コメント

自動採点システムは入試に採用できるか
—石岡論文へのコメント—

佐藤勝昭（東京農工大学）

石岡恒憲「日本語小論文の自動採点システムJessについて」(以下，石岡論文と称す)についてコメントを付したい。

石岡論文では，小論文試験の客観的採点をめざす試みとして開発している自動採点システムJessについて述べている。Jessシステムは，アメリカの教育用テスト機関Educational Testing Serviceが開発したE-raterという評価システムを参照して試作したE-rator日本語版ともいうべきものである。Jessは自然言語処理の手法に従って，修辞，論理構成，内容についての評価を行っており，計算機統計学の論述としては，それなりの学術的価値が認められる。

JessあるいはE-ratorによる採点が十分な意味を持ちうるのは，受験者のwriting abilityの評価においてである。修辞，構文の論理構成など「形式的」な部分については，ある程度の評価が可能であると思われる。石岡氏らは，毎日新聞の社説，コラムなどのデータベースをテストピースとしてJessに学習させ，これを模範として減点法でE-ratorのデモにある回答パターンを評価し，ほぼE-raterの採点結果と比較しうる評価結果を得ているので，Jessを文章の形式部分についての評価に用いることには，それなりの意味があると考えられる。

筆者は，石岡論文に，上述のwriting ability評価における有効性を認めながらも，自動採点システムを入学試験の小論文採点に利用することの有効性については，必ずしも肯定的な評価をすることができないので，この点についてコメントをしたい。

第1は，このシステムでは，内容の評価をすることがそもそも可能であるのか，はなはだ疑わしいという点である。石岡論文では，内容については，LSI解析を用い，問題文との一致度rを2つの文書ベクトル間の角度の余弦という尺度で測っている。問題文中のキーワードをちりばめただけの無意味な文章でも評価してしまう可能性を捨てきれない。また，科学技術用語の正確な使い方まで学習させるのは，極めて困難であろう。

第2は，この採点システムを用いた場合，修辞学的，構文的な点で評点してしまうので，このシステムを第一段評価に用いるような場合，「文章は下手だが，知識・ものの考え方・理解度などについて優れた点をもつような論述」を低く評価し，はねてしまう懼れがあるということである。科学技術の論述では，文章力よりも発想力などを重視すべきだとする議論も多い。

第3の疑問は，人間の判断を加えないで純粋にこのシステムだけで小論文の採点をするのはむずかしいのではないか。石岡氏の別論文(石岡，亀田2003)に記載されていることであるが，経営大学院の入学試験GMATにおける小論文の採点では，人間とコンピュータが独立に採点し，両者の得点差が開いた場合のみ人間の評点者が再評価するという形で試験採点の手間を半減させる目的で利用しているのである。

以上のように，小論文の自動採点は限られた分野における文章表現力の技術的評価には，ある程度の有効性をもつであろうが，これを論述内容の採点に利用することはかなり困難であり，危険ですらあると筆者は考える。適用限界の精査が必要であろう。

参考文献

石岡 恒憲・亀田 雅之，2003，「コンピュータによる小論文の自動採点システムJessの試作」『計算機統計学』16(1)（掲載予定）.