2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)

TALS: A Framework For Text Analysis, Fine-Grained Annotation, Localisation and Semantic Segmentation

Shatha Jaradat, Nima Dokoohaki, Ummal Wara, Mallu Goswami, Kim Hammar, Mihhail Matskin
Department of Software and Computer Systems
KTH - Royal Institute of Technology
Stockholm, Sweden
{shatha,nimad,ummul,mallu,kimham,misha}@kth.se

***Abstract*—With around 2.77 billion users using online social media platforms nowadays, it is becoming more attractive for business retailers to reach and to connect to more potential clients through social media. However, providing more effective recommendations to grab clients' attention requires a deep understanding of users' interests. Given the enormous amounts of text and images that users share in social media, deep learning approaches play a major role in performing semantic analysis of text and images. Moreover, object localisation and pixel-by-pixel semantic segmentation image analysis neural architectures provide an enhanced level of information. However, to train such architectures in an end-to-end manner, detailed datasets with specific meta-data are required. In our paper, we present a complete framework that can be used to tag images in a hierarchical fashion, and to perform object localisation and semantic segmentation. In addition to this, we show the value of using neural word embeddings in providing additional semantic details to annotators to guide them in annotating images in the system. Our framework is designed to be a fully functional solution capable of providing fine-grained annotations, essential localisation and segmentation services while keeping the core architecture simple and extensible. We also provide a fine-grained labelled fashion dataset that can be a rich source for research purposes.**

***Index Terms*—deep learning, word embeddings, natural language processing, annotations, fine-grained, localisation, semantic segmentation, dataset**

I. Introduction

With the advent of the numerous social networking platforms, people have found more ways to be socially connected, and to share their opinions and activities. At the same time, business retailers and marketers have found effective methods for advertising their products and connecting with larger audience for a better recognition of their brands. For example, social ads is one way that helps a business strengthen the bond with clients and establish new ties with potential ones. If clients are given suitable recommendations that match what they are looking for, the possibility of additional purchases increases considerably. That's why understanding the target audience is very crucial for better sales. According to recent statistics [1], 71% of US businesses use Instagram, 60% of users discover products on Instagram, 75% of users take action, and 90% of the top brands are on Instagram.

By providing recommendations through social media, business retailers can target specific users with predetermined characteristics such as gender and age for selling specific products [2]. These are just some of the variables which allow them to accurately identify their target audience. Nowadays, digital influencers are playing a major role in advertising products and increasing the brands' visibility in Instagram. The clicks and interactions that users do while browsing influencers' posts can be efficient metrics for measuring the value of the brand and for collecting information about the users' preferences. However, for better conclusions about what they are attracted to, a deeper analysis of the images and posts they like or write would provide better input for smarter recommendation systems. For example, when a user interacts with a certain fashion influencer' posts in a frequent basis, this might be due to her interest in the overall style of that influencer or/and of the details and combinations of clothing items in their outfits. For reaching such smart conclusions and grab users' attention, advanced deep learning architectures for text and images should be employed together with the recommendation systems.

An analysis of an image and its text and comments such as a fashionista's Instagram post defines the following tasks: (a) detection and localisation of the elements that appear in the image, (b) pixel-by-pixel semantic segmentation which is more precise than object localisation as an outline of every object in the input image is decided and classified for semantic segmentation, (c) classification of an image into a large space of possible fashion brands, clothing categories, subcategories and style attributes. For example: Dress (category) - Casual (sub-category) - Satin, Floral (Attributes) from Zara (brand). Given that there might be more than one clothing item in the image, the task requires a multi-class multi-labeled classification, (d) semantic analysis of the text associated with the image (post's text and comments). To train deep learning architectures to achieve the mentioned tasks, a detailed dataset with hierarchical annotations and localisation and semantic segmentation meta-data is required. In this paper, we describe our framework that prepares a complete annotated, localized, and semantically segmented dataset. We also apply our word embeddings architecture

978-1-7281-2607-4/19/$31.00 ©2019 IEEE
DOI 10.1109/COMPSAC.2019.10207

described in [3] to the text collected with images to give additional information to guide the annotators during their job.

Our contributions in this paper are the following: (a) we present a complete web solution where an annotator can tag each image with multiple categories, and for each category, different attributes are selected. In our framework, the annotator localizes the fashion items, and performs semantic segmentation with user-friendly interface. The framework can be easily extended for different purpose other than fashion clothing meta-data[1]. (b) We apply neural word embeddings for semantic analysis of posts' text and comments, and we show the detected clothing categories, attributes, and style information from the text to facilitate the annotator's job. (c) We describe our rich-hierarchical fashion dataset that can be used for training advanced deep learning architectures for research purposes [2].

II. Related Work

In this section, we focus on the related work of deep learning for text mining, and we compare our dataset to the publicly available fashion datasets for research.

A. Deep Learning for NLP

Natural Language Processing (NLP) tools have been adapted widely together with machine learning in social media for information extraction [4], classification [5] and conversation modeling [6]. The major part of text mining research in online social networks focused on Twitter as an important source for information spreading. An unsupervised approach to event extraction and categorization in Twitter has been proposed in [4] where a supervised tagger is used to identify events which later get categorized using a latent variable model. Many researchers have analyzed the challenges of hashtags' clustering and classification, as they introduce more challenges in information extraction and classification tasks. In particular, in [7], authors represent the hashtags by the concatenation of the tweets in which they appeared. Then, they apply K-means clustering based on the concept of co-occurrence of terms and hashtags. Word embeddings have shown to be a great asset for information extraction. When embedding words, they are transformed from being isolated symbols into mathematical vectors that can be operated on. Word2vec [8] is one of the popular deep learning language models that produce word embeddings and capture semantic similarities between them. An example of work that applied Word2vec for text clustering is [9], where authors first train their data in Word2vec model, and then evaluate the words' similarities and use them for categorization. In [10] the utility of word embeddings for named entity recognition on twitter is evaluated with results that demonstrate improvements when using word embeddings compared to baseline methods. Most of the work has been focused on Twitter with little attention to such image sharing platforms as Instagram, and to our knowledge, no prior assessment has been made on complex, multi-label, hierarchical extraction and classification in social media. In our work [3], we focus on the task of extracting fashion attributes, such as clothing style, from text on Instagram, and we utilize semantic word knowledge to classify Instagram posts into hierarchical clothing categories based on the associated text and hashtags. The extracted knowledge is used in our annotations framework described in this paper for facilitating the visual images' annotation. The extracted and classified text is provided as hierarchical labels for our dataset that we explain in this paper. The dataset can be also a rich source for fashion trends detection and clothing recommendations evaluation for research purposes.

B. Fine-Grained Data Sources For Deep learning

The available fashion datasets in research community are focused on four sources: (a) social fashion networks, (b) fashion-show images, (c) fashion photographs from the web, and (d) online shopping websites. [11], [12] are examples of research that have focused on Chictopia.com and Pose.com as social fashion networks which content is mostly visual, consisting of the "outfit of the day" pictures uploaded by users. Other researchers gathered their data from fashion show images, such as [13] where they analyzed New York Fashion Show images for the years (2014-2015) for training their machine learning clothing attribute approach to learn clothing style features. As the web contains lots of fashion model photographs, [14] formed a large scale annotated dataset (Fashion-136) from fashion model web photos. Street2SHop dataset [15] has street photos (more than 20K) and shop photos (more than 404K) from ModCloth.com which is a large online retail store specialized in vintage style fashion. [16] introduced a dataset that has more images and attributes of a higher resolution, and supports view-invariant retrieval from several online shopping websites such as Amazon.com, Zapos.com and Shopbop.com. Most of the traditional computer vision methods for describing clothing, have focused on a small set of coarse-grained attributes.

Massive attributes lead to better partition of the clothing feature space, facilitating the recognition and retrieval of cross domain clothes images [17]. DeepFashion [17] is a large scale clothes dataset with comprehensive annotations. It contains over 800K images, which are richly annotated with attributes, clothing landmarks, and correspondence of images taken under different scenarios including store, street snapshot, and consumer. In our work, an efficient and collaborative web-solution is implemented to facilitate large-scale hierarchical category-based detail-level annotation of Instagram data. Our solution focuses on tagging each image not just by high-level categories, but also attributes per each category level. The extracted text is used to assist the annotators during their annotation. This helps us in performing better analysis of clothing features and producing better recommendations. To the best of our knowledge, all the available fashion data

[1]The code available on: github.com/shatha2014/FashionRec_DataCollection
[2]The dataset is provided upon request

sources don't contain neither such level of hierarchy nor the added value of user interactions analysis.

III. Framework Description

In this section, we describe our framework components in detail. We present our approaches to text mining, taxonomy annotation, localisation and semantic segmentation.

A. Text Analysis

Instagram hosts large volumes of unstructured user-generated text along with images. Specifically, an Instagram post can be associated with an image caption written by the author of the post, comments written by other users, and by hashtags in the image that refer to other users or brands. We have formulated the task of extracting fashion details from the text associated with Instagram posts as a mixture of information extraction and a ranking problem. As text on Instagram often is multi-lingual and noisy, the usefulness of syntactic matching and linguistic rules is reduced. For this reason, we have made use of word embeddings as a central component to extract fashion details from Instagram posts.

The extraction is carried out as follows. First, the text is tokenized with NLTKs [18] TweetTokenizer, that is designed to recognize text from social media (a tokenizer that can handle online-specific tokens such as emojis and emoticons). Then the text is normalized by removing stopwords, lemmatizing and lower-casing all tokens, extracting hashtags, emojis, and user-handles using regular expressions, and segmenting hashtags using the segmenter presented in [19]. Next, we use word embeddings trained on a large corpora of Instagram text to semantically relate the tokens in the text to a fashion ontology that contains brands, items, patterns, materials, and styles. Formally, we match tokens to the ontology based on the cosine similarity between the tokens in the text and terms in the ontology. In this mapping, we combine the cosine similarity with several other factors using a linear combination that includes the Term Frequency-Inverse Document Frequency (TF-IDF), a term-score, and a lookup in the Probase API [20] for ambiguity resolution. TF-IDF is used to give more weight to words that are characterizing for the particular Instagram post. Further, the purpose of the term-score is to give a different weight to tokens depending on whether the token occurred in the caption, a usertag, a hashtag, or in a comment, with the intuition that tokens written by the author of the post tend to be more accurate than tokens occurring in user comments. Moreover, by looking up a token in Probase we can get a sense of how ambiguous a token is based on how many possible meanings it can have in different contexts. With the output of Probase, tokens that have multiple meanings, like the homonym felt that can refer to many things, including a clothing fabric, are given less weight than non-ambiguous words, like polyester. Finally, after this mapping between the input text and the ontology, the terms in the ontology are ranked based on their accumulated matching with the input text. An example from our text analysis output is shown in

Fashion Vocabulary	Related Word
brands	nike:39.18%, converse:27.26%, adidas:18.85%, kickers:14.71%
hashtags	#treatyoself, #liketkit, #linkinbio, #ltkshoecrush, #nikewomen,
item_category	shoes:84.68%, trouser_and_shorts:6.98% jumpers_and_cardigans:4.68%, all_accessories:3.66%
item_sub_category	sneaker:71.86%, shoe:17.54%, legging:6.43%, sweater:4.16%
materials	leather:36.04%, lace:24.46%, denim:23.58%, gold:15.92%
patterns	checked:32.13%, herringbone:24.54%, striped:22.61%, colourful:20.72%
styles	sporty / casual / easy/ practical -style:34.38%, trendy / creative / unique/ fashion-forward -style:26.21%, classic / conservative / timeless/ traditional / crisp -style:19.74%, girly style:19.68%

Fig. 1. An example of text analysis results from our system. In the picture, the system detects that the fashionista wears a sneakers (shoes), legging (trousers and shorts), sweater (jumpers and cardigans)

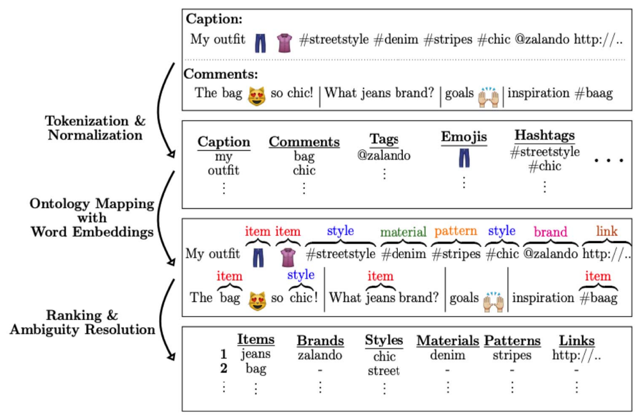

Fig. 2. Illustration of the workflow we follow for text mining in our framework. More information is available on [3]

Figure 1. The workflow for text mining in our framework is illustrated in **Figure** 2.

B. Taxonomy Annotation

An effective recommendation of clothing to users requires a detailed level annotation of their images. With the large number of clothing categories, sub-categories and attributes, this task becomes more difficult. For example, a clothing category such as dresses can have multiple **sub-categories** such as: casual, cocktail, denim, and jersey. Attributes of dresses can be related to the **patterns** such as: floral, striped, **materials** such as: lace, silk, **occasion** such as: party, business, **collar-type** such as: boat-neck, high-collar, **length** such as: long, mid. Different categories might have different attributes. For example: trousers don't have collar type attributes. To facilitate large-scale hierarchical category-based annotation of Instagram fashion images, we have developed a complete web solution where the annotator can tag each image with multiple categories, and for each category, different sets of attributes are displayed. Having such level of detailed attributes we expect to produce a better understanding of users' exact taste in fashion. Each level of details can guide us into better recommended choices of clothing and brands which will make customers more satisfied and business more successful in understanding their taste and providing them with more personalized and accurate recommendations. **Figure** 3 illustrates an example of the options from which the user can choose during annotations. Deep learning in our web application plays a vital role in aiding the annotators doing their job. The effective text mining approach that is explained in the previous section gives the annotators hints on what is available inside the image and the hashtags are very effective as well in deciding the overall style of the outfit in the image (illustrated in **Figure** 4).

C. Tools for Semantic Segmentation Compared to Our Approach

For the development of localisation and segmentation modules in our application, a research study on various algorithms and tools has been conducted. Segmentation can be defined as pixel-wise semantic classification of an image. As compared to object detection and localisation, evaluation and training of segmentation Convolutional Neural Networks (CNN) architectures is more complex and expensive. The reason behind this is that obtaining pixel-level classification is challenging. Algorithms for segmentation can be classified as: weakly supervised, supervised or semi-supervised. Seed, Expand and Constraint (SEC) [21] is a weakly supervised segmentation algorithm, which builds on VGG16 architecture [22]. SEC model works on the principle of seeding with weak localisation cues, to expand objects based on the information about which classes can occur in an image, and to constrain the segmentation to coincide with object boundaries. Another weakly supervised segmentation algorithm is "Tell me what you see, and I will show you where it is" [23]. This instance prediction framework works on guessing the absence or presence of a semantic label class. The core objective of the model is to learn from image labels where pixel-wise labelling is not required. The mentioned weakly supervised algorithms are trained only to detect general class labels e.g. cat, dog, sky. However our project required fashionista class labels to be semantically segmented. Examples of algorithms which were specifically trained and tested on fashionista datasets are Paper Doll and Fashion Parser [24]. Paper doll is a complete cloth parsing and recommendation tool which utilizes the best practices of image processing and deep learning. It relies on a set of modules such as pose estimator, a neural network which dynamically learns from the set of databases provided for fashionista. It consists of two networks: one CNN pre-trained on clothes which will work as a global network, and a second network is local and dynamic which learns instantly based on given input and retrieval-based approach. One major drawback of this model is its size which is approximately 70 GB, as it has 14 different CNN combining local and global ones mentioned above. Fashion Parser combines Fully Convolutional Network (FCN) segmentation model [25] with an outfit encoder, to make the model more flexible for fashion parsing and recognition.

To build our segmentation module to perform in a fast and efficient way, we have used Simple Linear Iterative Clustering (SLIC) [26] and Watershed algorithms for the generation of super pixels and their refinement. Super pixels are known to be powerful in image processing. The complexity of processing is reduced to a greater extent when an image is divided into super pixels. SLIC produces remarkable results in super pixel generation in a simple and uniform way. SLIC generates super pixels by clustering pixels in an image, belonging to similar colour and proximity. Sometimes generated clusters are not accurately differentiated from one another, and refining on the edges is required. Watershed algorithm is a gradient ascent method which refines the clusters edges by iterating over previous result until convergence [26]. **Figure** 3 shows an example of semantic segmentation module in our framework.

D. Tools for Object localisation Compared to Our Approach

There are a plenty of annotation tools built for object detection and localisation. Bbox annotation [3] tool written in Coffee built for object detection produces the output in JSON format consisting of bounded box dimensions. Bbox annotation is a very basic and simple tool that we considered as a core foundation for our application. Another popular tool is Pilab annotator [4] which is written in Python and processes images in series and produces output in the form of XML files. It has a richer GUI as it is built using PyQt and NumPy libraries. It supports drawing rectangle boxes for classification of objects. LabelMe [27] is another GUI rich annotation solution built using Python and PyQt. VGG image annotator is another solution for annotating and labelling images. It is an open source annotation tool based on HTML, JS, CSS and supports drawing rectangle, circle, ellipse, polygon and point. It also uses JavaScript canvas to parse images and output can be exported in JSON or CSV format. LabelImg [5] is another tool which falls in the same category, and produces annotation outcome in XML format and supports many sizes, shapes. Sloth [6] is a tool built for computer vision research purposes supports labelling of videos along with images as well. Most of the solutions use python, HTML, JavaScript as building block programming tools. A few of them also uses C++ libraries, such as LEAR - an image annotation tool. Though annotation results were satisfactory on this tool, yet C++ made it quite rigid and complex for future stages of development. For our project we have chosen the best features in terms of complexity and reliability after detailed study of the available tools. Our application is designed to be fully functional solution capable of providing essential localisation and segmentation services while keeping the core architecture simple and extensible (illustrated in **Figure** 6).

[3] https://github.com/kyamagu/bbox-annotator
[4] https://code.google.com/archive/p/pilab-annotator/
[5] https://github.com/tzutalin/labelImg
[6] http://sloth.readthedocs.io/en/latest/

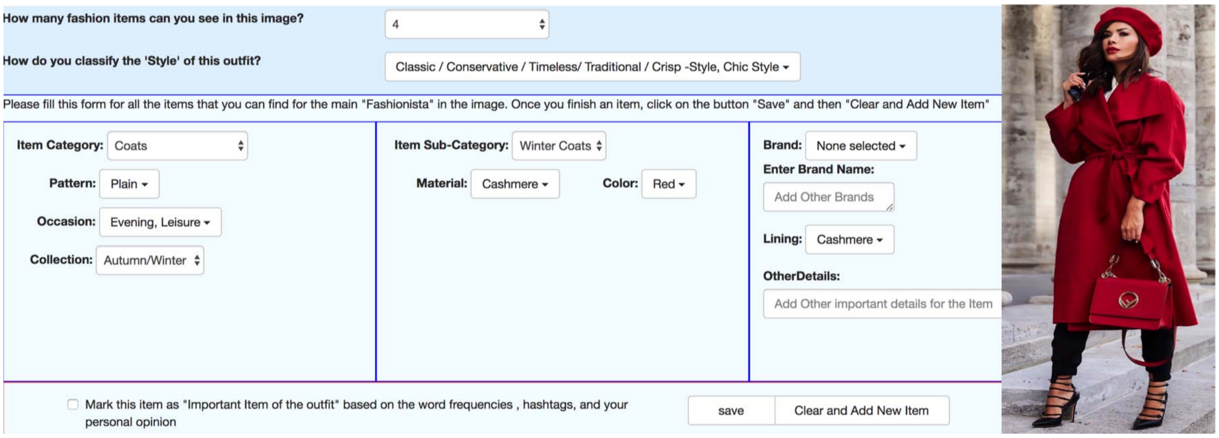

Fig. 3. A screenshot of the annotation box from which the annotator can tag multiple fashion items per image. In this example, the annotator started with the coat appearing in the image.

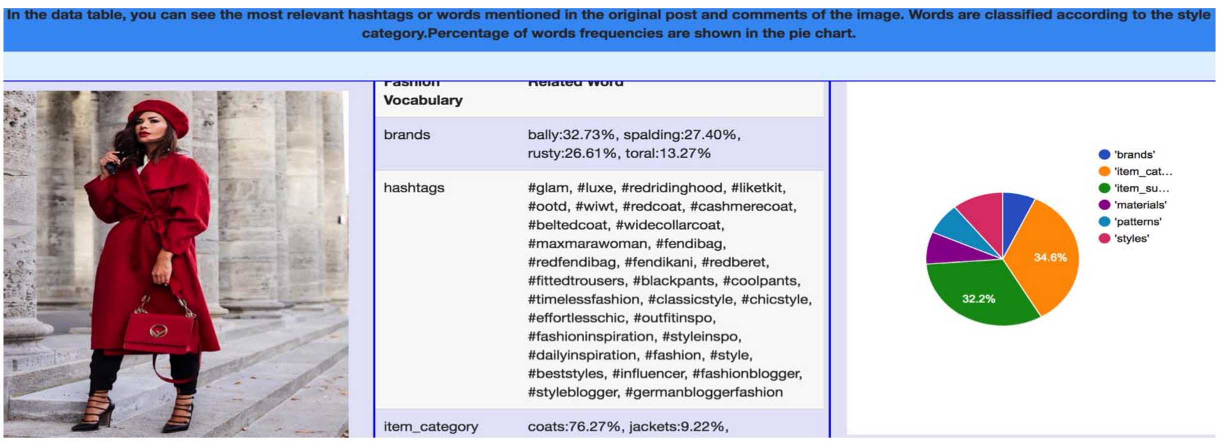

Fig. 4. Illustration of a partial set of the text analysis information that is used to aid annotators. As previously explained, we segment the hashtags to add additional weights for the detected fashion brands or clothing attributes classifications. In this example, the fashionista added hashtags such as: cashmere coat, belted coat, wide collar coat, MaxMara, Fendi, red, bag, pants, fitted trousers, classic styles and many other useful information that enhanced the classification.

Fig. 5. A screenshot of an example from the semantic segmentation module in our framework. Different colors based on the clothing categories.

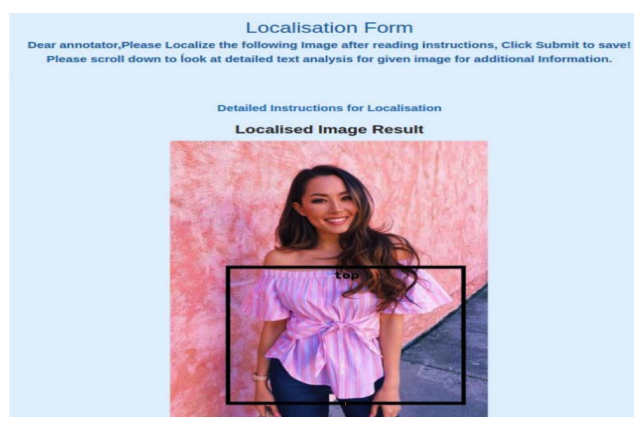

Fig. 6. A screenshot of an example from the localisation module in our framework. Based on the need, a specific localisation box can be drowned on clothing categories.

E. Fine-Grained Fashion Dataset

Our dataset contains around 70K images collected from 47 public fashionistas' Instagram accounts [7], along with hierarchical labels that represent the clothing categories (4 per image), sub-categories (4 per image), materials (4 per image), patterns (4 per image), brands (4 per image), styles (2 per image) and hashtags. In total we have 13 categories, 155 sub-categories, 176 attributes (materials and patterns), and 20 styles. These labels were generated by following the approach explained in section III.A. The developed annotation web-solution is utilized and evaluated by the team members and Amazon Turk Service for detail-level annotation, localisation, and semantic segmentation. This dataset will be a great asset for the researchers community who intend to work on a detail-annotated fashion based dataset of a social network.

F. Conclusions and Future Work

We present a complete framework for detail-level annotation, localisation and semantic segmentation. In our framework, we apply neural word embeddings approach to extract useful guides from text to facilitate the annotators job. We also provide a fine-grained fashion dataset that can be useful for training advanced neural archiectures. We are in the process of generating localisation and semantic segmentation meta-data to enrich our dataset.

References

[1] Hootsuite.com, Clarke, Todd. "24+ Instagram Statistics That Matter to Marketers in 2019". Available: https://blog.hootsuite.com/instagram-statistics/. Accessed: Feb 25, 2019.

[2] CyberClick.com, "https://www.cyberclick.es/en/advertising/advertising-on-social-media". Accessed: Feb 25, 2019.

[3] Hammar, Kim, Shatha Jaradat, Nima Dokoohaki, and Mihhail Matskin. "Deep text mining of instagram data without strong supervision." In 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI), pp. 158-165. IEEE, 2018.

[4] Ritter, Alan, Oren Etzioni, and Sam Clark. "Open domain event extraction from twitter." In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 1104-1112. ACM, 2012.

[5] Lee, Kathy, Diana Palsetia, Ramanathan Narayanan, Md Mostofa Ali Patwary, Ankit Agrawal, and Alok Choudhary. "Twitter trending topic classification." In 2011 IEEE 11th International Conference on Data Mining Workshops, pp. 251-258. IEEE, 2011.

[6] Ritter, Alan, Colin Cherry, and Bill Dolan. "Unsupervised modeling of twitter conversations." In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp. 172-180. Association for Computational Linguistics, 2010.

[7] Muntean, Cristina Ioana, Gabriela Andreea Morar, and Darie Moldovan. "Exploring the meaning behind twitter hashtags through clustering." In International Conference on Business Information Systems, pp. 231-242. Springer, Berlin, Heidelberg, 2012.

[8] Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. "Distributed representations of words and phrases and their compositionality." In Advances in neural information processing systems, pp. 3111-3119. 2013.

[9] Ma, Long, and Yanqing Zhang. "Using Word2Vec to process big text data." In 2015 IEEE International Conference on Big Data (Big Data), pp. 2895-2897. IEEE, 2015.

[10] Cherry, Colin, and Hongyu Guo. "The unreasonable effectiveness of word representations for twitter named entity recognition." In Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 735-745. 2015.

[11] Yamaguchi, Kota, Tamara L. Berg, and Luis E. Ortiz. "Chic or social: Visual popularity analysis in online fashion networks." In Proceedings of the 22nd ACM international conference on Multimedia, pp. 773-776. ACM, 2014.

[12] Nogueira, Keiller and Veloso, Adriano Alonso and dos Santos, Jefersson A. "Statistical and Deep Learning Algorithms for Annotating and Parsing Clothing Items in Fashion Photographs." Dissertation, Universidade Federal de Minas Gerais. 2015.

[13] Chen, Kuan-Ting, and Jiebo Luo. "When fashion meets big data: discriminative mining of best selling clothing features." In Proceedings of the 26th International Conference on World Wide Web Companion, pp. 15-22. International World Wide Web Conferences Steering Committee, 2017.

[14] Jagadeesh, Vignesh, Robinson Piramuthu, Anurag Bhardwaj, Wei Di, and Neel Sundaresan. "Large scale visual recommendations from street fashion images." In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 1925-1934. ACM, 2014.

[15] Hadi Kiapour, M., Xufeng Han, Svetlana Lazebnik, Alexander C. Berg, and Tamara L. Berg. "Where to buy it: Matching street clothing photos in online shops." In Proceedings of the IEEE international conference on computer vision, pp. 3343-3351. 2015.

[16] Liu, Kuan-Hsien, Ting-Yen Chen, and Chu-Song Chen. "Mvc: A dataset for view-invariant clothing retrieval and attribute prediction." In Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval, pp. 313-316. ACM, 2016.

[17] Liu, Ziwei, Ping Luo, Shi Qiu, Xiaogang Wang, and Xiaoou Tang. "Deepfashion: Powering robust clothes recognition and retrieval with rich annotations." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1096-1104. 2016.

[18] Loper, Edward and Steven Bird. Nltk: The natural language toolkit. In Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics - Volume 1, ETMTNLP 02, pages 6370, Stroudsburg, PA, USA, 2002. Association for Computational Linguistics.

[19] Baziotis, Christos, Nikos Pelekis, and Christos Doulkeridis. Datastories at semeval-2017 task 4: Deep lstm with attention for message-level and topicbased sentiment analysis. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), pages 747754, Vancouver, Canada, August 2017. Association for Computational Linguistics.

[20] Wu, Wentao, Hongsong Li, Haixun Wang, and Kenny Q Zhu. Probase: a probabilistic taxonomy for text understanding. In Proc. 2012 ACM SIGMOD Int. Conf. Manag. Data, SIGMOD 12, pages 481492, New York, NY, USA, 2012. ACM.

[21] Kolesnikov, Alexander, and Christoph H. Lampert. "Seed, expand and constrain: Three principles for weakly-supervised image segmentation." In European Conference on Computer Vision, pp. 695-711. Springer, Cham, 2016.

[22] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

[23] Xu, Jia, Alexander G. Schwing, and Raquel Urtasun. "Tell me what you see and i will show you where it is." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3190-3197. 2014.

[24] Yamaguchi, Kota, M. Hadi Kiapour, and Tamara L. Berg. "Paper doll parsing: Retrieving similar styles to parse clothing items." In Proceedings of the IEEE international conference on computer vision, pp. 3519-3526. 2013.

[25] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431-3440. 2015.

[26] Achanta, Radhakrishna, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Ssstrunk. "Slic superpixels." Ecole Polytechnique Fdral de Lausssanne (EPFL), Tech. Rep 149300 (2010): 155-162.

[27] Russell, Bryan C., Antonio Torralba, Kevin P. Murphy, and William T. Freeman. "LabelMe: a database and web-based tool for image annotation." International journal of computer vision 77, no. 1-3 (2008): 157-173. SWQ XSQW

[7] Fashionistas belong to the liketoknowit style reward community