Why most subsets of Σ^* do not have a solution of membership problem?

Alëna Aksënova
LIN629 – Learnability

September 21, 2017

1 General logic of the argument

- A power set of Σ^* – $\mathbb{P}(\Sigma^*)$ – contains all subsets of Σ^*, i.e. all languages.
- Having a solution of membership problem for x in $\mathbb{P}(\Sigma^*)$ means that it is possible to check whether x is or is not in $\mathbb{P}(\Sigma^*)$.
- It means that the set $\mathbb{P}(\Sigma^*)$ must be countable, or enumerable: each entry must come some finite amount of places after the first one.
- The set Σ^* is enumerable.
- The set $\mathbb{P}(\Sigma^*)$ is not enumerable (*Cantor's theorem*).
- The set of all languages $\mathbb{P}(\Sigma^*)$ is not learnable.

2 Enumerability

A set is *enumerable*, or *countable*, if its members can be arranged in an ordered list where each member sooner or later will be encountered. For example, consider this list:

$$1, 2, 3, 4, ...$$

Here, every positive natural number will appear in some finite amount of time. Every element can be *enumerated* – a number can be assigned to it. For example, the function can assign the value n to each positive nth integer.

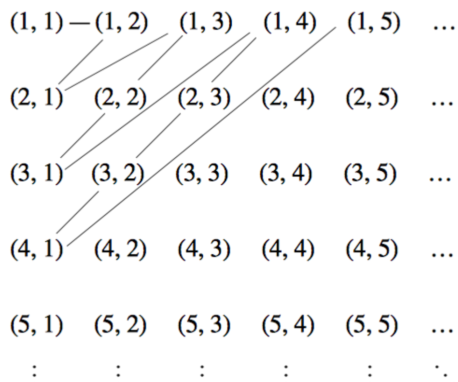
However, this is a bad list:

$$1,\ 3,\ 5,\ ...,\ 2,\ 4,\ 6,\ ...$$

It will not be possible to assign an index to even numbers: they will not be encountered in finite amount of time. **In an acceptable list, each item must appear sooner or later as the nth entry, for some *finite* n.**

To say that the set A is enumerable is to say that there is a function all of whose arguments are positive integers and all of whose values are members of A, and that each member of A is a value of this function.

More complex sequences are countable as well: for example, the set of ordered pairs of natural numbers. Imagine them being ordered in *Cantor's zig-zag manner*:



For such an arrangement, a natural number can be assigned to every pair, and every natural number can be mapped to a single pair of natural numbers. Among others, such sets as Σ^*, rational numbers, finite sets of positive integers, etc. are enumerable.

So what's wrong with the subsets of Σ^*, i.e. with its power set $\mathbb{P}(\Sigma^*)$?

3 Cantor's theorem

Cantor's theorem The set of all sets of positive integers is not enumerable.

The logic of the proof:

1. Introduce operation that finds not-yet-encountered set of integers.
2. Introduce the *diagonalization method* that allows to construct such a set.
3. Show that the number of such sets is not countable.

3.1 How to find a not-yet-encountered set of natural numbers?

The list L is the list of sets of positive integers that have already been constructed:

$$L = S_1,\, S_2,\, S_3,\, \ldots S_n$$

Then we can construct a new set S_{new} following the next rule: **for each positive integer n, n is in S_{new} iff n is not in S_n.** For example, if given are $L = \{1, 2, 4, 5, ...\}, \{1, 3, ...\}$, then the $S_{new} = \{2, 3, 4, ...\}$.

3.2 Are we sure that is will not be the same as any previous list?

Yes. Imagine that this S_{new} already appeared somewhere before, i.e. some $S_m = S_{new}$. But if the number m is in S_m, it cannot be in S_{new}, and the opposite! Contradiction.

3.3 How to find more and more of such new lists?

The list of already encountered sets can be represented as such rectangular array:

	1	2	3	4	
s_1	$s_1(1)$	$s_1(2)$	$s_1(3)$	$s_1(4)$	
s_2	$s_2(1)$	$s_2(2)$	$s_2(3)$	$s_2(4)$	
s_3	$s_3(1)$	$s_3(2)$	$s_3(3)$	$s_3(4)$	
s_4	$s_4(1)$	$s_4(2)$	$s_4(3)$	$s_4(4)$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Here, each row s_n refers to the nth set of numbers in the list L. $s_n(m)$ takes values 0 or 1 depending on whether the number m is present in s_n. Each column is a positive integer.

In this case, the way to form S_{new} is to take a diagonal sequence, and change all ones to zeros, and all zeros to ones, thus obtaining the **antidiagonal** sequence. Changing the values to the opposite will ensure that this sequence has never beed observed before – for every S_m, its mth element will *never* be in S_{new}.

3.4 So why is it impossible to enumerate this L?

A sequence of natural number is possible to enumerate because **each entry comes some *finite* number of places after the first**.

Assume that the set of all subsets of positive natural numbers is countable as well. Then all its elements can be written as enumeration S_1, S_2, S_3, ... S_n – but there is always something to add! It means that no such list can be constructed – **no device is available for arranging all the sets of positive integers into a single infinite list**.

3.5 How is it relevant to languages?

The set Σ^* is countable, because positive natural numbers can be assigned to every member of it. All subsets of positive natural numbers are not countable, as was proven before in the Cantor's theorem, so neither are all the subsets of Σ^*. It means that **the set of all languages is not enumerable**.

The *membership problem* is the problem of deciding whether a string belongs to a set, but in 3.4 it is shown that no such set can be constructed. The result is the following one:

> The set of all languages (all subsets of Σ^*) is not learnable, i.e. it does not have a solution of membership problem.