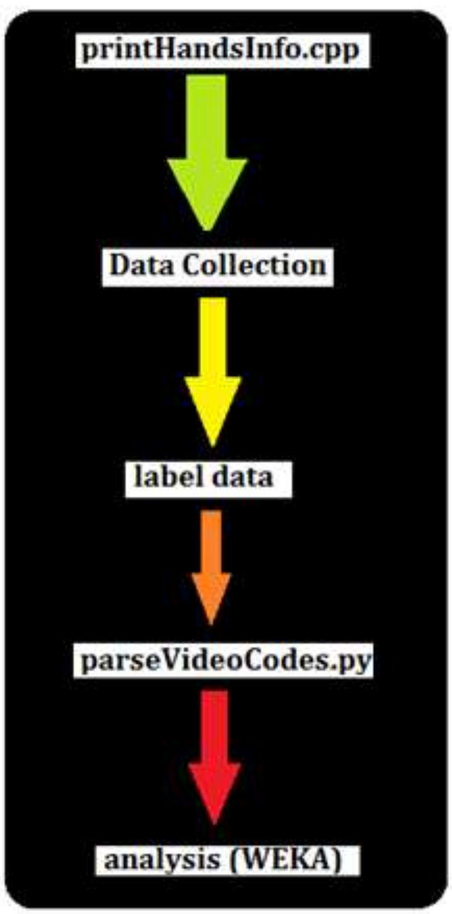
How We Build

Micah Arvey

Abstract

How do we go about building things? What steps do we take, and in what order? These are the fundamental questions addressed in this project. Using information gathered from real people (via the Xbox kinect, recorded by a web cam, and labeled by hand) machine learning algorithms are trained on this data to label the actions of unknown builders. Subjects are asked to build simple (albeit intricate) designs to satisfy few constraints. Given straws, tape, and a paper plate, participants are asked to build a structure to hold a pingpong ball, without taping anything to the table. Immediately they begin testing their items and preparing to build and are recorded every step of the way. Building is a part of everyday life. From construction of makeshift desk aids to architecture, there is a lot to be learned from the theory of building. With more information comes the possibility of enhancing our abilities and bringing new light to the world of construction.

Introduction



The aim is to automatically label building sessions based on hand motion. How well a computer can classify actions based on physical cues? We will specifically be focusing on upper body cues, and even more specifically on the right and left hands. The Xbox kinect has the capability to capture the x, y, and z positions of the right and left hands and is just waiting to be hacked.

However, some challenges present themselves at this stage. For instance, which body parts will be most useful? Where do we set the kinect? How will we measure the data (discrete? Continuous? Time steps?). Data is the most important piece in this puzzle (aside from appropriate analysis) so these questions are important. The next question is how to organize this data. There is manually labeled data, kinect recorded data, video data, and semantic data for every experiment.

There are many Machine Learning algorithms which could be run – to wit there are many available libraries which do a lot of the work, but it is first necessary to get and format the input data. One available tool is called WEKA, from the University of Waikato. The graphical user interface makes it easy to apply various machine learning algorithms and discover underlying trends, as well as generate charts.

Data Gathering

In such an experiment, the training data is very important and was undertaken carefully. I was very excited about the prospect of gathering completely new data rather than using stock data from a database. This process would prove to be a time consuming one. First to hack the Xbox kinect, then to record real people, label the data, and then to synthesize the data into something useful.

The kinect API was fairly easy to work with. I downloaded the API from Microsoft online and began development. The script is a little longer than 100 lines of code. I specified which joints to measure (only upper body) and generated time stamps every 100 milliseconds. This would all output to a comma separated value file.

-0.06908	-0.34671	2.31807	-0.05483	-0.34597	2.35563	-0.0078
0ms						
-0.02588	-0.3644	2.2985	-0.01969	-0.35894	2.33887	-0.00963
100ms						
0.001119	-0.3808	2.27902	-0.00256	-0.36801	2.32397	-0.01774
200ms						
-0.01017	-0.32962	2.26112	-0.00731	-0.29869	2.28744	-0.02087
300ms						
-0.02172	-0.30355	2.2448	-0.0159	-0.26505	2.25619	-0.03535
400ms						
-0.05084	-0.29066	2.22848	-0.04767	-0.25105	2.22596	-0.08883
500ms						

```
[00:02:05.27] Selecting Materials
[00:02:09.14] Single Object Examination
[00:02:12.27] Single Object Examination
[00:02:17.06] Single Object Examination
[00:02:21.10] Single Object Examination
[00:02:28.08] Protyping Mechanism
[00:02:42.06] Organizing
[00:02:45.15] Protyping Mechanism
[00:02:54.14] Think w/object
[00:03:00.12] Protyping Mechanism
[00:03:04.16] Think w/o object
[00:03:09.11] Single Object Examination
[00:03:13.24] Single Object Examination
[00:03:20.10] Single Object Examination
[00:03:25.10] Think w/object
[00:03:32.17] Selecting Materials
[00:03:38.11] Protyping Mechanism
[00:03:45.19] Protyping Mechanism
[00:04:16.06] Building
[00:04:39.26] Building
```

Very special thanks go out to Marcelo Worsely in the Transformative Learning Technologies Lab for bringing high school students in to test and gather data. We labeled this data by watching videos and creating time sheets with mappings to the intended action as determined by us.

The choice of features becomes an important consideration at this point. The features desired include hand position and hand displacement (delta position from previous time stamp). Formulation of when to label is also important. The decision here was that it would be easy enough to manipulate the data form python after recording it by hand in the way displayed above.

Data Manipulation

Python was used to translate the data from unformulated into a comma separated value files ready for analysis. NumPy is a statistics package for python and was used for the creation of aggregated data. The main concern during this step was whether to display the data on a "by time interval" basis (aka use the intervals delineated by our labeling) or to have a fixed time interval (½ second for example). Here lay the decisions of features which was in turn fueled by the analysis.

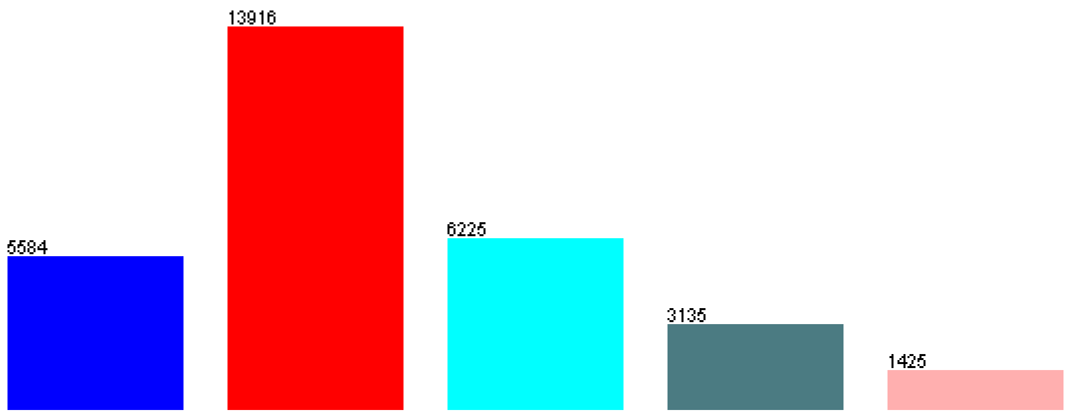
action	r_avg	r_std	r_sum	l_avg	l_std	l_sum	avg_diff
protyping mechanism	0.090986	0.087619	21.47258	0.081627	0.076091	19.26398	0.065463
building	0.055849	0.051341	6.143355	0.074113	0.063996	8.152452	0.060883
adjusting	0.094156	0.078697	1.224031	0.064492	0.040562	0.838399	0.06047
building	0.12237	0.05958	1.346071	0.10329	0.083153	1.136194	0.060739
adjusting	0.093383	0.07173	7.003704	0.085591	0.072578	6.419292	0.061552
testing mechanism	0.114629	0.07478	1.031665	0.09434	0.110464	0.84906	0.062387
adjusting	0.064654	0.059756	2.392198	0.069843	0.076844	2.584173	0.061434
building	0.03001	0.036192	0.57019	0.051703	0.044633	0.982364	0.060543

Analysis

After training, the classification algorithm should be able to classify each action a person takes during the building process. The strengths of this approach are clear; if we choose a good domain of classifications and good features, we can hopefully solve our problem. An apparent weakness is that choosing correct features and formulating them in conjunction with our time stamps is hard.

WEKA was used for analysis. Originally, nearly all of the ½ second intervals were classified as building. This resulted in a false sense of accomplishment as it most of the intervals actually were "building" as shown below.

1	protyping mechanism	5584
2	building	13916
3	adjusting	6225
4	testing mechanism	3135
5	undo	1425



With fixed time intervals therefore, it was difficult to convince the machine to choose anything but building. A new approach was to aggregate the data and use the time intervals delineated by the label files. The domain space of classification was shrunk to 5 items (listed above) This proved to be a valid approach when the correctly classified percentage surpassed the percentage of "building" predicted intervals. 34.3% accuracy was determined by the Naïve Bayed algorithm and shown below from the output of WEKA. 34.3% is 175% better than the baseline of 20% accuracy when based off of random guessing.

```
=== Summary ===

Correctly Classified Instances         260               34.3461 %
Incorrectly Classified Instances       497               65.6539 %
=== Confusion Matrix ===

   a   b   c   d   e   <-- classified as
  15  34   1 119   0 |   a = protyping mechanism
  12 103   4  82   0 |   b = building
   6  47   3 110   0 |   c = adjusting
   5  11   8 139   0 |   d = testing mechanism
   1   3   1  53   0 |   e = undo
```

Conclusion

Though the result is not staggering it shows that there is (obviously) insight on what action people are taking based on physical action motions. For further analysis, an HMM (Hidden Markov Model) may be appropriate because it is obvious that once a person starts a task, they will likely continue that task for a certain period of time and therefore it can be construed that the previous action has some influence on the next action. Also information could be gathered from open CV to give data about how close the hands are to the objects.

References

Special thanks to Marcelo Worsely for gathering data

Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009); The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.