

Skill Evaluation

BY SHASHANK SRIKANT, ROHIT TAKHAR,
VISHAL VENUGOPAL, AND VARUN AGGARWAL

UPPWARD OF FOUR million graduates enter the labor market every year in India alone.

India boasts of a large services economy, wherein a single company hires thousands of new employees every year. Meanwhile, product companies and small and medium enterprises (SMEs) look for a few skilled people each. This requires cost-effective and scalable methods of hiring. Interviewing every applicant is not a feasible solution.

On the other hand, graduates from 30,000+ institutes of higher education spread across 20+ Indian states face a constant challenge in signaling their competence to potential employers. Companies, most of which are located in the top 20 biggest cities in the country, bias their search by relying on proxies like university name and the city a college is located in. Applying such crude filters results in meritorious



students from various demographics being ignored. Further, these students have no mechanism to get feedback on how their skills compare to those required by the industry.

Having systems that can intelligently and scalably assess a wide variety of skills is essential to addressing this broader problem affecting every modern-day

labor market. *Aspiring Minds* was formed 10 years ago to address this challenging problem. We have developed a scalable platform to deliver standardized assessments to test job skills. The platform tests more than two million students every year and is used by 5,000+ companies including 100+ Fortune 500 companies.

A particular challenge in designing scalable assessment technologies is evaluating subjective, open-ended responses to questions. Such questions directly simulate a skill or a job task within the constraints of a test-like environment and are generally more informative than multiple-choice questions (MCQs). For instance, it is almost necessary to evaluate programming or

spoken skills using such formats over MCQs. Evaluating such responses is an expensive, time-consuming process involving human graders, and suffers from standardization concerns. Automated grading has the potential to address these issues and impact millions of job seekers, trainers, and corporations.

At *Aspiring Minds*, we have, over the last decade, distilled a framework to cast the question of subjective assessments as problems in computer science, and specifically, in machine learning (ML).¹ In it, candidate responses are data points in a high dimensional space, from which we predict their true, latent, underlying score. This is a different paradigm altogether that we envision. While there

A particular challenge in designing scalable assessment technologies is evaluating subjective, open-ended responses to questions.

exist solutions and products that evaluate language skills subjectively, most solutions provided by established, international educational testing and assessment organizations focus on testing general aptitude skills and adopt traditional testing formats like MCQs.

We illustrate the broad industry verticals we have developed tools for, each highlighting a research problem it addresses and the associated innovative intervention we devised.

► **Programming and software engineering.** Automata^{6,8,9} uses ML models to automatically score computer programs on parameters such as functional correctness, complexity, and style. These models use intelligent features extracted from programs, which can signal correctness even when they fail to compile. Importantly, we designed them to be independent of the task the program solves, thus allowing to scale assessments to a wide variety of questions. There have been attempts by other research groups^{2,7} at analyzing programs solving introductory programming problems. They, however, focus on providing automated feedback. Our work differs in that we focus on grading programs on a rubric. To achieve this, we extract key data flow properties in programs that capture their *meaning* and use them as features in an ML model; the problems we model are significantly more involved than introductory problems and exist in multiple languages.

► **Customer service.** The IT-enabled services (ITeS) market in India employs four million people and is a US\$181-billion industry. Spoken English skills are central to this industry.

SVAR^{3,4} evaluates speaking skills at scale. Applicants call a phone number, have a conversation with an automated interactive system, and on hanging up, receive a score on their spoken skills such as pronunciation and fluency. It draws from speech and signal processing technologies and uses ML to predict these scores. To reduce evaluation time, and to improve model accuracy, we innovated by crowdsourcing parts of our feature extraction and model evaluation.

► **Blue-collar jobs.** Four-and-a-half million employees in India are estimated to be employed in blue-collar jobs. However, no automated means existed to assess motor skills, a key requirement in these jobs. Akin to how computers serve as a medium to test cognitive skills, we showed how touch devices can be used to assess motor skills.⁵ This requires a person to use their fingers and wrists to play specific games designed for tablet apps. We have shown their performance on these tasks to correlate with on-job performance.

► **Professional communication.** Email correspondence has become an integral part of the communication tool chain in any organization. To test professionals' email writing skills, we employ deep learning and NLP to assess various aspects like grammar, content, and structure.

To our knowledge, this is the first attempt at designing and productizing such ML-driven technologies to assess these specific skills.

► **Domain knowledge.** In consultation with subject-matter and industry experts, we have designed 300+ tests for domain knowledge across various industry verticals such as IT, ITeS, retail,

To our knowledge, this is the first attempt at designing and productizing such ML-driven technologies to assess these specific skills.

manufacturing, BFSI, hospitality, and telecom. Backed by statistical techniques such as item response theory, these tests provide standardized assessments in specific topics, helping create a level playing field for job applicants.

Over the years, we have gathered a database of applicants' performance in the various verticals discussed here. This has helped us quantify the state of employability in India, and study a year-on-year change in employability conditions. Since 2010, Aspiring Minds has released annual National Employability Reports, which have now become the gold standard for tracking the quality of higher education in India, aiding and informing policy formulation.

Besides these opportunities, we have also identified a number of challenges in using CS/ML for grading. These include issues around quality of labels (expert grades), low sample sizes, sample characteristics, standards for acceptable errors in models, among others. Several key issues are in developing models that are causal and addressing issues of fairness and bias in grading. These form areas of active research.

References

1. Aggarwal, V., Srikant, S., and Shashidhar, V. Principles for using machine learning in the assessment

- of open response items: Programming assessment as a case study. In *Proceedings of the Workshop on Data Driven Education*, 2013.
2. Gupta R.R. et al. DeepFix: Fixing common C language errors by deep learning. In *AAAI 2017*.
3. Shashidhar, V., Pandey, N., and Aggarwal, V. Spoken English grading: Machine learning with crowd intelligence. In *Proceedings of the 21st ACM SIGKDD Intern. Conf. Knowledge Discovery and Data Mining*, KDD '15.
4. Shashidhar, V., Pandey, N., and Aggarwal, V. Automatic spontaneous speech grading: A novel feature derivation technique using the crowd. In *Proceedings of the 53rd Annual Meeting of the Association of Computational Linguistics and the 7th Intern. Joint Conf. Natural Language Processing*.
5. Singh, B.P. and Aggarwal, V. Apps to measure motor skills of vocational workers. In *Proceedings of the 2016 ACM Intern. Joint Conf. Pervasive and Ubiquitous Computing*.
6. Singh, G., Srikant, S., and Aggarwal, V. Question independent grading using machine learning: The case of computer program grading. In *Proceedings of the 22nd ACM SIGKDD Intern. Conf. Knowledge Discovery and Data Mining*, 2016.
7. Singh, R., Gulwani, S., and Solar-Lezama, A. Automated feedback generation for introductory programming assignments. In *Proceedings of the 34th ACM SIGPLAN Conf. Programming Language Design and Implementation*, 2013.
8. Srikant, S. and Aggarwal, V. A system to grade computer programming skills using machine learning. In *Proceedings of the 20th ACM SIGKDD Intern. Conf. on Knowledge Discovery and Data Mining*, 2014.
9. Takhar, R. and Aggarwal, V. Grading uncompileable programs. In *Proceedings of the Innovative Applications of Artificial Intelligence Conf. Assoc. Advancement of Artificial Intelligence*, 2019.

Shashank Srikant is a Ph.D. candidate at Massachusetts Institute of Technology, Cambridge, MA, USA.

Rohit Takhar is a research engineer at Aspiring Minds, Gurugram, India.

Vishal Venugopal is a senior software engineer at Aspiring Minds, Gurugram, India.

Varun Aggarwal is co-founder and Chief Technology Officer of Aspiring Minds, Gurugram, India.