Contextual Code Completion
Subhasis Das, Chinmayee Shah

My code completion so dumb

```
int n, i;
for(i = 0; ... 
```

What goes here?

Me

```
int n, i;
for(i = 0; i < ???; i++) {
```

My computer

```
int n, i;
for(i = 0; ???
```

Window Based Approach

Last W tokens to predict the next token

Dictionary = Key + Positional
- frequently occurring key words
- language key words
- functions
- variables

```
int n , i ;
k2 p2 k10 p4 k1
for ( i = 0 ; ... 
k7 k5 p4 k21 p10 k1
```

Modeling Tokens

word2vec: token → dense vector

```
k1 { 0.3, 0.1, -0.5, ...}
p1 { -0.8, 0.9, 0.4, ...}
```

Total # tokens = # keywords + # positions

Learning Approaches and Results

Dense 3-Layer Model

Attention Model

GRU Model

Training and Test Sets

Linux - kernel source, C project (50-50 train/test split)
Twisted - networking library, python project (50-50)

<table>
<thead>
<tr>
<th>Dataset/Method</th>
<th>Accuracy (Known)</th>
<th>Top-3 Accuracy</th>
<th>Key word Accuracy</th>
<th>Non-key Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linux/Dense-3</td>
<td>64.5%</td>
<td>81.8%</td>
<td>78.2%</td>
<td>43.6%</td>
</tr>
<tr>
<td>Linux/Dense-4</td>
<td>63.2%</td>
<td>82.0%</td>
<td>72.8%</td>
<td>41.8%</td>
</tr>
<tr>
<td>Linux/Attn</td>
<td>67.6%</td>
<td>83.6%</td>
<td>80.0%</td>
<td>48.4%</td>
</tr>
<tr>
<td>Twisted/LSTM</td>
<td>41.6%</td>
<td>59.1%</td>
<td>64.4%</td>
<td>4.9%</td>
</tr>
<tr>
<td>Twisted/Dense-3</td>
<td>46.3%</td>
<td>64.8%</td>
<td>64.0%</td>
<td>14.3%</td>
</tr>
<tr>
<td>Twisted/Dense-4</td>
<td>38.9%</td>
<td>53.9%</td>
<td>56.4%</td>
<td>7.3%</td>
</tr>
<tr>
<td>Twisted/Attn</td>
<td>46.6%</td>
<td>64.7%</td>
<td>62.5%</td>
<td>18.5%</td>
</tr>
</tbody>
</table>