Grounding Test Result – Cross Validation

# Instruction

Figure 1. The dataset used in the experiment. All the commands are manually labeled as ground truth

In this test, I examined my algorithm by cross validation. I manually labeled the ground truth grounding for each chunk in the 149 template corpus which has been turn into chunking tree structure by Tatiana (Figure 1). There are 820 chunks in my dataset. In the test I randomly selected 36 descriptions with around 200 chunks from template corpus for training and put the left ones into test group. I performed 30 tests and the result is shown in Table 1. The result is also shown by histogram diagram in Figure 2 and Figure 3.

# Result

Table 1. Cross validation result. Chunk Grounding correct rate is the portion of chunks that correctly grounded. Command grounding success rate is the portion of commands that successfully parsed which means all the chunks in the commands are correctly grounded. Command grounding success number is the number of commands that successfully parsed.

|  |  |  |  |
| --- | --- | --- | --- |
| Index | Chunk Grounding Correct Rate | Command Grounding Success Rate | Command Grounding Success Number |
| 1 | 0.976829 | 0.885906 | 132 |
| 2 | 0.931707 | 0.771812 | 115 |
| 3 | 0.947561 | 0.805369 | 120 |
| 4 | 0.979268 | 0.885906 | 132 |
| 5 | 0.967073 | 0.85906 | 128 |
| 6 | 0.936585 | 0.778523 | 116 |
| 7 | 0.981707 | 0.899329 | 134 |
| 8 | 0.94878 | 0.738255 | 110 |
| 9 | 0.957317 | 0.872483 | 130 |
| 10 | 0.992683 | 0.959732 | 143 |
| 11 | 0.982927 | 0.90604 | 135 |
| 12 | 0.947561 | 0.765101 | 114 |
| 13 | 0.967073 | 0.845638 | 126 |
| 14 | 0.968293 | 0.838926 | 125 |
| 15 | 0.990244 | 0.946309 | 141 |
| 16 | 0.97561 | 0.865772 | 129 |
| 17 | 0.945122 | 0.785235 | 117 |
| 18 | 0.965854 | 0.812081 | 121 |
| 19 | 0.971951 | 0.852349 | 127 |
| 20 | 0.987805 | 0.932886 | 139 |
| 21 | 0.945122 | 0.744966 | 111 |
| 22 | 0.979268 | 0.885906 | 132 |
| 23 | 0.980488 | 0.892617 | 133 |
| 24 | 0.960976 | 0.818792 | 122 |
| 25 | 0.982927 | 0.90604 | 135 |
| 26 | 0.969512 | 0.872483 | 130 |
| 27 | 0.985366 | 0.919463 | 137 |
| 28 | 0.964634 | 0.805369 | 120 |
| 29 | 0.987805 | 0.932886 | 139 |
| 30 | 0.976829 | 0.872483 | 130 |



Figure 2 Histogram of chunk grounding correct rate



Figure 3 Histogram of command parsing success rate

# Discussion

The result shows that the current grounding algorithm is able to parse chunked commands to robot navigation. The cross validation proves the reliability of our work. However, I found that the current Levenshtein Distance (LD) which is used to measure the similarity between arrays is not very suitable to our current work. We use text in chunks from training commands as template and compare them with the text from a sample chunk to find the best matching grounding of that sample chunk. However, the traditional LD just measure similarity from chunk structure and words. It does not consider the meaning of the words in the chunks. It sometimes declared larger similarity on a chunk which has a very similar structure but an opposite meaning to the sample chunk. In my following work, I will try to give different weights to the words in chunks so that the new distance equation can judge similarity between chunks from meaning.