

1154-68-1302

Amber Hu* (amber.hu@yale.edu), **Rodrigo Leonardo** (rodleonardoc@outlook.com) and **Mohammad Uzair** (muzair.bsccs15seecs@seecs.edu.pk). *Fusing Visual and Textual Information to Determine Content Safety*.

In advertising, identifying content safety is a significant concern since advertisers do not want brands to be associated with unsafe content. At the same time, publishers would like to open as much inventory as possible. Thus, a fine balance must be achieved to satisfy both advertisers and publishers. In this talk, we propose a multimodal machine learning framework that fuses visual and textual information from web pages to produce content safety predictions. The primary focus is on late fusion, a multimodal approach that involves combining final outputs of separate modalities, such as computer vision and natural language processing, to arrive at a decision. We developed a fully automated framework for multilabel classification using late fusion techniques. We also introduce work in early fusion, which involves extracting and fusing intermediate features from separate models to produce predictions. Our algorithms are applied to data provided by GumGum, Inc., a company that leverages machine learning in the advertising industry. Both of our late and early fusion methods obtain significant improvements over state-of-the-art algorithms. This work was completed at RIPS-Los Angeles. (Received September 14, 2019)