TAIG Competition Winners Describe Research

Tommy Jones and Yeseul Jeon were recently named winners of the 2021 Text Analysis Interest Group (TAIG) presentation competition for talks they gave at the 2021 Joint Statistical Meetings.

Jones is a PhD candidate at George Mason University, a senior member of the technology staff at In-Q-Tel, and vice president of Data Community DC. Here, he summarizes his research:

Tommy Jones is a PhD candidate at George Mason University.

Yeseul Jeon is a PhD student at Yonsei University.

> My research is developing some statistical theory for analyzing language, focused on latent Dirichlet allocation (LDA). Statistical theory for language, which I call "corpus statistics," can allow us to measure linguistic phenomena with the same rigor that we use to measure, for example, economic phenomena like unemployment and prices. Then, businesses, researchers, and policymakers can incorporate cultural zeitgeist into their analyses in a principled way. LDA is a great model for corpus statistics because it's a Bayesian probability model, allowing us to leverage existing best practices and embedding language into a probability space where relationships between points are interpretable and well-defined.

Jeon is a PhD student at Yonsei University. She investigated effective estimation and visualization of topic interactions in the context

of COVID-19 research. She describes her research below:

> I have been interested in text data mining and modeling. Since late 2019, under guidance of my adviser, Dr. Ick Hoon Jin, and Dr. Dongjun Chung, both of whom have considerable experience and research in statistical modeling and biomedical big data mining, I had investigated COVID-19 using text data mining and modeling. This was a really exciting problem for me because of its worldwide importance and a large inflow of researchers across the globe to publish numerous papers on the subject.

Jones and Jeon were invited to elaborate on their prize-winning research during an upcoming Data Science DC (DSDC) Meetup (*www.meetup.com/Data-Science-DC*), sponsored by the TAIG and Washington Statistical Society.

TAIG serves as the bridge between the mainstream statistical community represented by the ASA and the growing field of text analysis, defined broadly (e.g., text mining, natural language processing, computational linguistics, web scraping, sentiment analysis, topic modeling, GAN text generation, automated translation, etc.).

For more information about TAIG or to join, visit *https://bit.ly/3dZM0m8*. Questions can be emailed to the executive committee at *asataig@gmail.com*. ∎

EDITOR'S NOTE
The views expressed here are the authors' and do not necessarily represent those of their organizations.