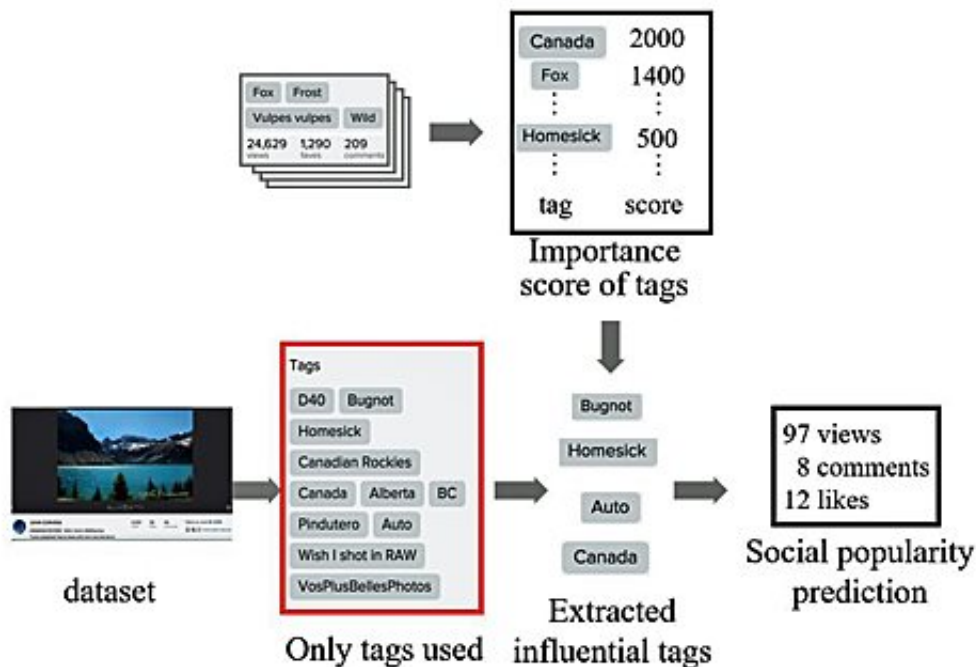# Swimming in a deluge of user generated content

October 27 2015, by Sean Kuno

The Internet is awash in user generated content (UGC)—from blogs, reviews, and Q&As, to wikis, tweets, and Facebook posts. And let's not forget photo- and video-sharing sites: every second, one hour of video is uploaded to YouTube, and an average of more than 80 million photos get added to Instagram every day. It seems we can't get enough of posting our own content and reading or viewing other people's. With technological advances making it ever easier to create UGC, the

phenomenon is only getting bigger. Also getting bigger is the task of navigating all that UGC. With so much content to sift through, how can people find quality UCG? Conversely, how can UCG creators attract more eyeballs to their content?

Toshihiko Yamasaki, an associate professor of information and communication engineering at the University of Tokyo, has been researching the challenges posed by the UGC deluge and how people can generate better, more popular content.

Working in collaboration with Microsoft Research, Yamasaki has focused on the text tags that are commonly used to classify and identify a given UGC item. His research project, named CORE 10, is unique in that it uses algorithms to quantify which tags—such as Facebook "likes"—contribute to the popularity of the content. His algorithms address two challenging problems: (1) estimating the numbers of views, comments, and favorite ratings of images and videos on content sharing services, using only the text tags associated with the content, and (2) recommending the addition (or, in some cases, the deletion) of tags to bolster the popularity of the associated content.

The lead researcher from the Microsoft side, Tao Mei of Microsoft Research Asia, praises the spirit of cooperation and collegiality that has characterized the project. "I enjoy the collaboration and hope Professor Yamasaki can make a bigger impact through the collaboration with Microsoft Research Asia," Tao says.

That "bigger impact" seems all but ensured, as Yamasaki's algorithms are now being applied to analyze the popularity of spoken presentations, such as TED video talks. His research group has proposed a method to predict, accurately, the impact of multiple impression tags for a given video talk. The validity of the proposed method is currently being evaluated against a collection of TED talks, measuring 14 different

impression types. In tests using 1,646 TED presentations, the proposed method provided accurate predictions 93.3% of the time, outperforming conventional methods.

Yamasaki's spoken presentation analysis is the product of collaboration with Danushka Bollegala, a senior lecturer in the Department of Computer Science at University of Liverpool in the United Kingdom. An expert on artificial intelligence and computational linguistics, Bollegala formerly worked at the University of Tokyo and has collaborated on natural language processing studies with Microsoft Research Asia. Commenting on his work with Yamasaki, Bollegala explains, "I have been mainly involved in the machine learning and text processing side of this project. Our collaboration has been fruitful, and we have written a paper together which has been accepted [for presentation at] a workshop during [the] ACM Multimedia [Conference]."

One of the most promising and innovative applications of the research is an automatic feedback system for personal presentation talks and MOOC (massive open online course) lecture videos. The method predicts multiple impression categories such as "inspiring," "informative," "unconvincing," and so on for a single presentation. The predicted impression categories can be used for content analysis as well as for content recommendation or retrieval purposes. The predicted impressions can also provide the presenters with useful feedback to help them improve the quality of their talks.

So, what's next for Yamasaki and CORE 10? E-commerce and the TV industry. Yamasaki is currently using his algorithms to analyze product descriptions, identifying those words that drive likeability and thus potentially increase product sales. He is also conducting preliminary work to forecast TV ratings based on the cast, popularity, and synopsis of a program before it airs—essential information for producers and

sponsors.

Yamasaki has been elated by the collaboration with Microsoft Research. "They not only give us financial support, but they also provide us opportunities to collaborate with great researchers in Microsoft Research Asia. Their events are also good opportunities to get to know Japanese researchers in different disciplines. Thanks to the attractive collaborations with Tao Mei, Danushka Bollegala, and other researchers, our projects are more fruitful than I first expected. Our collaborations are still expanding to other research topics."

 More information: www.acmmm.org/2015/

Provided by Microsoft

Citation: Swimming in a deluge of user generated content (2015, October 27) retrieved 1 May 2024 from https://phys.org/news/2015-10-deluge-user-content.html