's proof is the diagonalization technique.

My Guess about Turing's Thought Process (3/3)

- One thing I didn't borrow from Gödel's proof is the diagonalization technique.
- Yet, Gödel's diagonalization does not fit with TM because it was about assertion formulas. How about Cantor's diagonalization?

My Guess about Turing's Thought Process (3/3)

- One thing I didn't borrow from Gödel's proof is the diagonalization technique.
- Yet, Gödel's diagonalization does not fit with TM because it was about assertion formulas. How about Cantor's diagonalization?

And he proves the halting problem is not computable by TM.

My Guess about Turing's Thought Process (3/3)

- One thing I didn't borrow from Gödel's proof is the diagonalization technique.
- Yet, Gödel's diagonalization does not fit with TM because it was about assertion formulas. How about Cantor's diagonalization?

And he proves the halting problem is not computable by TM.

Then he proves the goal:

My Guess about Turing's Thought Process (3/3)

- One thing I didn't borrow from Gödel's proof is the diagonalization technique.
- Yet, Gödel's diagonalization does not fit with TM because it was about assertion formulas. How about Cantor's diagonalization?

And he proves the halting problem is not computable by TM.

Then he proves the goal:

- $\exists$VERI $\implies \exists H$
  - run VERI and wait & see
  - VERI will eventually print either "M halts" or "not(M halts)"

My Guess about Turing's Thought Process (3/3)

- One thing I didn't borrow from Gödel's proof is the diagonalization technique.
- Yet, Gödel's diagonalization does not fit with TM because it was about assertion formulas. How about Cantor's diagonalization?

And he proves the halting problem is not computable by TM.

Then he proves the goal:

- $\exists$VERI $\implies \exists H$
  - run VERI and wait & see
  - VERI will eventually print either "M halts" or "not(M halts)"
- Contradiction, because $\nexists H$.

My Guess about Turing's Thought Process (3/3)

- One thing I didn't borrow from Gödel's proof is the diagonalization technique.
- Yet, Gödel's diagonalization does not fit with TM because it was about assertion formulas. How about Cantor's diagonalization?

And he proves the halting problem is not computable by TM.

Then he proves the goal:

- $\exists$VERI $\implies \exists H$
  - run VERI and wait & see
  - VERI will eventually print either "M halts" or "not(M halts)"
- Contradiction, because $\nexists H$.
- Hence $\nexists$VERI. QED

Conclusion

My guess about Turing's 1935, on how he came up with his 1936 paper:

Conclusion

My guess about Turing's 1935, on how he came up with his 1936 paper:

- Each step of the paper matches with underlying ideas of Gödel's proof

Conclusion

My guess about Turing's 1935, on how he came up with his 1936 paper:

- Each step of the paper matches with underlying ideas of Gödel's proof
- Turing replayed them in his machine world.
  - Did the 1936 paper come from an epiphany available only to "geniuses"? No fear, maybe not.

Conclusion

My guess about Turing's 1935, on how he came up with his 1936 paper:

- Each step of the paper matches with underlying ideas of Gödel's proof
- Turing replayed them in his machine world.
  - Did the 1936 paper come from an epiphany available only to "geniuses"? No fear, maybe not.
- Turing's 1936 paper, by a collective work than by a single "genius"

Conclusion

My guess about Turing's 1935, on how he came up with his 1936 paper:

- Each step of the paper matches with underlying ideas of Gödel's proof
- Turing replayed them in his machine world.
  - Did the 1936 paper come from an epiphany available only to "geniuses"? No fear, maybe not.
- Turing's 1936 paper, by a collective work than by a single "genius"
  - a teacher who taught the details of Gödel's proof

Conclusion

My guess about Turing's 1935, on how he came up with his 1936 paper:

- Each step of the paper matches with underlying ideas of Gödel's proof
- Turing replayed them in his machine world.
  - Did the 1936 paper come from an epiphany available only to "geniuses"? No fear, maybe not.
- Turing's 1936 paper, by a collective work than by a single "genius"
  - a teacher who taught the details of Gödel's proof
  - a student with self-esteem ("chutzpah" spirit, maybe)

Conclusion

My guess about Turing's 1935, on how he came up with his 1936 paper:

- Each step of the paper matches with underlying ideas of Gödel's proof
- Turing replayed them in his machine world.
  - Did the 1936 paper come from an epiphany available only to "geniuses"? No fear, maybe not.
- Turing's 1936 paper, by a collective work than by a single "genius"
  - a teacher who taught the details of Gödel's proof
  - a student with self-esteem ("chutzpah" spirit, maybe)
  - helps at every its crisis from Gödel's proof

Conclusion

My guess about Turing's 1935, on how he came up with his 1936 paper:

- Each step of the paper matches with underlying ideas of Gödel's proof
- Turing replayed them in his machine world.
  - Did the 1936 paper come from an epiphany available only to "geniuses"? No fear, maybe not.
- Turing's 1936 paper, by a collective work than by a single "genius"
  - a teacher who taught the details of Gödel's proof
  - a student with self-esteem ("chutzpah" spirit, maybe)
  - helps at every its crisis from Gödel's proof
  - inventing "universal machine" as a tool during this process

Conclusion

My guess about Turing's 1935, on how he came up with his 1936 paper:

- Each step of the paper matches with underlying ideas of Gödel's proof
- Turing replayed them in his machine world.
  - Did the 1936 paper come from an epiphany available only to "geniuses"? No fear, maybe not.
- Turing's 1936 paper, by a collective work than by a single "genius"
  - a teacher who taught the details of Gödel's proof
  - a student with self-esteem ("chutzpah" spirit, maybe)
  - helps at every its crisis from Gödel's proof
  - inventing "universal machine" as a tool during this process
  - a teacher who archived Turing's down-to-earth-style proof

Conclusion

My guess about Turing's 1935, on how he came up with his 1936 paper:

- Each step of the paper matches with underlying ideas of Gödel's proof
- Turing replayed them in his machine world.
  - Did the 1936 paper come from an epiphany available only to "geniuses"? No fear, maybe not.
- Turing's 1936 paper, by a collective work than by a single "genius"
  - a teacher who taught the details of Gödel's proof
  - a student with self-esteem ("chutzpah" spirit, maybe)
  - helps at every its crisis from Gödel's proof
  - inventing "universal machine" as a tool during this process
  - a teacher who archived Turing's down-to-earth-style proof

Thank you.