Proposal for CIE6023(MDS6232) Final Project 2021

Names & IDs: Chuxuan Fan (219012002), Peng Deng (220041042), Jiadong Lou (220041045)
Title: Multi-modal Online Product Similarity Assessment

Description:

- **Task and goal**

Our project is based on a Kaggle contest called Shopee - Price Match Guarantee[1]. The **aim** of this task is to **match product** (according to images and titles), so that when a customer makes purchase decisions, he can make sure that he gets the product from the cheapest retailer.

- **Dataset**

We are provided with a dataset with 34250 training images, each of them is matched with some other product images. Additionally, we have titles for each image from publishers.

- **Previous work and why previous methods not practical in this task?**

In image similarity assessment research, conventional way is manually-designed features descriptor like SIFT (Lowe, D. etal, 2004), DAISY (2010), pHash (Zauner, C. etal, 2010), SSIM (Wang, Z. etal, 2004) etc. However, compared with deep learning, which is an end-to-end feature representation, hand-designed descriptor performs worse. With development of deep learning, Siamese network is proposed (Bromley, J. etal, 1993; Chopra, S. etal, 2005). With top network serving as a similarity function, image pair is fed into the two branches function as two descriptors. Many methods are raised to adjust traditional Siamese Network to improve model performance (He, K. etal, 2015; Zagoruyko, S. etal, 2015).

In sentence similarity assessment research, the methods can be classified into three main categories: word based, structure based and vector based. In order to compare semantic similarity, learning based vectors using a simple modification on LSTM can be used (Thyagarajan, A. etal, 2016).

However, current methods cannot reach the expectation of Shopee. It is very common that two different images of similar wares may represent the same product or two completely different items. They believe that with additional title, product matching performance can be enhanced.

- **Our model and potential innovation**

In order to combine the similarity of images and titles, multimodal fusion at feature level can be used to extract the joint feature presentation (Baltrusaitis, T. etal, 2017). Then the joint presentations can be fed into adjusted Siamese Network to assess similarity.

- **Expected results**

Our model is assessed by F1-Score. With the multimodal model, we expect he features from images and titles can complement each other and then to achieve a well performance. Some different images (but the same products) can get high similarity with the help of title description.

Tentative Timeline/To-do lists:

- Apr 4 – Apr 11: Paper Review
- Apr 11 – Apr 18: Code Review and Study in Kaggle Discussion
- Apr 18 – Apr 25: Code Implementation
- Apr 26 – May 9: Adjustment and Improvement
- May 9 – May 16: Paper Writing

[1] https://www.kaggle.com/c/shopee-product-matching/overview

Reference

Lowe, D. . (2004). Distinctive image features from scaleinvariant keypoints. International Journal of Computer Vision, 60.

(2010). Daisy: an efficient dense descriptor applied to wide-baseline stereo. Tpami, 32.

Zauner, C. . (2010). Implementation and Benchmarking of Perceptual Image Hash Functions. revista musical chilena.

Wang, Z. . (2004). Image quality assessment : from error visibility to structural similarity. IEEE Transactions on Image Processing.

Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., & Shah, R. (1993). Signature verification using a" siamese" time delay neural network. Advances in neural information processing systems, 6, 737-744.

Chopra, S. , Hadsell, R. , & Lecun, Y. . (2005). Learning a Similarity Metric Discriminatively, with Application to Face Verification. null. IEEE Computer Society.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 37(9), 1904-1916.

Zagoruyko, S., & Komodakis, N. (2015). Learning to compare image patches via convolutional neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4353-4361).

Thyagarajan, A. . (2016). Siamese Recurrent Architectures for Learning Sentence Similarity. Thirtieth Aaai Conference on Artificial Intelligence. AAAI Press.

Baltrusaitis, T. , Ahuja, C. , & Morency, L. P. . (2017). Multimodal machine learning: a survey and taxonomy. IEEE Transactions on Pattern Analysis & Machine Intelligence, PP(99), 1-1.