Case ID:M16-052P^

Published: 2/26/2020

Inventors

Jundong Li

Xia (Ben) Hu

Jiliang Tang

Huan Liu

Contact

Shen Yan
shen.yan@skysonginnovations.com

Unsupervised Streaming Feature Selection in Social Media

The explosive and viral nature of social media creates massive amounts of high-dimensional data. Features such as hashtags, keywords, or slang words are created every day and quickly become popular within a short period of time. However, social media data is not fully structured and its features are usually not predefined. One way to resolve this problem is feature selection, which aims to select a subset of relevant features for a compact and accurate representation of the data. Traditional feature selection assumes that all features are static and known in advance. However, this assumption is invalid in social media where features are generated dynamically, new features are sequentially added, and the size of features is unknown in most cases. Currently, streaming feature selection (SFS) is used to efficiently process candidate features and rapidly adapt to changes. The vast majority of existing streaming feature selection algorithms are supervised and utilize label information to guide the feature selection process. However, social media data is often unlabeled and requires extensive time and labor to obtain or assign labels for supervised SFS processing. Therefore, there needs to be an efficient method for processing unlabeled social media data.

Researchers at Arizona State University have created a method that is able to process unlabeled dynamically-changing features without supervision. The framework, called USFS (unsupervised streaming feature selection), allows for the effective and efficient selection of features at a higher accuracy than existing unsupervised algorithms. This novel method determines and selects relevant features in streaming social media data by exploiting posted link information. The model has also been empowered to evaluate and add newly arrived features, as well as remove existing features automatically. This method can be used for increased accuracy and efficiency in the mining of social media data.

Potential Applications

- Social Media
- Data Analytics
- Journalism and News Reporting

Benefits and Advantages

- Improved accuracy – Increases accuracy for feature selection and produces more relevant data for the user.
- Increased efficiency – This unsupervised method reduces time and effort needed to process unlabeled data.
- Ideal for dynamic systems - An alternative technique for extracting features

of potential interest from dynamically changing data sets.

For more information about the inventor(s) and their research, please see

[Dr. Huan Liu's directory webpage](#)