Data Analytics to Improve Co-Operative Education

Shivangi Chopra, Yuheng Jiang, Andrew Toulis, Lukasz Golab
University of Waterloo
Waterloo, Ontario, Canada
{s9chopra,y29jiang,aptoulis,l golab}@uwaterloo.ca

ABSTRACT
In this paper, we summarize our recent research on applying data analytics to a new application area: co-operative education. Many post-secondary institutions currently offer co-operative programs in which students alternate between on-campus classes and off-campus work terms. We observe that the co-operative process produces a variety of interesting data including job advertisements and performance evaluations. We discuss novel data science methodologies we applied to these datasets and the business insights we obtained.

1 INTRODUCTION
According to the World Association for Cooperative and Work-integrated Education, 275 institutions from 37 countries offer co-operative education (co-op) programs, also known as work-integrated learning. Students enrolled in a co-op program typically alternate between on-campus classes and off-campus work terms/internships. Co-operative education has become popular worldwide as it provides an enhanced learning experience for students and a talent pipeline for employers.

Related Work. Research on co-operative education considers three perspectives: of the student, of the employer and of the educational institution. From the student’s perspective, the focus has been on the impact of co-op on skill and career growth, and on characterizing the attributes that make co-op students successful based on survey data and workplace supervisor evaluations. From the employer’s perspective, there has been work on studying employer expectations (see, e.g., [6, 9, 10, 17, 22, 23]). From the institution’s point of view, the focus has been on assessing the effectiveness of and improving co-operative academic programs (see, e.g., [12, 18, 24]).

Our Approach. Much of the prior work on co-operative education uses data obtained by surveying students or employers. Since surveys tend to suffer from low response rates, datasets used in prior work contain on the order of 100 datapoints or fewer. We observe that a co-op process at a large university generates a large amount of data that can be collected and analyzed: textual data such as job descriptions, relational data denoting which student applied to/interviewed with which employer, and numeric data such as workplace evaluations. Based on this observation, we have recently collected these datasets and initiated a new research direction on data-driven analysis of co-operative education. In this paper, we provide an overview of our research agenda, the datasets and methodologies we have used, our results so far, and directions for future work. We believe that co-operative education is an important new application area that showcases the power of data analytics and data-driven decision making.

We classify our research so far into the following four topics.

(1) Job analysis: we perform text mining on job advertisements to understand what types of co-op jobs are available and what skills employers are looking for.

(2) Competition analysis: we represent students and employers as graphs, with edges between students who interviewed with the same employer and edges with employers who interviewed the same students. This allows us to find densely-connected subgraphs of jobs and students who compete with each other.

(3) Satisfaction analysis: we analyze employers’ evaluations of students’ workterm performance and students’ evaluations of employers to determine whether the participating parties are satisfied with each other. Furthermore, since employers rate students on multiple criteria such as productivity, communication and leadership, we can identify what co-op students are good at and what areas need improvement.

(4) Entrepreneurship analysis: we identify co-op jobs created by local startup companies to quantify the effect of entrepreneurship on the co-op market.

Roadmap. The remainder of this paper is organized as follows. Section 2 gives an overview of the co-operative education process and the datasets used in our research. Sections 3 through 6 discuss the four research topics mentioned above. For each topic, we present the motivation, followed by our data-driven methodology and the resulting business insights for students, employers and the institution. Section 7 concludes the paper with directions for future work.

2 PROCESS & DATA OVERVIEW
In traditional post-secondary programs, an academic year is divided into two or three semesters, and students spend some or all semesters on-campus taking classes. In co-operative (co-op) programs, students alternate between on-campus study terms and off-campus work terms, with each work term possibly taking place at a different employer. Thus, in any one semester, some students may be taking classes on campus whereas others may be away on work terms. In order to graduate with a co-op degree, students must take the required number of courses and also be away on work terms. In any one semester, some students may be taking classes on campus whereas others may be away on work terms. In order to graduate with a co-op degree, students must take the required number of courses and also complete a required number of work terms (e.g., at least three or five). Work terms may be one or two semesters long.

In a typical post-secondary institution, the undergraduate co-op process takes place every semester for students currently on campus who are seeking a co-op job in the upcoming semester. At the beginning of a semester, employers post job advertisements. Students apply to jobs by uploading their resumes and grade transcripts, and employers interview selected candidates. Finally, hiring decisions are made before the end of the current semester. Then, at the end of the work term (next semester), students and employers evaluate each other.
We have collected ten years of co-op data from a large North American university, having the following schema:

- Student data: student id, academic program
- Employer data: employer id, employer name
- Job data: job id, employer id, semester, location, job title, job advertisement text, salary
- Interview data: student id, job id, academic year of the student at the time of the interview, a binary attribute denoting whether or not the student obtained the job
- Employer evaluations of students: job id, overall numeric evaluation, numeric evaluations on various criteria (communication, problem solving, initiative, etc.)
- Student evaluations of employers: job id, overall numeric evaluation

Our dataset spans from 2006 till 2015 and contains over 138,000 job advertisements, over 37,000 students and over 12,000 employers.

Real datasets usually contain errors and inconsistencies. In our case, the salary field was problematic. Some job postings did not include a salary, perhaps because the salary was negotiable. Some jobs included what appeared to be hourly salary, whereas others specified larger numbers which appeared to be monthly or whole-semester salaries.

In the remainder of the paper, we discuss our analysis of jobs (Sections 3 and 6), interviews (Section 4) and evaluations (Sections 5 and 6).

3 JOB ANALYSIS

3.1 Motivation

We begin with an analysis of job advertisements. We observe that job descriptions are a rich source of information about desired skills, company culture and working environments. Thus, our goal is to extract informative terms from job descriptions: technical skills, soft skills, perks (e.g., free food or proximity to public transit) and other terms indicating the nature of the job. We aim to understand employers’ talent needs and to let students know what types of co-op jobs are available to them. We only use the most recent data (from 2015) for this analysis.

3.2 Methodology

Figure 1 shows an anonymized example of a job description. It includes the following useful information:

- Technical skills: Javascript, Ruby on Rails
- Soft skills: team player, ability to learn
- Job duties: architecting and implementing UI designs
- Desired mindset and attitude: obsessed with technology
- Perks: ping-pong and foosball table, free lunch
- Company culture: casual environment

However, job descriptions are not standardized or well structured, and include administrative and formatting elements such as URLs, contact emails, timestamps, and of course common English words. Our technical challenge, therefore, is to extract useful information from job descriptions.

We address this challenge by designing a parser that extracts job-related attributes from unstructured job descriptions. To remove unnecessary words, we build a vocabulary, call it List A, consisting of publicly available lists of common English words:

$$\text{http://www.lexicat.ca/freq/lists_download/longman_3000_list.pdf}$$

3.3 Insights

Below, we give two examples of insight that can be obtained by comparing groups of job descriptions; see [4] for full analysis.

First, we compare jobs obtained by Information Technology (IT) students with those obtained by Finance students. The word clouds with frequently occurring terms in IT and Finance jobs are shown in Figure 2. Soft skills are highlighted in green. We note that soft skills such as communication, teamwork and learning are frequent in both types of jobs; this emphasizes the importance of soft skills in post-secondary curricula. However, hard skills are different: IT jobs mention C++ and Java whereas Finance jobs are more likely to mention MS Excel and accounting. Upon closer inspection, we found that the top five sought-after programming languages in IT jobs are Java (mentioned in 33 percent of job postings), C++ (33 percent), JavaScript (31 percent), C (24 percent) and Python (22 percent). We also found interesting differences


$$\text{https://media.gcflearnfree.org/casestudies/modules/48/common_abbr.png}$$

$$\text{https://www.thebalance.com/list-of-the-best-skills-for-resumes-2062422}$$

$$\text{http://nec.xde.ge.ca/English/nec/welcome.aspx?ver=1.16}$$

Figure 1: A sample job description.
in the descriptions of mindsets and work environments: IT jobs are more likely to mention passion, creativity and love (of technology) whereas Finance jobs mention client relationships and interpersonal skills.

Next, we show two Venn diagrams in Figure 3, which characterize the overlap between junior jobs (obtained by lower-year students in years 1 and 2) and senior jobs (obtained by upper-year students in years 3 and 4). Again, IT is on the left and Finance is on the right. All IT and Finance jobs require soft skills such as communication and collaboration. However, junior IT jobs require scripting and HTML whereas senior IT jobs mention advanced technologies: distributed and scalable systems and security. Furthermore, common terms in junior Finance jobs include file, arrange, update and MS Office, which suggests clerical and data entry positions. On the other hand, senior Finance jobs are more likely to mention risk managing, statistics, modelling and investing. These results can help manage the expectations of junior students: it may take until senior years to obtain a co-op position that leverages advanced skills and technologies.

4 COMPETITION ANALYSIS

4.1 Motivation
The previous section discussed job description mining to understand what skills employers are looking for. After advertising jobs, the next step in the co-op process is to select candidates for interviews. In this section, we analyze interview data to determine which groups of students and employers compete with each other. Characterizing the extent of competition is an important business problem. For example, employers may not have a good understanding of the available talent pool and may not be allocating their recruiting resources effectively. Likewise, students may not be aware of the extent of competition for various types of jobs and therefore they may not know which jobs are realistically within their reach. Again, we only use the most recent data for this analysis.

4.2 Methodology
We use a graph mining methodology to characterize competition. We construct two graphs from interview data: a student graph, in which two students are connected if they interview for at
least one job in common, and a job graphs, in which two jobs are connected if they interview at least one student in common. Next, we run community detection on both graphs using the Louvain Method. The goal of community detection is to cluster the nodes in a graph such that nodes belonging to the same cluster/community are strongly connected while nodes in different communities are sparsely connected [3].

We illustrate our methodology with a simple example in Table 1, drawn from [21], which describes interviews of nine students (labelled 1-9) for eight jobs (labelled A-H). Figure 4 shows the corresponding student and job graphs. The job graph contains two communities, coloured blue and red. We can then colour the communities in the student graph based on the job communities in which the students had the most interviews. For example, student community 1, containing students 1–5, is blue because these students interviewed for jobs in job community 1 which is also blue.

In addition to community detection, we identify nodes with high closeness centrality, i.e., nodes with the smallest average shortest path length to other nodes. These nodes (jobs) are interesting as they are likely to be multi-disciplinary positions that interview a diverse set of students and compete with a diverse set of other jobs for these students.

4.3 Insights

Below, we describe selected results on the competition in the Information Technology sector; see [21] for full details and see [14] for a graph-mining study on the competition for co-op jobs among academic programs.

Table 1: Example table of interviews

<table>
<thead>
<tr>
<th>Job ID</th>
<th>Student IDs</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1, 2</td>
</tr>
<tr>
<td>B</td>
<td>1, 2</td>
</tr>
<tr>
<td>C</td>
<td>2</td>
</tr>
<tr>
<td>D</td>
<td>1, 3, 4, 5</td>
</tr>
<tr>
<td>E</td>
<td>5, 6</td>
</tr>
<tr>
<td>F</td>
<td>6, 7, 8, 9</td>
</tr>
<tr>
<td>G</td>
<td>7, 8, 9</td>
</tr>
<tr>
<td>H</td>
<td>7</td>
</tr>
</tbody>
</table>

The Louvain Method found eight clusters in our job graphs, three of which contained mostly IT jobs. Upon further inspection, we established a clear ranking of these three communities:

- The first community contained sought-after IT jobs at top companies such as Facebook and Google. Most of the students who interviewed for these jobs were senior (in their third or fourth years of study).
- The second community contained small IT companies and start-ups which mostly interviewed and hired junior students (in their second year of study).
- The third community had mostly quality assurance and software testing jobs, which are perceived by students as less desirable work. Most students competing for these jobs were in their first year of study and had little prior work experience.

When analyzing competition, we found that some small IT companies and start-ups from the second community interviewed the same students as top-tier companies from the first community. However, a majority of these top students accepted positions from top IT companies, and the smaller companies ended up hiring more junior students. We conclude that the smaller companies that are able to attract significant student attention are underestimating their competition and have difficulties competing for top co-op talent.

Interestingly, our centrality analysis revealed that the most central job in the top-tier IT community was a data scientist position, suggesting that data science roles are more multi-disciplinary than traditional IT positions.

5 SATISFACTION ANALYSIS

5.1 Motivation

Having analyzed what employers are looking for and which groups of employers (and students) compete with each other, we now turn to analyzing work term evaluations to understand whether students and employers are satisfied with each other. Additionally, analyzing evaluation sub-categories suggests what students are good at and what areas need improvement (as perceived by their co-op employers). This analysis uses the most recent three years of data and only includes Engineering students (the largest co-op population at the university).
Table 2: Average scores of the 19 sub-categories of performance evaluations of co-op students, in descending order

<table>
<thead>
<tr>
<th>Category</th>
<th>Average score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response to supervision</td>
<td>3.65</td>
</tr>
<tr>
<td>Ability to learn</td>
<td>3.59</td>
</tr>
<tr>
<td>Interpersonal behaviour</td>
<td>3.54</td>
</tr>
<tr>
<td>Dependability</td>
<td>3.53</td>
</tr>
<tr>
<td>Adapting to org. rules &amp; policies</td>
<td>3.52</td>
</tr>
<tr>
<td>Handling conflicts</td>
<td>3.51</td>
</tr>
<tr>
<td>Interest in work</td>
<td>3.45</td>
</tr>
<tr>
<td>Quality of work</td>
<td>3.42</td>
</tr>
<tr>
<td>Quantity of work</td>
<td>3.40</td>
</tr>
<tr>
<td>Integration of prior learning</td>
<td>3.40</td>
</tr>
<tr>
<td>Goal setting</td>
<td>3.31</td>
</tr>
<tr>
<td>Initiative</td>
<td>3.30</td>
</tr>
<tr>
<td>Verbal communication</td>
<td>3.26</td>
</tr>
<tr>
<td>Judgement</td>
<td>3.23</td>
</tr>
<tr>
<td>Written communication</td>
<td>3.22</td>
</tr>
<tr>
<td>Problem solving</td>
<td>3.21</td>
</tr>
<tr>
<td>Planning &amp; organization</td>
<td>3.13</td>
</tr>
<tr>
<td>Creativity</td>
<td>3.01</td>
</tr>
<tr>
<td>Leadership</td>
<td>2.92</td>
</tr>
</tbody>
</table>

5.2 Methodology

The methodology for satisfaction analysis is simple: we compute average evaluation scores for different groups of students and point out statistically significant differences. We also pay attention to the fraction of Not Applicable (N/A) scores as employers have the option to enter N/A for any category that was not applicable to a particular work term.

5.3 Insights

We start with students’ evaluations of their employers (on a scale from one to ten; higher is better). We found that Engineering students gave their employers an average score of 7.55. This suggests that students are generally satisfied with their co-op experience. Interestingly, students tend to rate their first employers higher than subsequent employers, perhaps because their first co-op expectations are lower.

Next, we discuss workplace supervisor evaluations of students. Students receive an overall score from one to five corresponding to: unsatisfactory, satisfactory, good, very good and excellent. We found that Engineering students obtained an average score of 3.74, i.e., between very good and excellent. Senior students consistently obtained higher scores than junior students, and furthermore, senior students were more likely to take a job abroad and be satisfied with it.

Additionally, students are rated on 19 criteria, with each rating being from one to four (higher is better). Table 2 shows the average score for each of the 19 criteria, from highest to lowest. Students tend to excel at Response to supervision and Ability to learn, but are not rated highly on Creativity and Leadership. We speculate that it may be difficult to display leadership and creativity in limited-term co-op positions with well-defined tasks. Students may be focused on completing their tasks before the end of their work term rather than trying out new approaches.

The two criteria with the most N/A scores were Conflict management and Leadership (in fact, nearly half the ratings were N/A). However, the percentage of N/A ratings for Integration of prior learning, Goal setting, Leadership and Written communication decreases from first year through fourth year. This suggests that senior students enjoy more opportunities for leadership and independence. On the other hand, the percentage of N/A ratings for other criteria does not change significantly over time.

Interestingly, we found that the average Problem Solving scores improved the most from first year to final year: they increased from 3.07 to 3.23, which is statistically significant at the 95 percent confidence level.

6 ENTREPRENEURSHIP ANALYSIS

6.1 Motivation

Entrepreneurship can lead to job creation and economic growth. As a result, there has been private and public emphasis on fostering entrepreneurship: examples include tax credits and establishing supporting entities such as startup incubators which are often paired with universities. Furthermore, there is evidence that innovative universities can contribute to growth in the regional economies. Thus, it is natural to ask how entrepreneurship impacts the co-operative process. In this section, we give an overview of our study of the impact of entrepreneurship on co-operative education and job creation [1].

6.2 Methodology

For this analysis, we combine the co-op dataset described earlier with a list of 472 companies started by 746 of the institution’s current or former Engineering students and faculty members. To integrate these two datasets, we matched company and founder names in the startup dataset with employer and student names in the co-op dataset. To deal with alternative name spellings (e.g., “XYZ Inc.” vs. “XYZ Systems Inc.” or “Jim Smith” vs. “James A. Smith”), we identified potential matches using approximate string matching and verified correct matches using publicly available data such as LinkedIn profiles. At the end of this process, we identified:

(1) Co-op placements at the institution’s startup companies, including salaries and students’ and employers’ evaluations

(2) Students in the co-op dataset who at some point were enrolled in a co-op program at the institution and went on to start a company (we refer to these students as future founders)

We then summed up the salaries at the aforementioned co-op placements to quantify the economic impact of entrepreneurship on the institution’s co-op system. For placements with no salary data, we imputed the missing salary with the mean salary across all startup companies.

Since the startup dataset may not be complete, our results should be interpreted as lower bounds on the true number of (and salaries paid by) the institution’s companies. Furthermore, we only consider co-op placements for the institution’s own students, not the total number of jobs created by the institution’s companies.

6.3 Insights

We start with our economic impact analysis. We found that over the past ten years, nearly half (223 of the 472) known companies started by the institution’s Engineering students and professors have participated in the institution’s cooperative process. These

223 companies hired over 5,800 distinct students from the institution, which is 15 percent of all students, for a total of over 9,000 co-op placements, which is 6.5 percent of all placements. We estimate that the salaries paid at these placements add up to over $116 million. The institution can view this as data-supported evidence of the economic impact of the entrepreneurship of its members on co-operative education. Furthermore, these results can be used by institutions to motivate programs and initiatives that encourage entrepreneurship.

We then examined the employer and student evaluations corresponding to placements at these 223 companies. We found that both are statistically significantly higher compared to those at other placements.

Finally, we analyzed the co-op histories of future founders (i.e., students who went on to start companies). We located 221 of the 746 founders in the co-op dataset (the others are faculty or staff members, or students who were not enrolled in a co-op program within the past ten years). Only five percent of these 221 founders are female; in future work, we want to understand why this is the case and to determine if the trend is improving. Notably, future founders were more likely to give and receive higher work term evaluations compared to other students. In particular, future founders were rated more highly than other students for their soft skills such as Initiative, Creativity and Communication (recall Table 2). This suggests a possible link between success in co-operative education and entrepreneurship.

7 CONCLUSIONS

In this paper, we presented a new application area for data analytics: improving co-operative education. We explained the datasets that arise in the co-op process, ranging from textual job advertisements to interview relationships and numeric performance evaluations. We then outlined the data-intensive methodologies that may be applied to produce actionable insight for students, employers and institutions. The methodologies included text mining, graph mining and integrating multiple data sources through approximate string matching.

Our research so far has led to new data-driven insight, but there is more that can be done. Below, we list several potential directions for future work on analyzing co-operative data, possibly combined with other datasets:

- Analysis of co-op and post graduation data: Does co-op employment with a given employer lead to full-time employment with the same employer after graduation? This requires correlating co-op data with postgraduate employment data, which could be obtained, e.g., from LinkedIn profiles.

- Analysis of co-op and secondary school data: Does secondary school work/extracurricular experience help students obtain post-secondary co-op jobs? This requires correlating co-op data with undergraduate admission records.

- Combining competition analysis with satisfaction analysis: do top-tier jobs receive higher evaluations by students?

- Gender equity in co-operative education: Are female students in traditionally male-dominated academic programs such as Computer Science satisfied with the co-op experience?

- Trend analysis: Have sought-after skills changed over time? Have evaluation scores (of students and of employers) changed over time?

- Employer/Employee recommender systems: Can text matching or graph mining techniques such as link prediction be used to recommend potential students to potential employers?

REFERENCES


