#### Towards an Automatic Approach for Assessing Program Competencies

Xinyuan Chang, Bingxin Wang, Bowen Hui

Computer Science University of British Columbia Okanagan

### Motivation

- Skills analysis in economics, education, policy making
  - Understanding economic needs
  - Observing and forecasting skill trends
  - Aligning industry needs and training
  - Developing re-skilling programs
  - Ministries push towards competency based education

### Motivation

- Skills analysis in economics, education, policy making
  - Understanding economic needs
  - Observing and forecasting skill trends
  - Aligning industry needs and training
  - Developing re-skilling programs
  - Ministries push towards competency based education
- Recent interest from computer science
  - Fast-changing skills
  - Role in developing scalable solutions

### Motivation

- Skills analysis in economics, education, policy making
  - Understanding economic needs
  - Observing and forecasting skill trends
  - Aligning industry needs and training
  - Developing re-skilling programs
  - Ministries push towards competency based education
- Recent interest from computer science
  - Fast-changing skills
  - Role in developing scalable solutions
- Our goal: To develop scalable approach for evaluating program competencies
  - Develop automatic skills extraction system
  - Identify skill gaps in a CS program

### **Related Work**

- Definition of "skill" [Green 2013; Payne 2017; Duckworth & Yaeger 2015]
  - Cognitive and non-cognitive abilities
  - Emotional abilities
  - Discipline-specific behaviors and workplace contexts

### **Related Work**

- Definition of "skill" [Green 2013; Payne 2017; Duckworth & Yaeger 2015]
  - Cognitive and non-cognitive abilities
  - Emotional abilities
  - Discipline-specific behaviors and workplace contexts
- Skill assessments [Rios et al. 2020]
  - Self-reported questionnaires, employer surveys
  - Teacher/observer reports
  - Performance tasks, job analysis data
  - Expert theoretical synthesis

### **Related Work**

- Definition of "skill" [Green 2013; Payne 2017; Duckworth & Yaeger 2015]
  - Cognitive and non-cognitive abilities
  - Emotional abilities
  - Discipline-specific behaviors and workplace contexts
- Skill assessments [Rios et al. 2020]
  - Self-reported questionnaires, employer surveys
  - Teacher/observer reports
  - Performance tasks, job analysis data
  - Expert theoretical synthesis
- Major approaches in skills analysis with labor market data (cf. paper)
  - Content analysis coded by domain experts
  - Automatic approach using keyword matching
  - Domain and language knowledge via external resources
  - Machine learning algorithms for generalizability

### Job Skills Landscape

 Occupational classification systems built by government organizations (e.g., O\*NET, [Peterson et al. 1995])

### Job Skills Landscape

- Occupational classification systems built by government organizations (e.g., O\*NET, [Peterson et al. 1995])
- Aligning supply of skills with demand for employment [McGuinness et al. 2018]

### Job Skills Landscape

- Occupational classification systems built by government organizations (e.g., O\*NET, [Peterson et al. 1995])
- Aligning supply of skills with demand for employment [McGuinness et al. 2018]
- Changing skills in IT
  - Across time [Todd et al. 1995; Smith & Ali 2014]
  - Across career stages [Kappelman et al. 2016]
  - Across company size [Nelson et al. 2007]

### Skill Gaps in CS Education

• Changes introduced by Industry 4.0 and transition to new technologies, data analysis, design/research skills [da Motta Reis et al. 2020;Pinzone et al. 2017]

### Skill Gaps in CS Education

- Changes introduced by Industry 4.0 and transition to new technologies, data analysis, design/research skills [da Motta Reis et al. 2020;Pinzone et al. 2017]
- Increase in digital divide may dampen participation by females [West et al. 2019;Papyrina et al. 2021]

### Skill Gaps in CS Education

- Changes introduced by Industry 4.0 and transition to new technologies, data analysis, design/research skills [da Motta Reis et al. 2020;Pinzone et al. 2017]
- Increase in digital divide may dampen participation by females [West et al. 2019;Papyrina et al. 2021]
- Impact within on curriculum [Patacsil & Tablatin 2017; Radermacher et al. 2014;Börner et al. 2018; Restuccia 2019]
  - Gap in student skills and industry needs
  - CS: Prioritization in communication, testing, project experience, problem-solving in practical settings, specialized tools
  - Börner et al. used keyword matching to extract longest matching skill from publication abstracts, course syllabi, job postings

## **ML** Techniques in Skills Analysis

• Beyond keywords/strict syntax

# **ML** Techniques in Skills Analysis

- Beyond keywords/strict syntax
- Automatic extraction of skills
  - Specialized track in TREC (2005+) focus on expert search
  - LinkedIn [Skomoroch et al. 2012]: deep NLP analyses, clustering, crowdsourcing
  - SKILL by CareerBuilder [Hoang et al 2015;Javed et al. 2017;Zhao et al. 2015]: skills taxonomy generation, deep NLP analyses, Wikipedia categories
  - ScholarLens [Sateli et al. 2017]: extract competencies from publications to create researcher profiles
  - Others [Bernabé-Moreno 2019; González-Eras & Aguilar 2019]: linguistic phrase structure, additional resources, clustering

# **ML** Techniques in Skills Analysis

- Beyond keywords/strict syntax
- Automatic extraction of skills
  - Specialized track in TREC (2005+) focus on expert search
  - LinkedIn [Skomoroch et al. 2012]: deep NLP analyses, clustering, crowdsourcing
  - SKILL by CareerBuilder [Hoang et al 2015;Javed et al. 2017;Zhao et al. 2015]: skills taxonomy generation, deep NLP analyses, Wikipedia categories
  - ScholarLens [Sateli et al. 2017]: extract competencies from publications to create researcher profiles
  - Others [Bernabé-Moreno 2019; González-Eras & Aguilar 2019]: linguistic phrase structure, additional resources, clustering
- Classification of job skills
  - Skills database to explore employee turnover [Liu et al 2018]
  - WoLMIS [Boselli et al. 2018] classified job ads to occupational codes

- No fixed keywords or strict text format requirement
- Major steps:
  - Seed phrases
  - Preprocessing
  - Linguistic patterns analysis
  - ML text classification models
  - Evaluation

- Seed phrases
  - Research database <u>edgemap.ok.ubc.ca</u> with 202 students in Digital Citizenship course self-reported 1,966 skills
  - Manually extracted skills from 1,700 job postings from indeed.com
  - Intercoder reliability (1<sup>st</sup>: 82%, 2<sup>nd</sup>: 94% agreement)
  - Result: 6,972 phrases with 4,886 skills and 2,086 non-skills
- Preprocessing
- Linguistic patterns analysis
- ML text classification models
- Evaluation

- Seed phrases
- Preprocessing
  - Standard NLP steps to process job description sections
  - Removed HTML tags, tokenization by punctuation and conjunctions, removed stop words, lemmatization, POS tagging
- Linguistic patterns analysis
- ML text classification models
- Evaluation

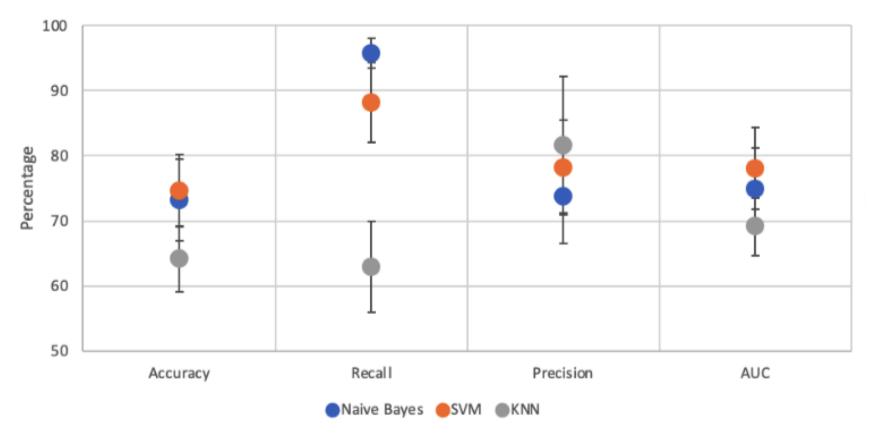
- Seed phrases
- Preprocessing
- Linguistic patterns analysis

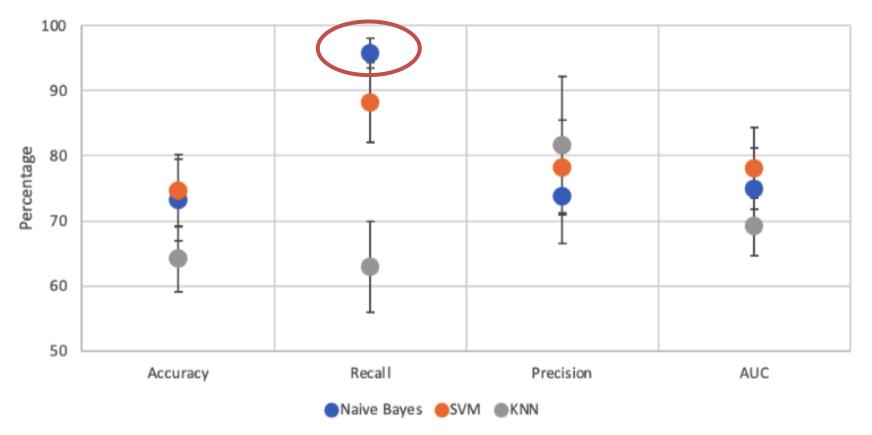
Linguistic Pattern	Examples
Noun phrase	Java, Microsoft Word, time management, strong programming skills
Verb/Gerund + Noun phrase	programming websites, design user interfaces
Noun phrase + Gerund	server hosting, software testing

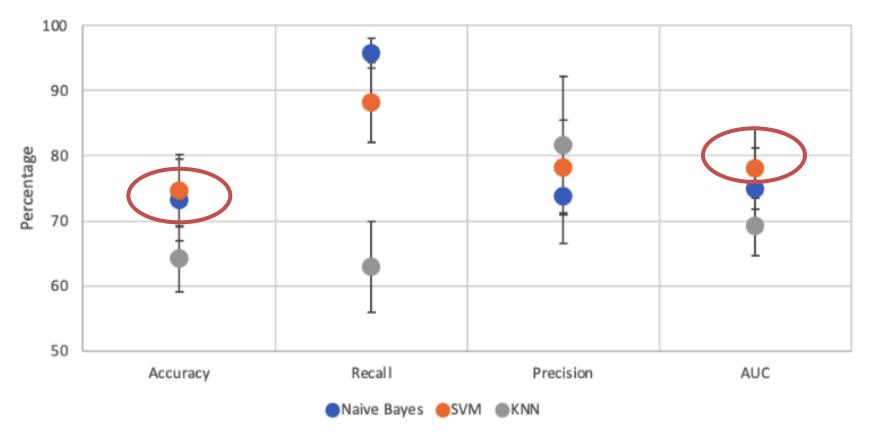
- ML text classification models
- Evaluation

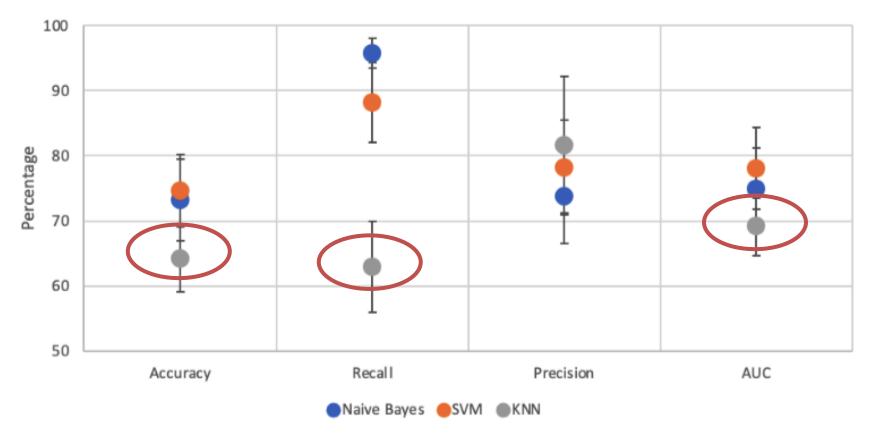
- Seed phrases
- Preprocessing
- Linguistic patterns analysis
- ML text classification models
  - Used seed phrases as a labeled dataset for text classification
  - Extracted all linguistic patterns from job postings
  - Text classification models: Naïve Bayes, SVM linear, k-NN
- Evaluation

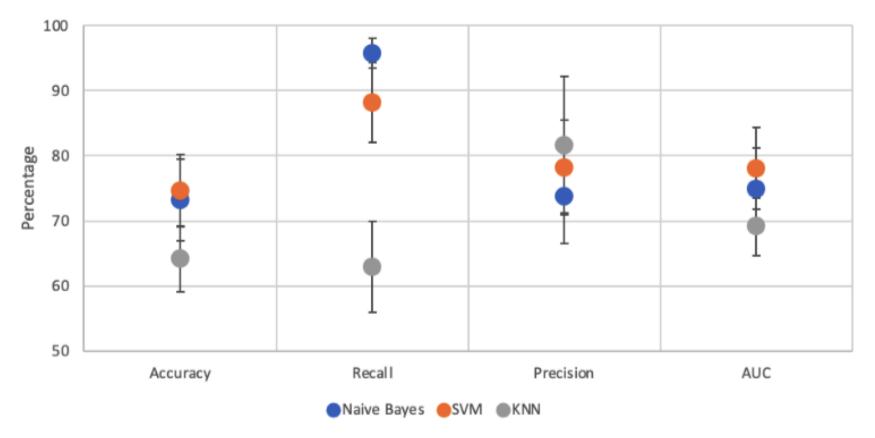
- Seed phrases
- Preprocessing
- Linguistic patterns analysis
- ML text classification models
- Evaluation
  - 10-fold cross validation
  - Measured accuracy, recall, precision, AUC











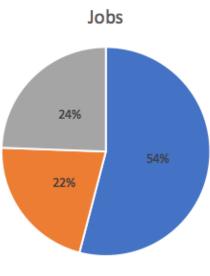
#### Analysis of Student Skills, University Courses, and Jobs

- Cross-sector perspective across 20 NAICS sectors
- CS discipline case study

#### Analysis of Student Skills, University Courses, and Jobs

- Cross-sector perspective across 20 NAICS sectors
- CS discipline case study
  - 202 students with 1,966 skills
  - 26 course syllabi in a CS program
  - 13,493 software developer job ads



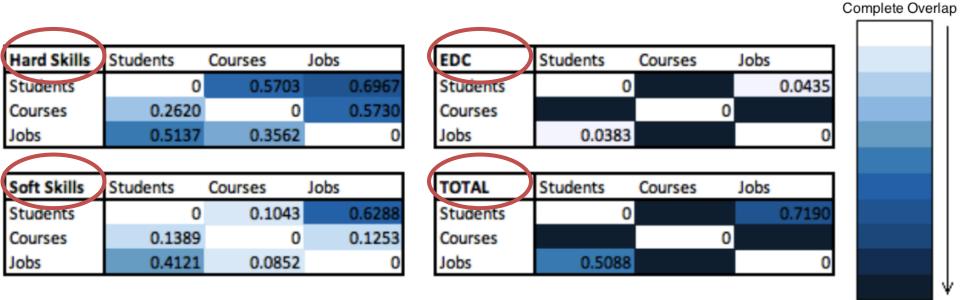


## **High-Level Findings**

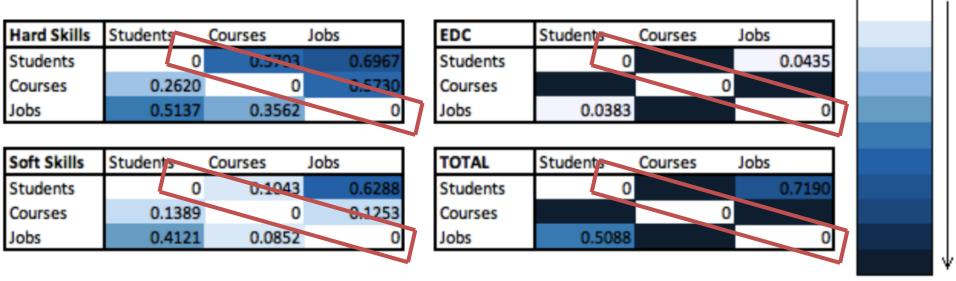
- Software developer jobs demand mostly technical skills
- CS programs should incorporate more soft skills
- Students have a high proportion of soft skills

- Questions remaining:
  - Do students have the soft skills that jobs need?
  - Are courses teaching the hard skills that jobs need?

- How much overlap is there between the skills in each category?
  - KL-divergence to measure entropy between two distributions
    - KL=0: Identical skill sets
    - Higher KL: No overlap

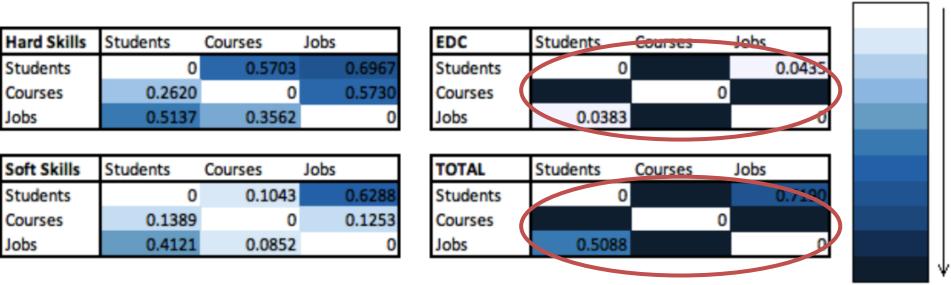


- How much overlap is there between the skills in each category?
  - KL-divergence to measure entropy between two distributions
    - KL=0: Identical skill sets
    - Higher KL: No overlap



Complete Overlap

- How much overlap is there between the skills in each category?
  - KL-divergence to measure entropy between two distributions
    - KL=0: Identical skill sets
    - Higher KL: No overlap



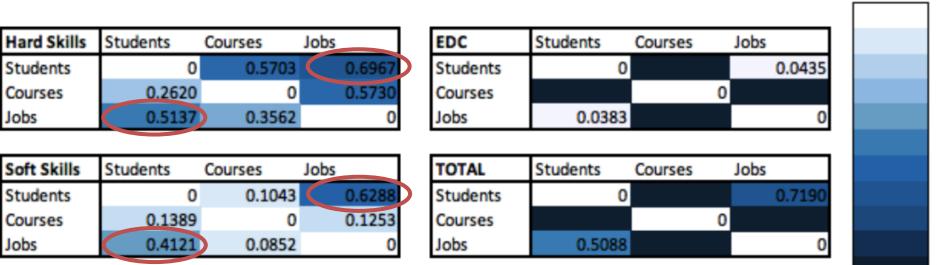
Complete Overlap

- How much overlap is there between the skills in each category?
  - KL-divergence to measure entropy between two distributions
    - KL=0: Identical skill sets
    - Higher KL: No overlap

Hard Skills	Students	Courses	Jobs	EDC	Students	Courses	Jobs
Students	0	0.5703	0.6967	Students	0		0.0435
Courses	0.2620	0	0.5730	Courses		(	D
Jobs	0.5137	0.3562	0	Jobs	0.0383		0
Jobs	0.5137	0.3562	0	Jobs	0.0383		0
Jobs Soft Skills	0.5137 Students	0.3562 Courses	0 Jobs	Jobs TOTAL	0.0383 Students	Courses	0 Jobs
			Jobs				
Soft Skills	Students	Courses 0.1043	Jobs	TOTAL		Courses	Jobs

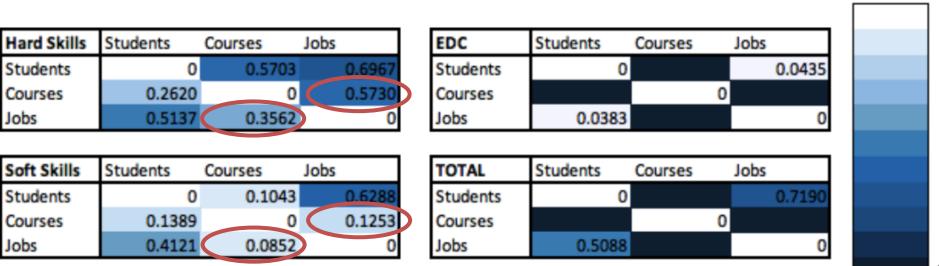
Complete Overlap

- How much overlap is there between the skills in each category?
  - KL-divergence to measure entropy between two distributions
    - KL=0: Identical skill sets
    - Higher KL: No overlap



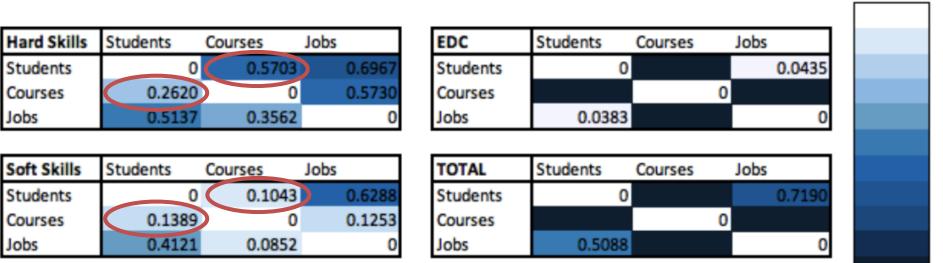
Complete Overlap

- How much overlap is there between the skills in each category?
  - KL-divergence to measure entropy between two distributions
    - KL=0: Identical skill sets
    - Higher KL: No overlap



Complete Overlap

- How much overlap is there between the skills in each category?
  - KL-divergence to measure entropy between two distributions
    - KL=0: Identical skill sets
    - Higher KL: No overlap



Complete Overlap

#### A Visual Comparison

ability able address aderant advanced agile algorithms analysis analyze animation app applications art assembly **Dasic** building class cms coding collect computer courses css cat data database design determine development editing efficient experience fluent functions game gas git hardware hosting html identify information installing internet ip **Java** javascript knowledge language machine matlab mobile model modify mysgl network object oil operate oracle order oriented packages parts pc php prezi proficient programming python requirements research resident router sampling scientific security server simple sketchup skills software sources sql statistical surveys system technical testing tools trends understanding unity variety video voip Web website xml year

algorithms analysis analyze applications apply approaches approximation architecture assembly basic code compare computer data database define demonstrate describe design development different differential direct discrete equations error evaluate experience explain exposure factorization fundamental gain game graph Idea identify imaging implementation Intelligence Interaction introduction issues java key language limitations line linear list main management manipulation medical memory methods models network newton numerical object optimization organization parallel perform planning point practice problems processes programming project proof real search select simulation software solution solving systems standard structures study techniques terminology test theory tree trust understand unit used variable various vision VS work write

access agile analysis applications code cleaning automated architecture computer control creating data databases design development distributed documentation engineering environment excellent existing frameworks health healthcare highly information javascript knowledge languages learning legacy machine methodologies multiple object-oriented occasional patterns operating procedures programming products **Drojects** python quality relational reports research reviews science sensitive services **SOTTWARE** solutions standards structures systems technical technologies testing tools versioning webwebsite winning workflows

(a) Students

(b) Courses

(c) Jobs

### **Conclusions and Future Work**

- Summary contributions and findings
  - Auto-extract skills to analyze student skills, course syllabi, job postings
  - Within-sector analysis on programming jobs and training provided curricular insights on CS program
- Limitations
- Next steps

### **Conclusions and Future Work**

- Summary contributions and findings
- Limitations
  - Job postings is a secondary data source
  - Does not use external resource
  - Student skills database contained more diverse content than desired
  - Syllabi content is a proxy to course competencies
- Next steps

### **Conclusions and Future Work**

- Summary contributions and findings
- Limitations
- Next steps
  - Extend labeled dataset and retrain classifier
  - Explore other text classification models
  - Additional within-sector analyses