Large Scale Taxonomy Mapping: From E-commerce Platforms to Google

Ma Yunshan, Ming Zhaoyan, Chua Tat-Seng
yunshan@comp.nus.edu.sg, dcsming@nus.edu.sg, dcscts@nus.edu.sg

What is E-commerce Taxonomy

Google helps online retailers to better advertise their products if their products are labelled with google taxonomy.

It will be much easier for consumers to find what they want to buy.

Each online retailer has large number of products, yet manually label will cost a lot of money and manpower.

If we could automatically map the taxonomy of each platform to google’s, all the products will be labelled to google taxonomy.

How large are they

<table>
<thead>
<tr>
<th></th>
<th>Google</th>
<th>Amazon</th>
<th>Walmart</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>21</td>
<td>38</td>
<td>28</td>
</tr>
<tr>
<td>Level 2</td>
<td>191</td>
<td>262</td>
<td>421</td>
</tr>
<tr>
<td>Level 3</td>
<td>1330</td>
<td>1642</td>
<td>2136</td>
</tr>
<tr>
<td>Level 4</td>
<td>2135</td>
<td>3212</td>
<td>None</td>
</tr>
<tr>
<td>Level 5</td>
<td>1318</td>
<td>2942</td>
<td>None</td>
</tr>
<tr>
<td>products</td>
<td>not available</td>
<td>20,000,000</td>
<td>6,000,000</td>
</tr>
</tbody>
</table>

We do the mapping

Ongoing work

- Manually map Walmart taxonomy level 3 to Google level 4, and have a whole understanding of the taxonomies.
- Train textCNN based classifiers using products’ title and description, to see the potential of automatic taxonomy mapping via deep learning techniques.
- We just treat each of the leaf nodes as independent classes, while there’s a serious problem because of unbalanced data. After using some methods of data augmentation, the macro accuracy improves significantly.

Next Step

- Use a part of the manually labelled mappings to predict other mappings.
- Use both text and image information to collaboratively do the mapping and classification.
- Nodes that cannot map to google taxonomy should be added to it, then where to put these new nodes is a question.

Accuracies of the products classification based on textCNN.

Micro Accuracy on Training Set
Macro Accuracy on Test Set
Micro Accuracy on Test Set

Scan the QR code to view some of the taxonomies online