Citation for the hugging face course:

@misc{huggingfacecourse,

author = {Hugging Face},

title = {The Hugging Face Course, 2022},

howpublished = "\url{https://huggingface.co/course}",

year = {2022},

note = "[Online; accessed <today>]"

}

Searching history:

**robot commands with natural language processing**

sort by most citings.

Did not find anything interesting. The search response consists of articles about the theory behind HRI and general communication methods. As this is not the focus of my thesis, then the search is deemed unfruitfull.

# robot language commands using neural network

sort by citations [by papers]

Publication topics:

* Speech recognition
* Natural language processing
* Neural nets (recurrent, convolutional, and feed forward)
* AI
* Intelligent robots
* Natural language
* Natural language interfaces.
* Mobile robots
* Control engineering
* Gesture recognition
* Robot vision

# Automatic Translation of Spanish Natural Language Commands to Control Robot Comands Based on LSTM Neural Network

# [Automatic Translation of Spanish Natural Language Commands to Control Robot Comands Based on LSTM Neural Network | IEEE Conference Publication | IEEE Xplore (sdu.dk)](https://ieeexplore-ieee-org.proxy1-bib.sdu.dk/document/8675641)

Uses a RNN encoder decoder structure to translate natural language into an intermediate action language called Robot Control Language (RCL), this has the purpose of translating the language into a non-interpretable text, which is then usually used for mobile robotics.

“This study is meant to propose a high level layer component able process commands in natural language with the purpose of generating an intermediate action language, a formal representation that can be used by downstream layers in a robotic architecture. The intermediate action language that was chosen is Robot Control Language (RCL), this is a motion oriented language made of route instructions through an indoor environment containing objects and landmarks [12].”

The key difference in this article compared to my thesis, is that the neural network is used directly to translate the text to an action language which is more easily parsed into a robot. While I use a parser to create the robot action commands.

Another difference is the use of LSTM’s instead of my proposed model of choice, which is a transformer model.

The model was trained by using seven different categorizable actions which all sentences would fall into, then base sentences were made for each of the actions, which served as a data tag and data set respectively.

Afterwards different words in the sentence were exchanged with their synonyms to create a bigger database of words corresponding to the correct actions.

“using the “related works” section, I found more articles that matched my searching criteria”

[Commanding mobile robot movement based on natural language processing with RNN encoder­decoder | IEEE Conference Publication | IEEE Xplore (sdu.dk)](https://ieeexplore-ieee-org.proxy1-bib.sdu.dk/document/8391185)

[Indonesian natural voice command for robotic applications | IEEE Conference Publication | IEEE Xplore (sdu.dk)](https://ieeexplore-ieee-org.proxy1-bib.sdu.dk/document/7352577)

# Approximate Decision Making by Natural Language Commands for Robots

[Approximate Decision Making by Natural Language Commands for Robots | IEEE Conference Publication | IEEE Xplore (sdu.dk)](https://ieeexplore-ieee-org.proxy1-bib.sdu.dk/document/4153761/authors#authors)

This article explains how a fuzzy controller system used along with a fuzzy voice command can be used in combination with a probabilistic neural network to perform bin picking actions. Were the main learning, comes from figuring out what the voice commands are, and translating it into where on the table, the object is.

The theory behind fuzzy voice commands, and how the neural network is implemented in the fuzzy controller is unknown, but the core gist, is that the user can control the robot with voice (fuzzy controller part), and correct the robot, if it moves incorrectly, such that the robot learns to move in the correct way in the future (neural network part)

The main part of this paper is that it uses user feedback to train the neural network if the robot does incorrect actions. Therefore, enabling human correction of the robot on the fly.

# Development of Indonesian Speech Recognition with Deep Neural Network for Robotic Command

[Development of Indonesian Speech Recognition with Deep Neural Network for Robotic Command | IEEE Conference Publication | IEEE Xplore (sdu.dk)](https://ieeexplore-ieee-org.proxy1-bib.sdu.dk/document/8937275)

Using a regular Deep neural network, the authors could create a natural language processing model, which was capable of categorizing words directly into one of six different actions. The final model achieved a success fail ratio of 89.57% based on 10.521 datasets.in which the testing set consisted of 10% of the data, meaning, 1052 datasets.

The paper focuses more on the audible part of the process of natural language processing, making it slightly different from the focus I want to put on my thesis. Though their results using a DNN is very interesting, as the success fail ratio clearly states, that the DNN is capable of being a good natural language processing model, given that the actions are simplistic (only six of them)

**Commanding mobile robot movement based on natural language processing with RNN encoder­decoder**

[Commanding mobile robot movement based on natural language processing with RNN encoder­decoder | IEEE Conference Publication | IEEE Xplore (sdu.dk)](https://ieeexplore-ieee-org.proxy1-bib.sdu.dk/document/8391185)

Based on an overview, what they have done, is that they have split the task into many smaller tasks, so that the RNN decoder has less work to train.

First, they have a pre-processing step which has the purpose of splitting the sentence, such that each part only has one robot command action with its parameters. Furthermore, it also serves to filter out non important text.

Secondly, they use an RNN encoder decoder model to classify each of the sentences by their probability of being one of the actions.

Last of all there is a post processing function that maps the output from the RNN into actual robot code.

The interesting part of this paper is that it uses a pre-processing part, to make sure that the input for the RNN is only considered to be a single language command. Furthermore, the pre-processing serves to filter out uninteresting words which has no meaning for the text.

For the data, they use their own, by making volunteers write down their preferred natural language command given a picture of a robot and a direction. This sums up to 1600 unique datasets, in which they achieve an accuracy of 79.23% and 73.65 % on sentences with multiple commands.

**Indonesian natural voice command for robotic applications**

[Indonesian natural voice command for robotic applications | IEEE Conference Publication | IEEE Xplore (sdu.dk)](https://ieeexplore-ieee-org.proxy1-bib.sdu.dk/document/7352577)

Instead of using a parser to analyse the text grammatically, they use their own setup to create a low-level text analyser using an RNN. Based on the fact, that one hot vector methods work to slow, because of the computational complexity, an alternative method is produced, where a neural network is used, such to exploits its effectiveness, such to make methods close to one hot encodings work. Instead of using a vector, they use a Vector space model, which is more akin to a matrix.

The RNN is used to map each word into its correct location in the vector space, outputting a matrix of locations, based on the words, as its output. Words that are close to each other is then approximated to mean the same thing. With” Go”, and ”move” as an example. A decoder is then used to translate the VSM RNN output to robot actions.

It should also be noted, that there is a lot of theory that is not completely understood in this article. How the voice to text part is made, though it is not important for my thesis. How does the decoder translate from VSM to action is not stated.

Diagram, engineering drawing

Description automatically generated

Figure 5, is completely disregarded, as it is never mentioned, and that it counteracts the part of the article that states, that the RNN is used to create the VSM.

It does seem, that their focus lies more on the voice to text part, than the natural language processing part, which means it is of less importance to my thesis than the other articles.

# From Jannik

# [The Illustrated Transformer – Jay Alammar – Visualizing machine learning one concept at a time. (jalammar.github.io)](https://jalammar.github.io/illustrated-transformer/)

Transformers can be used instead of convolutions for images too.

Transformers is an encoder decoder model. There are more different kinds. Like the decoder models, which specialise in answering, like GPT models. Grammar analysis models are encoder based like BERT.

Unsupervised pre-training method. Predict the next word in a sentence given an arbitrary large input word size. The output is then compared to the correct output (probably using word embeddings), such that the error can be fed through the neural network.

The two major component of the multiheaded attention layer.

The self-attention layer which combines words to each other, such to create understanding of substitute words like “it”, “him”, “that”, etc. the other component is the feed forward network which is used to compute the output from the self-attention layer.

Language models don’t see words, they only see and work with lists of numbers.

Token ID’s = vocabulary index number

Word embeddings = multi-dimensional space point unique per word