

Shakespeare and cancer diagnoses: how bard can it be?

July 24 2013

Shakespeare's plays and cancer: two seemingly unrelated topics with an underlying common thread.

The techniques that [computational linguistics](#) and computer scientists use to analyse the Bard's works are also used in cancer diagnostic procedures – and it's all down to the quantification of subtle variations of attributes present in large amounts of data.

In last month's published collaboration in the journal *PLoS ONE*, we applied a simple and novel ranking method to a dataset involving plays of undisputed authorship from the Shakespearean era.

We ranked the frequency of words by playwrights John Fletcher, Ben Jonson, Thomas Middleton and William Shakespeare, testing all 55,055 unique words used in 168 plays.

The results of using this new method were very encouraging. For some authors, such as Shakespeare, the slight under-use of particular words provided better markers of individuation than over-used words. We found Shakespeare's four lowest ranked words to be:

- all
- to (infinitive)
- now
- ye

The last one was also among the top 20 lowest ranked scores for Jonson and Middleton, but interestingly, was the top highest score for Fletcher. His preference for the use of "ye" over the average of the plays of that do not belong to him is now very clear.

These are quantifiable markers that can objectively measure an author's creative mind at work.

The idea that variations on the use of words over time can give clues about psychological problems or even markers of depression in the work of suicidal poets has already been discussed.

But this simple idea for a new scoring method may also give expected dividends in other areas, such as diagnostics, medical algorithms and [molecular medicine](#).

Information-based medicine

In a study from 2009, Shakespeare and other English Renaissance authors were studied using methods based on information theory (the scientific field that leads with the quantification of information).

They observed that Shakespeare's work seemed remarkable for its homogeneity on the probability of use of common words and for its closeness to overall average use of words at the time. This naturally triggers a central question:

Would it be possible to find some distinctive signatures of individual authors by looking at the fluctuations of the observed frequencies of words used?

So, you may be asking yourself:

Why would this be a question of interest for the analysis of biomedical data?

The identification of biological markers is critical for information-based medicine. Such biomarkers are quantitative indicators that can be objectively measured and indicate normal biological processes, the existence of pathogenic processes, or altered pharmacologic responses to a therapeutic intervention.

Biomarkers are needed for cancer diagnostics and early screening (for example, levels of the enzyme Kallikrein-3, also known as PSA or prostate-specific antigen, are often elevated in men with prostate cancer or other prostate disorders).

Biomarkers are central for the core aims of personalised medicine and the quest to individualise risk, identification, therapies and the post-treatment monitoring of possible recurrence.

But controversies exist about the use of a single biomarker (in fact, this is already happening even with established biomarkers such as PSA for prostate cancer) so current medical research advocates for finding panels of biomarkers.

Statistical scores are usually employed to rank and identify the best biomarkers when individually tested. But to identify panels it is important to find the best combination of biomarkers. Other mathematical methods are needed.

Our team uses combinatorial optimisation (the branch of computer science and discrete applied mathematics that deals with these optimal selection problems) approaches to do so, not only in cancer and the selection of therapeutic combinations but also in multiple sclerosis and in Alzheimer's disease.

Using panels of biomarkers it is possible to improve the classification accuracy of the tests, boosting sensitivities and specificities to approximately 90% as we have recently shown in studies in Alzheimer's Disease.

Finding the best fit

This is not the first time that combinatorial optimisation has been used at the University of Newcastle's Centre for Bioinformatics, Biomarker Discovery and Information-based Medicine (CIBM) both in cancer and in literature and linguistic studies.

In a different paper published in 2006, combinatorial optimisation methods were used to produce a consensus phylogenetic tree of 84 Indo-European languages. In that same study, we showed how to generate a classification of several different cancer cell lines.

Again, our approach was heavily based on combinatorial optimisation.

The application of these more sophisticated methods is necessary for personalised medicine as they can be used to subtype different types of cancers at the molecular level by analysing patterns of variations across different samples.

While our team's work concentrates on developing molecular signatures of disease states based on a combination of biomarkers (as opposed to single scores like the novel one used in our study) we also recognise the usefulness of this new score, presented in the analysis of Shakespeare's works, for a rapid preliminary analysis of large biomarker datasets.

Our team now routinely analyses large biomedical datasets with this new method. As in the Shakespeare study mentioned above, it has served to identify potentially mislabelled samples, outliers of a major class of

interest of a disease, and other potential pitfalls identifiable and avoidable during early processing of the data.

For our institution our new contribution accounts as one of those success stories of collaboration across faculties and disciplines, a rare curiosity-driven basic research endeavour that generally does not get the nod from national funding agencies that only look to support translational medical research with simplistic definitions.

These "unthinkable quests" are vital to spin-off breakthrough translational research.

They need to be protected, supported and developed as computer science provides the core expertise that may lead to new scalable ways to address the tidal wave of data coming from the life sciences that may ultimately result in a blessing for your health.

More information: www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0066813

This story is published courtesy of [The Conversation](#) (under Creative Commons-Attribution/No derivatives).

Citation: Shakespeare and cancer diagnoses: how bard can it be? (2013, July 24) retrieved 24 April 2024 from <https://medicalxpress.com/news/2013-07-shakespeare-cancer-bard.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--