The Career CHARTING App

Beverly P. Woolf and Andrew S. Lan, University of Massachusetts Amherst
01/07/2019

We propose the career CHARTING (Career Help via Analysis, Recommendation, and TrainING) application for workforce analysis and reskilling. The CHARTING application will i) help workers to identify and self-select new jobs and even career paths according to their skill sets and interests and ii) provide guidance on how they can prepare for new jobs that require new skills. CHARTING will analyze a worker's skills, strengths and preferences, propose future careers, suggest growth paths, and provide complementary training programs for workforce retraining. Initially, the application will be made available to personnel at NSF and then to workers at NIH. These workers will chart the path of their own career and dedicate time each day to build skills needed for their future career. Eventually, we will make the application available to anyone, at any time, and make it accessible from any device, in any location, and from any economic or educational background. Initially, it will train workers in the fields of data science and cybersecurity, from among workers who have been identified as being successful in computational thinking.

CHARTING consists of three phases: i) skill analysis, ii) career path recommendation, and iii) a workforce retraining program that cut across agencies and other stakeholders.

Phase I: Skill Analysis

Phase I involves skill analysis and predictive models that suggest whether workers would be likely to succeed in careers that require *computational thinking, language arts, management* skills, and other skills. This is based on a system that profiles a worker by evaluating their CV and analyzing extensive clickstream data gathered during a brief diagnostic instrument. This environment uses interactive simulations to observe workers' problem-solving processes and assess their skills. This instrument is NOT a standard assessment tool and only contains a set of worksheets or questions/answers. In contrast, it invites workers to demonstrate their thinking and reasoning processes in an interactive environment while dealing with a real-life task. Our key innovation is to use clickstream behavior data to assess a user's skill level and predict a worker's future success; this innovation is motivated by our previous findings in the ASSISTments data mining competition [Heffernan et al., 2018], which suggests that students' future career choice (STEM/non-STEM) can be predicted by their behavior at an early age, when they learn within a computerized learning platform.

<u>Computational Thinking</u>. The *computational thinking* component of the diagnostic instrument tracks the workers' behavior while solving problems using variables such as "time taken to enter first attempt", "number of attempts", "correctness," and "hint requested." The workers' usage is tracked along with information about the skills being tested, including pre-requisite, co-requisite and post-requisite skills revealing the cognitive skills associated with each problem.

<u>Language Arts</u>. The *language arts* component of the diagnostic instrument identifies workers who are strong in phonemic awareness, literacy, reading and writing. This instrument measures workers' comprehension and capacity for use of written and oral language. It is designed to further enhance workers' literacy (this includes speaking, and sometimes, even listening skills), building critical thinking skills through reading comprehension, making

inferences, asking questions, summarizing, comparing and contrasting, analyzing characters, and identifying cause and effect. Learning multiple meanings, synonyms, antonyms, prefixes, suffixes, parts of speech, and using context clues help workers broaden their oral expression, writing, and speaking skills.

Management Skills. The *management skills* component of the diagnostic instrument records how workers analyze raw business cases. In contrast to traditional cases, which digest components of an event, distill it down to a few pages and record a correct answer, raw cases provide information in a variety of ways – reports, articles, interviews, videos, photographs, original documents, and links to other websites. Raw cases mimic the real world, where information is scattered and sometimes contradictory. Part of the challenge for workers comes from sorting through raw information to come to reasonable conclusions. Variables in the log data of cases track the worker's behavior such as "navigation", "number of variables viewed", "additional resources viewed," and "videos accessed." Raw cases require workers to provide a multi-dimensional approach to the analysis of data. The instrument examines workers' ability to explore data, make decisions, coordinate facts, recognize pitfalls suggested in the case, and draw conclusions. Each clickstream record contains information about the workers' behavior within the environment.

Soft skills. The soft skills component of the diagnostic instrument captures worker's other skills, including communication skills, their ability to work in a team, etc. Using natural language processing (NLP) tools, CHARTING will analyze conversation logs between a worker and their supervisors, advisees, and co-workers in order to estimate how effective they are in terms of conveying their ideas to others and understanding others' ideas. CHARTING will also study workers' records in collaborative projects and supervisor and/or peer evaluations to evaluate how effective they are when working in a team and what is the most important component they bring to a team. These "soft" skills are essential components for certain jobs and may be key factors in a worker's career success.

CHARTING will estimate the expected knowledge state of workers with respect to computational, literacy, and management skills using a dynamic model (e.g., deep knowledge tracing). In addition to estimating a worker's skill level (which is always evolving) over time, CHARTING will also analyze the worker's ability to learn skills of certain types. This component is of crucial importance when a worker wants to make a dramatic career change: using traditional models that only estimates a worker's skill level, one will not be able to generalize and predict how well the worker will develop new skills required for a different type of job. As a concrete example, for a sales representative who do not know programming but have demonstrated good mathematical skills in their early years and can learn key computational and reasoning concepts quickly, there is a good chance that they will become a good candidate for a programmer job. This component requires a new regime in psychometrics: in addition to traditional models that only use workers' skill levels [Lan et al., 2014], we need new models with two levels of latent variables: one level for skills, which dictates the worker's performance, and the other level for learning abilities, which dictates the worker's ability to develop new skills over time.

We have expertise in building tools for skill analysis and predictive models. For example, experiments show that a student's mastery level, or so-called the knowledge state, of mathematics has a potential power to distinguish between choosing STEM and non-STEM careers.

The ASSISTments Data Mining competition [Heffernan et al., 2018] developed models that predicted whether a student's first job out of college would be a STEM/non-STEM field [Yeung & Yeung, 2019]. Deep knowledge tracing estimated a student's knowledge state on different cognitive skills in mathematics by exploiting the extensive clickstream data collected from the blended learning platform used by students a decade earlier when they were in middle school. This analysis discovered important factors that potentially influenced a student's first job in a STEM/non-STEM field. The expected knowledge state was combined with the student profile to train a predictive model that classified whether a student will choose a STEM field in their first job. We also have expertise in developing new machine learning methods for student outcome prediction, e.g., approximate Kalman filter [Lan et al., 2014], and for behavior data analysis [Chen et al., 2018].

Phase 2: Career Path Recommendation

In addition to skill analysis, another important component in the CHARTING application is to analyze the job market landscape and make recommendations of future career paths to each worker. Due to the rapidly development of science and technology, the career landscape in science, technology, engineering, and mathematics (STEM) is constantly evolving: there are many new jobs that require skill sets traditional education programs do not provide. Therefore, analyzing what skills are required for a job is as important as analyzing whether a worker possesses a skill. CHARTING will build on existing works that leverage NLP tools to analyze job postings on popular job listings websites (indeed.com, glassdoor.com) and associate key skills with job listings. Moreover, CHARTING will analyze career paths of real professionals (e.g., using data from professional networks like LinkedIn) to build profiles for typical career paths and recommend next jobs to workers based on their past and current jobs and their current skill levels [Mimno and McCallum, 2008].

More importantly, the CHARTING application will not only recommend future jobs to a worker but also identify the skills they need to develop or improve on to qualify for the future jobs. CHARTING will analyze large-scale career path data from professional networks (especially the skill tags associated to each worker) and identify the crucial skills for each job. Then, coupling with the estimated skill levels of the worker, the CHARTING application will be able to lay out a detailed, targeted, and personalized training plan for the worker and guide them through the steps to become a highly qualified candidate.

The CHARTING application combines both NLP analysis and behavior data analysis and will thus significantly improve over existing career path recommendation tools that use only NLP.

Phase 3: Workforce Retraining

Once the first two phases are complete, the CHARTING application will recommend a training program to prepare workers for new careers. These courses will have a substantial emphasis on higher-order thinking skills, will combine effective use of video with animation, and will include a number of writing projects designed to help students achieve overall career readiness. We will incorporate appropriate massive open online courses (MOOCs) and also online courses provided by top universities into the training program. We detail two such opportunities at UMass below.

For example, UMass offers training programs in Cybersecurity and Data Science. The *Cybersecurity Institute* at UMass[1] is the multi-disciplinary focal point for security research and education at UMass. It brings together dozens of internationally recognized faculties to address the critical, cross-industry need for innovative security research and well-trained cybersecurity professionals. Working with partners in government, industry, and academia, the institute seeks to advance scientific and societal understanding of pressing issues related to the field. The Cybersecurity Institute offers online certificates at both the undergraduate and graduate levels. These courses provide training in IT security and policy, covering a wide range of topics including security and privacy challenges in networking and communications, embedded systems, software engineering, software systems, applied cryptography, etc. Students do not need to matriculate (or be enrolled at) UMass. All classes are accessible to remote students and they are allowed to take classes in their own schedule within a 12-month period. Once a student completes all courses, a certificate is awarded. Example courses include: Internet Law & Policy, System Defense and Test, Information Assurance, Digital Forensics, Secured Distributed Systems. Information Risk Management, Fraud Detention.

The *Data Science* degree[2] within the Informatics Program at UMass helps undergraduates develop computational thinking skills and to analyze massive quantities of data in areas outside of computer science, for example, in economics, voting, astronomy and epidemiology. It teaches workers about alternative ways to analyze, visualize, and reason about information and helps them gain knowledge and skills that link core computing ideas, e.g., computational thinking, to other disciplines. This program includes a technical core of six courses, plus six other courses.

Summary

The Career CHARTING application will detect a worker's skills, strengths and preferences, propose future careers paths, and provide a detailed training program for workforce retraining. Workers will chart the path of their own career and dedicate time each day to both executing their own current job functions and building skills needed for their future career. It will empower workers to refresh and modernize their skills toward future work that is one of continuous change and depends on a culture of continuous, lifelong learning. CHARTING will start by focusing on NSF employees within the fields of data science and cybersecurity and eventually generalize to the entire STEM workforce in the US.

---

1 More information about the UMass Cybersecurity Institute is located at https://infosec.cs.umass.edu/?_ga=2.266986446.1541486134.1546716227-950335311.1530712153

2 More information about the UMass Data Science Program is located at https://www.cics.umass.edu/ugrad-education/informatics-data-science-track

References

N. Heffernan, R. Baker, B. Woolf, and T. Patikorn. (2018). Proc. *Workshop on Scientific Findings from the ASSISTments Longitudinal Data Competition, International Conference on Educational Data Mining (EDM)*. https://sites.google.com/view/edm-longitudinal-workshop/home

C. Yeung, and D. Yeung. (2019). Incorporating features learned by an enhanced deep knowledge tracing model for STEM/non-STEM job prediction. *International Journal of Artificial Intelligence in Education (IJAIED)*, in press.

D. Mimno, and A. McCallum. (2008). Modeling Career Path Trajectories. https://mimno.infosci.cornell.edu/papers/07-69.pdf

A. S. Lan, C. Studer, and R. G. Baraniuk. (2014). Time-Varying Learning and Content Analytics via Sparse Factor Analysis. In Proc. *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 452–461.

W. Chen, A. S. Lan, D. Cao, C. Brinton, and M. Chiang. (2018). "Behavioral Analysis at Scale: Learning Course Prerequisite Structures from Learner Clickstreams," In Proc. *International Conference on Educational Data Mining (EDM)*, pp. 66-75.

V. S. Dave, B. Zhang, M. Al Hasan, K. AlJadda, and M. Korayem. (2018). A Combined Representation Learning Approach for Better Job and Skill Recommendation. In Proc. *ACM International Conference on Information and Knowledge Management* (CIKM), pp. 1997-2005.