Why Doesn't My Computer Understand Me?

Jason L. Hutchens

April 29, 1999

INTRODUCTION

Programming a computer to understand natural language is difficult.

Statistical approaches are often used, and have real-world applications.

I will discuss machine translation, conversation simulation, and results of some online language experiments.

Don't expect any technical content at all. Lack of time means that this talk will just need to be interesting instead!

BABELFISH TRANSLATION ENGINE

Altavista offer a machine translation service at `http://babelfish.altavista.digital.com/`

You can translate between English and French, German, Italian, Portuguese or Spanish.

I wrote a web page which translates from English to one of these five languages, then back again.

Lots of people from all over the world have used this page to test the quality of the translations, with amusing results.

FRENCH

“The meek shall inherit the earth” becomes “The soft one will inherit the ground”.

“Silly putty” becomes “Idiotic cement”.

“Jumping Jack Flash” becomes “instantaneous puppet”.

“Kill the peach” becomes “Destroy fishing”.

“race” becomes “mass storage subsystem”.

GERMAN

“Drink up!” becomes “Beverage above!”

“Bill Clinton is desperately trying to avoid impeachment now” becomes “Calculation Clinton tries hopelessly to avoid accusation now”.

“Press ANY KEY” becomes “Press EACH POSSIBLE KEY”.

“We’re all stars now in the dope show” becomes “We’re all asterisks now in the lubricant appearance”.

“Our church has Bingo on Sundays” becomes “Our church has (no meaning) on Sundays”.

ITALIAN

“Well bugger me dead!” becomes “The sink curses it completely!”

“Turn of the lights and come to bed” becomes “Turn of the lights and come in order to insert”.

The rest are far to rude to reproduce here!

PORTUGUESE

“A pot-bellied pig is the perfect pet” becomes “A potentiometer-swelled pig is the perfect animal of esteem”.

“Kick ass!” becomes “Donkey of the backtracking!”

“Only you can prevent forest fires” becomes “You only can hinder fogos of the forest”.

SPANISH

“The spirit is willing, but the flesh is weak” becomes “The alcohol is arranged, but the meat is weak”.

“I saw a deer” becomes “I saw a red deer”.

“have a nice day” becomes “be pleasant”.

CONCLUSION

Machine translation is difficult to do, and tends to fail on short, ambiguous sentences.

Translation is difficult enough for human beings, particularly when translating constrained writing, such as poetry.

As an example, there have been books written using all sorts of constrained writing. These books have actually been translated. (See Hofstadter's "Le Ton beau de Marot" for further information).

Machine translation is helpful for a guideline, but you wouldn't bet your life on the result.

NONI CONVERSATION SIMULATOR

The Loebner Contest is an instantiation of the Turing test.

I have entered the Loebner Contest four years running.

Each time I enter, I try harder to make a better program.

My results have been getting worse every year.

STATISTICAL APPROACH

NonI does it's best to simulate a human being statistically.

It creates a reply to a sentence by estimating $P(w_a|w_q)$.

Up to six candidate reply words are tried in various orders. An "in-gram" statistical model is used to 'smooth" these templates.

A standard Markov model is used to select the most promising sentence to display to the user.

A model of someone typing at a keyboard is used to print the reply, character by character, including errors.

TRAINING

Training the program was an incredibly boring process.

I made a list of thousands of potential questions, taken from a variety of sources.

I sat down, and typed answers to these questions, and the program used these to update all of its statistical models.

After all this effort, NonI came fourth in a field of six!

DISCUSSION

Not enough effort was put into the training. The models were too sparse to generate too many original responses.

The program was not able to maintain context during a conversation, which is a major problem.

Even so, it has come up with some humdingers:

“A symbol is a suspended brass disk which represents something else”.

ONLINE LANGUAGE EXPERIMENTS

I wanted to see how well human beings perform at the sort of language tasks computers are being used for.

Almost 700 people have participated in these experiments since they began in August, 1998.

The experiments consist of story writing, prediction, segmentation and smoothing.

I have recently only started to analyse the results.

STORY WRITING

Murder 2: “Detective Snadsworth ate some poisoned policeman who roughly shook the woman. Her dead limb fell off. Unconcious upon contact he strangled many necks.”

Freeform 5: “Never before has anyone, ever, attempted flight? Never. Man could dream about anything under the bed. Preposterously, nothing could stop his frantic hands from tearing newspaper apart.”

Murder 15: “Yet still I cannot believe that Richardson willfully destroyed the evidence. There must be some trace of an icky blot that she had seen. Clutching her purse nervously, she decided to kick open the briefcase.”

PREDICTION

Yesterday morning, I went out to collect my newspaper, and found that the...

Paper, Newspaper, Paperboy, Dog, Delivery, Rain, Mailbox

...delivery boy had...

Not, Thrown, Forgotten, Left, Put, Tossed, Been, Missed

...thrown it in the...

Bushes, Bin, Garden, Pool, Mailbox, Trash, Window, Bush

SEGMENTATION 1

HINGNOTHINGTHENTHENTHENHENOTHINGHEHETHEN
THENHENOTHINGTHENTHENTHENTHENHENOTHINGNO
THINGTHENTHENHEHENOTHINGTHENNOTHINGNOTHI
NGTHENNOTHINGTHENTHENNOTHINGHEHEHENOTHIN
GNOTHINGNOTHINGNOTHINGTHENHEHEHEHENOTHIN
GHETHENHEHETHENNOTHINGHEHEHENOTHINGTHENH
EHETHENHETHENNOTHINGHETHENHETHENTHENNOTH
INGTHENNOTHINGNOTHINGNOTHINGTHENHEHETHEN
THENHEHEHENOTHINGTHENHEHENOTHINGNOTHINGT
HENNOTHINGHENOTHINGTHENNOTHINGHETHENTHEN
HENOTHINGHEHENOTHINGNOTHINGNOTHINGTHENNO
THINGTHENTHENNOTHINGNOTHINGHETHENHEHETHE
NTHENNOTHINGTHENTHENNOTHINGNOTHINGNOTHIN
GNOTHINGHEHEHENOTHINGNOTHINGTHENNOTHINGT
HENNOTHINGHETHENNOTHINGHETHENHETHENNOTHI

SEGMENTATION 2

It is found that human beings segment this and other sequences fairly well.

In each of the five experiments, over 63% of people marked each of the correct words, and less than 44% of people marked words which were not part of the sequence.

Segmentation, therefore, would be possible via a simple thresholding process.

Some alphabets, such as “ABBA”, “AAA”, “CB” produced ambiguous sequences; even so, words such as “BAAB”, “AAAAAA” and “CBC” were correctly left out.

SMOOTHING 1

Chickadee,
I decree
a fine day.
Dart away
from your (BLANK)
and engage
in brave flight.

This is like playing blankety-blanks.

Have a guess at the missing word!

SMOOTHING 2

The word “cage” is, of course, the correct word.

On average, it took people 4 guesses to correctly identify it.

34.6% of people got it right on the first guess. If we fake a probability distribution from the guesses, we find the correct word has 1.53 bits of information.

Some examples of guessed words, sorted by decreasing frequency, are “cage”, “nest”, “tree”, “home”, “plight”, “play”, “fight”, “rage”, “light”, “night”, “branch”, “perch” and so on.

PREDICTION 3

And God made two great lights; the greater light to rule the day, and the lesser light to rule the night: he made the (BLANK) also.

On the first guess, words guessed included “stars”, “earth”, “sun”, “dark”, “night” and “sky”.

Some users didn’t guess the word correctly at all. By the ninth guess, they were really clutching at straws.

Some of the guessed words include “heavens”, “air”, “lawyers”, “rules”, “internet”, “photons” and “twilight”.

FEEDBACK

People really had a lot of fun participating, and often mentioned experiencing a sense of achievement in getting a prediction right.

They often had difficulty expressing how they were able to predict words so well, and often expressed surprise at their abilities.

Some sequences were random; in these cases, they claimed to have discovered patterns where no such patterns existed.

After completing a prediction experiment on artificial language data, one user commented "I don't know how I did that".

CONCLUSION

Natural language is a difficult form of pattern recognition. Often, near enough is not good enough. Subtle changes can drastically alter meaning.

Statistical approaches to natural language learning seems to have lots of potential, but still omits a lot of important contextual clues.

Human beings easily out perform computers. A text compression device based upon a human computer would perform remarkably well.

Language understanding is a classic AI problem. People have been working on it for decades, yet we're only just beginning to make slow progress, and there's a long way to go.