

Terms and Application Form of using CMCD

Beijing Institute of Technology & University of Science and Technology of China

Contacts: Yong Hu / email: huyong@bit.edu.cn

CMCD: The Cross-Modal Classification Dataset (CMCD) is a large-scale dataset constructed from Weibo, which consists of 85,860 tweets from 18 general categories, and all of them have been manually labelled and adversarial filtered. Tweets in the dataset has three modalities of text, image and video, and 64.4% of tweets contain images and 16.2% of tweets contain videos. We hope this dataset could promote the research on cross-modal classification on social network.

Terms of use:

1. Respect the privacy of personal information of the original source.
2. The original copyright of all the data of the CMCD belongs to users on Weibo and Weibo platform. Yong Hu et al. collect, organize, filter them. CMCD is free to the public for only academic research.
3. The dataset is only for the specified applicant or study groups for research purposes. Without permission, it should not be used for any commercial purposes.
4. If you want to use the dataset for further research, data providers, Yong Hu et al. should be identified in your results.
5. If the terms changed, the latest online version shall prevail.
6. If you want to use the corpus, please download the application form, fill sign it respectively, then e-mail the scan version to Yong Hu (email: huyong@bit.edu.cn)

Application Form:

Name: _____

Affiliation: _____

Email: _____

Address: _____

I have read the above terms and conditions, and are willing to accept the terms. If violated, I will be responsible for the consequences. The following part will be signed **by the person in charge of the research group** the applicant in.

Name: _____ Position: _____

Signature : _____ Date: _____