Business and Resume Analysis Engine
An Automated Approach to Classification and Identification

Adeoluwa Babatunde (axb131330@utdallas.edu), William J. Hester (wjh130030@utdallas.edu), Roman Lubarsky (ral110020@utdallas.edu), and Jassi Sikand (jxs131730@utdallas.edu)

Abstract

Bridge Alliance requires a tool to augment human efforts in “(curating) and (cultivating) startup information from various websites and documents”, and Zigatta requires a tool to “review IT related job orders.” Such a tool is needed because recruitment is a time-consuming, error-possible, manual effort and the current way allows for missed opportunities and incorrect rankings. We propose a tool general enough to solve both problems using machine learning algorithms: this tool will be built atop Facebook FastText. Such a tool will allow analysts at both companies to work with large sets of data more efficiently which will save recruiters potentially immense amounts of time as they will be more able to focus on interactions between companies and candidates.

Architecture

Polymer
- Web Design Framework which allows for Easy Dynamic Data Presentation
- Uses Android-based Material Design

Businesses
- You can add businesses to the database by submitting a name and a URL. Predictions from the scrape are displayed immediately.
- You can also search the database via labels, which will return a set of companies ordered by how they match the labels and then alphabetically.

Resumes
- You can submit a candidate for evaluation by submitting a name, a resume to analyze, as well as an optional URL to go scrape
- You can search for candidates by label as well, which will return results in the order of highest matching percentage, and then alphabetically.

Node.js Backend
- Text Extraction from Uploaded Documents (Convert)
- Text Scraping from a Given URL using PhantomJS (Scrape)
- Multi-class Text Classification using Facebook fastText (Predict)
- Retrieval of Documents by Predicted Class/Label (Search)

Impact

- Reduce the manual effort and labor required to browse and deduce meaning from resumes and companies for Zigatta and Bridge Alliance respectively.
- Provide a foundation for machine learning applications in many other fields such facilitating medical or governmental record keeping.

Metrics

- **Usability:** Survey sent to seven users. Mean scores (out of max. 5) reported below. Project met our goal of all mean scores >4.

<table>
<thead>
<tr>
<th></th>
<th>Usability</th>
<th>Accuracy</th>
<th>Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resumes</td>
<td>4.29</td>
<td>4.33</td>
<td>4.67</td>
</tr>
<tr>
<td>Companies</td>
<td>4.67</td>
<td>4.67</td>
<td>4.50</td>
</tr>
</tbody>
</table>

- **Machine learning:** fastText tested against validation set. On resumes task, model’s top prediction matched top label (of 7) 87.5% of the time.
- **Companies:** top prediction matched top label (of 44) 48.7% of the time
- Project met our goal of significantly exceeding accuracy of random classifier (for resumes, 14.28%; for companies, 2.27%)

Summary

- By creating the ability to submit documents for classification by a Machine-Learning Model via an easy-to-use website interface using a clean Polymer Framework, we have reduced the complexity of searching for particular candidates and companies.
- Now, recruiters can spend more time interfacing with candidates and companies, and less time having to deal with frustration of searching for applicable candidates.
- Since the Model depends on a specific training set, if this project was to be expanded, each customer/recruiter could have a solution tailored to how they decide what makes an applicable candidate for a position. This reduces the possibility of needing to question how successful the Model is for the Recruiter.
- According to one project sponsor, this solution ultimately changes the way that recruiting is done by simplifying the process immensely.