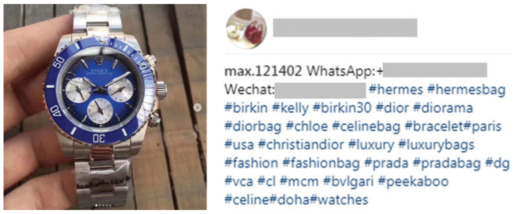
Proceedings of the Twelfth International AAAI Conference on Web and Social Media (ICWSM 2018)

Characterizing Clickbaits on Instagram

Yu-I Ha,*[*] Jeongmin Kim,[**] Donghyeon Won,[†] Meeyoung Cha,[**] Jungseock Joo[††]
[*]Graduate School of Culture Technology, KAIST, South Korea
[**]School of Computing, KAIST, South Korea
[†]Department of Electrical and Computer Engineering, UCLA, USA
[††]Department of Communication, UCLA, USA

Abstract

Clickbaits are routinely utilized by online publishers to attract the attention of people in competitive media markets. Clickbaits are increasingly used in visual-centric social media but remain a largely unexplored problem. Existing defense mechanisms rely on text-based features and are thus inapplicable to visual social media. By exploring the relationships between images and text, we develop a novel approach to characterize clickbaits on visual social media. Focusing on the topic of fashion, we first examined the prevalence of clickbaits on Instagram and surveyed their negative impacts on user experience through a focus group study (N=31). In a large-scale analysis, we collected 450,000 Instagram posts and manually labeled 12,659 of these posts to determine what people consider to be clickbaits. By combining three different types of features (e.g., image, text, and meta features), our classifier was able detect clickbaits with an accuracy of 0.863. We performed an extensive feature analysis and showed that content-based features are much more important than meta features (e.g., number of followers) in clickbait classification. Our analysis indicates that approximately 11% of fashion-related Instagram posts are clickbait and that these posts are consistently accompanied by many hashtags, thus demonstrating that clickbait is prevalent in visual social media.


[redacted]

max.121402 WhatsApp:+[redacted]
Wechat:[redacted] #hermes #hermesbag #birkin #kelly #birkin30 #dior #diorama #diorbag #chloe #celinebag #bracelet#paris #usa #christiandior #luxury #luxurybags #fashion #fashionbag #prada #pradabag #dg #vca #cl #mcm #bvlgari #peekaboo #celine#doha#watches

Figure 1: An example of clickbait on Instagram that contains many hashtags of famous brand names such as #hermes, #dior and #prada that are irrelevant to the image

Introduction

To capture people's attention in online media, some online publishers have begun to competitively use alluring headlines in news articles. Such posts, known as clickbait, are a form of misleading content marketing, which leads readers to believe that they can obtain relevant information that is not presented in the main text. In the field of journalism, the term 'clickbait' has been used to describe specific news headlines that have been crafted carefully to increase visits to a website. However, such news headlines became increasingly criticized for providing incomplete information about the main topic of the article (Beleslin, Njegovan, and Vukadinović 2017). Clickbait may provide false information regarding an article (Blom and Hansen 2015) and encourage clicking through techniques including building suspense, sensation, or teasing (Zheng et al. 2017). There are inconsistencies between the headlines or summaries of clickbaits and the actual content of clickbaits; hence, clickbaits degrade the quality of information-retrieval experiences of online readers.

Clickbait is a growing problem in all types of media markets, including visual-centric social media such as Snapchat, Pinterest, and Instagram. On these sites, an image, the main content, is typically accompanied by a short text providing additional information quickly and succinctly (Daer, Hoffman, and Goodman 2014). Sometimes, users are provided only images to click on (e.g., a search result showing thumbnails of images that match a given keyword). Therefore, people expect that the text content depicts the visual content well, and users expect that they can use text to search for certain content or interests. These two cues, however, often mismatch, posing a challenge regarding indexing and searching. Thus, clickbaits should be redefined in visual media as mismatching pairs of images and text because these posts are intentionally designed to reach a larger audience and to entice users to view ads, memes, and marketing campaigns.

Although the appearance of clickbait is commonly known on visual media (e.g., images with overlaid text "$100 free"), no study has carefully examined the other types of content that people *perceive* as uninformative or falsely enticing – information that is critical for designing platforms and algorithms. For example, Figure 1 shows a post by an online seller (which was unanimously labeled clickbait by our annotators), whose image does not contain any overlaid text. In fact, this clickbait would be difficult to detect using solely visual techniques because it would require algorithms to determine whether a given fashion product or particular brand mentioned in the text is actually referenced in the content. Our paper aims to characterize people's perceptions of clickbait on a visual social medium by focusing on a particular topic (i.e., fashion) and to identify clickbait features in terms of both visual and verbal dimensions. The topic of fashion was chosen because it not only is a popular topic of conversation on many visual media platforms but also attracts many clickbaiters and spammers.

Our research was conducted in two steps. The first step was a focus group study to qualitatively investigate the most prevalent types of clickbait, and the second step was large-scale data analysis aimed to quantitatively characterize clickbait and to develop an automated method to detect it. The key contributions are as follows:

1. We conducted an in-depth survey to understand the types of Instagram posts that may be perceived as clickbait by users.
2. To support large-scale analyses, we constructed a novel dataset of 12,659 Instagram posts regarding fashion and manually labeled whether these posts were clickbaits or not[1].
3. Using the dataset, we trained an automated clickbait classifier that combines multimodal feature channels, i.e., channels of visual, verbal, and meta features. Some of these features are general to social media, whereas some features are specific to fashion.
4. We applied the developed clickbait classifier to 450,000 fashion-focused Instagram posts, and confirmed that the dataset contained approximately 11% clickbait.
5. Since clickbait uses not only luxurious brand names but also high-street brand names in hashtags, indicators related to brand awareness were found to be good features for clickbait classification.

Our paper presents numerous new findings. We found that clickbait appears in a variety of patterns that are not always easy to characterize (e.g., zoom-in photos, scenery, and products of different brands) but nonetheless demonstrated that our combined algorithm can efficiently detect complex clickbait types. The key to this result is that the studied features are complementary to one another. Using image or text features alone can better detect clickbait than using meta features, although each type of feature enables the identification of different types of clickbait, and a combination of the features yields the best classification result. We found some brands to be more vulnerable to clickbait than others; clickbait represented 7% to 29% of the Instagram posts related to various fashion brands. Based on the results of the experiment, we provide design insights that can be used by brand marketers and researchers, in addition to service developers to better assist users on visual social media. While this study is limited to images in the fashion domain and cannot be generalized to other domains, it provides insight on how clickbaits are utilized on visual social media and what algorithms can detect them efficiently.

Background

Many online users often experience inefficient information retrieval. Such experiences involve retrieval of irrelevant content or low-quality content accompanied by incorrect titles (i.e., clickbait). The growing use of clickbaits on social media calls for a greater understanding of their effects on user actions and experience (Chakraborty et al. 2017). Spam is defined as unsolicited, irrelevant, or unwanted content or anything that hinders in performing work (Hayati et al. 2010; Patidar and Singh 2013; Verma et al. 2015), whereas clickbait is typically defined as content with headlines or summaries that are intended to entice readers and provide a small glimpse of what to expect from the article (Blom and Hansen 2015). To stimulate the curiosity of readers, exaggerated headlines or titles that appeal to the emotions are often used in clickbait articles; short messages that attract readers to click on a given link are also used by clickbait publishers (Potthast et al. 2016).

Prior studies have focused on the general form of clickbait, which features dissonance between the headlines and the accompanying body text; this characteristic is closely related to linguistic patterns (Chakraborty et al. 2016; Rony, Hassan, and Yousuf 2017; Chesney et al. 2017; Anand, Chakraborty, and Park 2017). The primary features used in classifying clickbait are the similarity between a title and the accompanying content, the informality of the post, and forward referencing, which creates information gaps (Biyani, Tsioutsiouliklis, and Blackmer 2016; Chen and Rubin 2017). One study has established types of clickbait based on the following features: exaggeration, teasing, inflammatory, formatting, graphic, bait-and-switch, ambiguous, and wrong (Biyani, Tsioutsiouliklis, and Blackmer 2016).

While existing methods may effectively detect clickbait in text media, little has been studied about visual clickbait in a visual medium such as Instagram. Images are a critical mode of communication in many social media and recent studies have demonstrated that computer vision based approaches can be used to systematically characterize various dimensions of visual communication and understand its roles in viral content prediction (Han et al. 2017), cultural preference diffusion (You et al. 2017), and social protests (Won, Steinert-Threlkeld, and Joo 2017). A recent study has attempted to classify clickbait in Twitter by integrating text and image cues but found an insignificant effect of images (Glenski et al. 2017). This study built on very generic object categories following the object detection literature, suggesting one would need to capture more relevant and specific features from images in clickbait classification.

Research Questions

The main objective of our paper is to characterize clickbait in visual social media in terms of the user experience in information retrieval. This task requires redefining the concept of clickbait, which has been mainly studied in text-based news media, in terms of the relationship between images and the accompanying text in visual media. For example, clickbaits in visual media may exhibit less consis-

[1] The final 7,769 labeling dataset adopted for this study was published in Harvard Dataverse (Title: Instagram's Clickbaits labeling data, https://tinyurl.com/yddqq939)

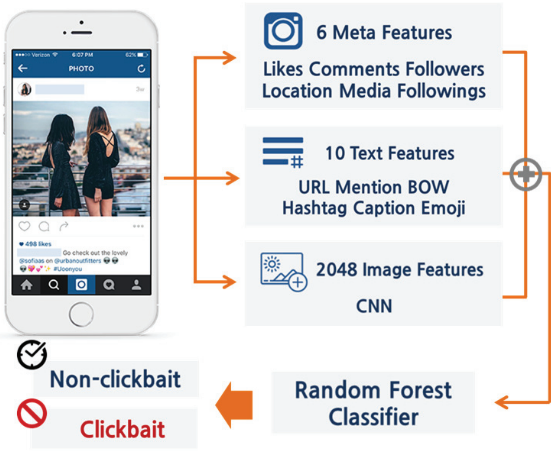


Figure 2: A diagram of the clickbait classification model applied to visual social media.

tency or relevance between the two channels – visual and verbal – whereas news clickbaits should be understood in terms of discrepancies between news headlines and the main text from a journalistic perspective. Additionally, any autonomous approach to clickbait classification in visual media will also necessitate a completely new framework that incorporates a different set of features. Our paper therefore formulates the problem of clickbait classification in a novel domain, investigates its characteristics, and seeks to develop an efficient automated classification method. Specifically, we are interested in answering the following two research questions:

- (RQ1) What types of images on visual social media are perceived by people as clickbait and how prevalent are such images?
- (RQ2) Can we automatically classify clickbait on visual social media? What types of post features are the most important for this classification task?

To answer these questions, we conducted a focus group study and established the types of posts that are perceived as irrelevant information, i.e., clickbait. We limited the scope of the study to fashion because of its popularity and its high relevance to marketing. Next, we collected annotations by asking the participants from CrowdFlower to evaluate more than 10,000 Instagram posts and to label the discrepancies between images and their accompanying text. From these data, four clickbait types, which are explained in detail later, were identified: bait-and-switch, graphic message, ambiguous, and landscape clickbait. We also trained a model based on the text-based, meta and visual features to classify clickbait posts on Instagram and compared the importance of these features for clickbait classification. This model was then applied to an unfiltered set of 450,000 fashion-related Instagram posts to examine the prevalence and patterns of clickbait usage in this medium. The overall architecture is shown in Figure 2.

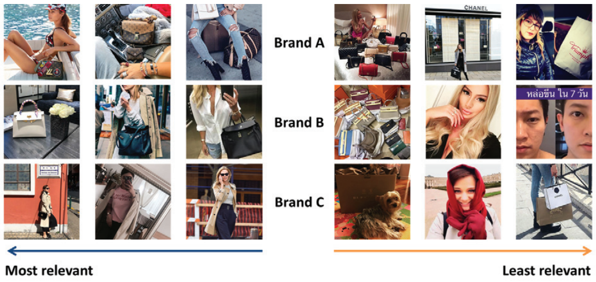


Figure 3: The top- and bottom-ranked images for three brands based on the study question on post informativeness. The respondents rated the images on the right as being the least informative regarding the given brand.

Focus Group Study: What is Clickbait?

Prior to building a clickbait classification model, we conducted a focus group study for pretesting to determine whether visual content affects clickbait perception. Although images posted on Instagram have various themes, our research focused on fashion-related information because both general users and large brands can generate images of this type of information, and such information has a high degree of popularity on Instagram. Study participants were targeted as having a high interest in fashion, being frequent users of Instagram and following at least one fashion account (a fashion brand, the owner of a local shop, a fashion journalist, or a fashion celebrity). A total of 31 users who were in their 20s or 30s were invited to in the study.

The questionnaire was designed to determine which types of photographs are considered by users to represent irrelevant posts and was composed mainly of image-based questions. The participants were shown sample search results for notable fashion brands with a set of nine images using a format similar to that used by Instagram and were asked to rate the pictures using the following question: *Which pictures are the most (or least) relevant to the mentioned brand?*

Many of the participants responded that selfies, non-fashion images, and body snaps with products from other brands were the least relevant posts in the search results for branded hashtags. In contrast, marketing images of the brand and body snap shots that included the mentioned brand products were considered to be the most relevant posts. Figure 3 shows the response examples associated with the relevance. A common characteristic of all of the least-relevant images was that the product mentioned in a hashtag (#brandproduct) presented was clearly not evident in the photograph. Therefore, the photographs that were rated as being least relevant to the mentioned brands can be considered as clickbait on Instagram, and certain visual attributes, such as faces, products, and logos in images, can be considered features used to classify clickbait on visual social media platforms.

Data Labeling

For the large-scale study, we collected fashion-related data from Instagram over a two-week period in July 2017 us-

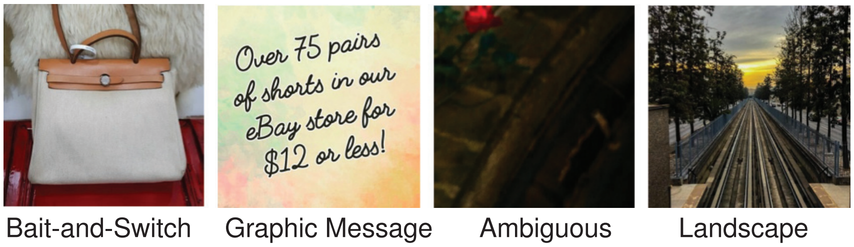


Figure 4: Four key types of clickbaits related to fashion posts and their sample images

ing the InstaLooter API, which collects only public content. We searched for public posts that included the names of famous fashion brands in hashtags. Our data include a total of 62 fashion brand names that ranged from internationally renowned high-end fashion houses such as Hermes, Loro Piana, and Louis Vuitton to high-street brands such as Uniqlo, Forever 21, and Zara. All the brands were popular on Instagram, and their official accounts had at least 50,000 followers. The final dataset contained more than 450,000 fashion posts. Each post included information about the user ID, user name, user profile picture URL, followings, followers, media count, searched brand name, hashtags, caption, image URL, likes, comments, creation time, link, and location.

We designed a task to label the gathered data and build ground truth indicating whether a given fashion image matched the associated hashtags. We employed annotators from the *CrowdFlower* crowdsourcing website (www.crowdflower.com). Three workforces were hired for each of the 12,659 posts randomly selected from the 450,000 initial posts. Upon posting a job, we limited participation to advanced contributors, which comprised a small group of more experienced contributors. Each contributor to CrowdFlower was assigned a level badge between 1 and 3, depending on how many test questions they completed and how consistent they were in answering the questions. We chose Level 2 or higher to acquire credible labeling results, as these individuals completed hundreds of different job types and had an extremely high overall accuracy. The annotators, who were paid between 3 and 5 cents for each task, assessed whether the provided image matched a hashtag.

The annotators were instructed to assign one of the three responses — yes, no, and not sure — according to whether the assigned hashtag (#fashion brand name) and image that we provided matched each other. The clickbait labeling instructions stated the following:

- Answer 'yes' if the fashion items (i.e., bag, shoes, and clothing) and text (brand logo or design) in an image match the assigned tag
- Answer 'no' if the image does not contain any fashion product or if it contains a fashion product that is not relevant to the assigned tag
- Answer 'not sure' if a fashion product is present in the image, but the annotator is not sure of the exact brand name of the product

A total of 37,977 assessments were performed by annotators for each of the 12,659 randomly selected posts. From the results of the labeling study, we employed only those posts that resulted in 100% agreement among all of the annotators as ground truth data, excluding those that did not reach a confidence level of 1.0 and those that were labeled 'not sure. Finally, we retained a total of 7,769 posts. Of these posts, 3,509 represented normal content, and 4,260 represented clickbait. Note that this ratio over-represents clickbaits and that the clickbait ratio in real systems is lower.

Types of Clickbait

The focus group study, presented earlier, identified certain image features (i.e., selfies, non-fashion images, and body snaps containing products from other brands) as *less relevant*. We investigated whether these image characteristics were commonly present in clickbaits. Images containing the above features were examined and labeled as clickbaits (yielding 4,260 samples) and then, via grounded theory, grouped into four main categories (Figure 4). Our categories indicated that clickbaits exhibit myriad patterns (not only for marketing purposes).

- Bait-and-switch clickbait, which aims to increase sales using famous-brand products or names as images or hashtags, was the most common type of clickbait that we observed. Posts of this type are intended to sell imitations of luxury-brand products or resold products.
- Graphic image clickbait demonstrates inconsistencies between the image and the hashtags. The images contain poetic phrases or promotional content (e.g., overlaid text), and the posts are accompanied by many hashtags (some of which appear legitimate).
- Ambiguous clickbait presents images containing an indistinct object. Examples include close-up photographs (of textiles or products). Annotators could not precisely identify the objects in these pictures.
- Landscape clickbait was the final type that we commonly observed. It does not contain any persons or products but features outdoor scenes that are unrelated to fashion brands.

There were a few other examples of mediating expressions found in the process of categorizing clickbait. We removed brand names that could have multiple meanings, including brand names such as #Gap, #Theory, #Coach, and #Mango, since their usage should not be judged as clickbait.

Table 1: Characteristics of the labeled data

Type	Variables	Normal Post (Average)	Clickbait (Average)	Cohen's d	Significance
Meta	Likes count	**81.7**	57.8	0.052	*
Meta	Comments count	**2.2**	1.4	0.067	**
Meta	Followings count of the posting user	901.7	**1023.6**	0.054	*
Meta	Followers count of the posting user	**6208.0**	3372.9	0.067	**
Meta	Media count of the posting user	**975.3**	906.6	0.028	p=0.206
Text	Length of hashtags	111.1	**180.7**	0.601	***
Text	Length of caption	200.5	**293.6**	0.367	***
Text	Top-100 hashtag usage count within Instagram	0.62	**1.00**	0.234	***
Text	Top-100 hashtag usage count within fashion data	1.62	**2.18**	0.247	***
Text	Count of co-mentioned hashtag pairs	0.49	**0.94**	0.197	***
Text	Emoji count in caption	1.75	**2.28**	0.086	***
Image	Selfie (whether contains a face that occupies 50% of height)	0.01	0.01	0.133	***
Image	Body snap (whether contains any body part)	**0.08**	0.02	0.412	***
Image	Marketing (whether contains runway or ceremony scenes)	**0.05**	0.04	0.243	***
Image	Product-only (whether contains products without persons)	**0.28**	0.08	0.663	***
Image	Non-fashion (whether missing fashion products)	0.59	**0.89**	0.953	***
Image	Face (whether contains frontal or side faces)	**0.09**	0.04	0.287	***
Image	Logo (whether contains any brand logo)	**0.83**	0.58	0.695	***
Image	Brand logo (whether contains specific brand logo)	**0.19**	0.11	0.664	***
Image	Smile (whether contains any smiling faces)	**0.02**	0.01	0.298	***
Image	Outdoor (whether contains outdoor background)	0.07	**0.25**	0.659	***
Type	Binary Variables	Normal Post	Clickbait	Chi-squared	Significance
Meta	Location existence	29.4%	**37.5%**	102.54	***
Text	URL included in caption	1.2%	**2.4%**	56.54	***
Text	Mention included in caption	16.9%	**18.1%**	47.21	***
Text	Emoji exists in caption	**42.7%**	37.8%	93.58	***

*p<0.05, **p<0.01, ***p<0.001

Clickbait Classification

The objective of clickbait classification is to determine whether a given post is clickbait. The problem is posed as a binary classification task, and we used a supervised learning approach with the training data explained in the previous section. The core issue in designing our classifier is the choice regarding relevant features. We utilized three types of features, i.e., visual, text, and meta features which we will elaborate on further in the following subsections.

Descriptive Statistics

We first examined the contrast between the normal post group and the clickbait group in our dataset with respect to the features in our model. To test for significant differences between the two groups, we used t-tests for numerical variables and chi-squared tests for binary variables.

Table 1 presents the significant differences between the two groups. Here, one can see that clickbait is more likely to have a location tag, URLs, a lengthy caption, and more hashtags. For certain fields, the trend will change when we use the median values. Our preliminary analysis confirms that many considered features yield statistically significant differences. Such features will be more effective in clickbait classification. We will further discuss each feature group in detail below.

Data Characteristics

Image Features As our focus group study indicated, image content plays a critical role in understanding and classifying clickbait. To systematically quantify the visual dimension of user posts, we consider two types of image features that can complement each other:

- Semantic image features: Through the user survey, we learned that fashion-related visual features are important to the relevance perception. For instance, a clean image containing a fashion product in the center is more informative than a selfie with no product image. To incorporate visual features closely related to fashion, we adopted 10 binary visual features, including selfie, body snap, marketing, product, and non-fashion, from a recent study (Ha et al. 2017). These features are semantically more meaningful.
- Generic image features: Clickbaits may also exhibit a variety of different formats, scene components and visual attributes, which may not be directly related to our fashion-related semantic features. To extract generic visual features, we adopted a popular ImageNet pretrained model (Krizhevsky, Sutskever, and Hinton 2012) and use the output of the last feature layer (2048-dimensions).

In both cases, we used a convolutional neural network based on a 50-layer Residual-Net architecture (He et al. 2016). To understand how the generic image features can help represent and cluster common visual traits in our

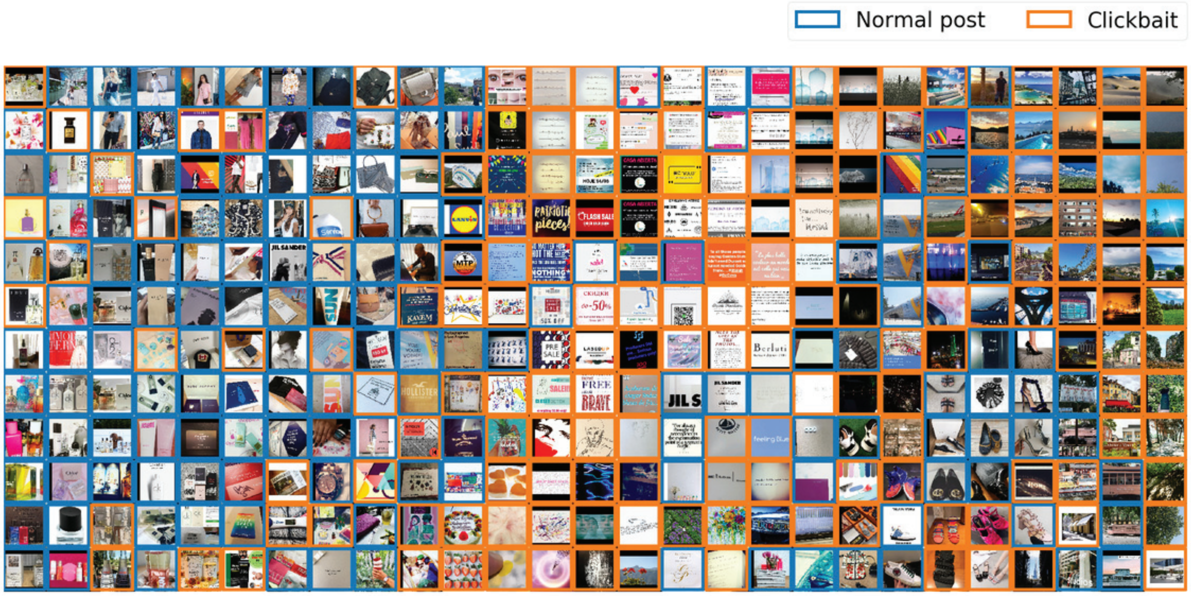


Figure 5: Visualization of t-SNE embedding of our images, obtained solely from generic image features of 2048-dimensions. The embedding is learned so that visually similar images can be placed in a similar region in the coordinate. Interestingly, the generic image features are very effective in grouping images of similar content. For example, we can see image groups for outdoor scenes (right), texts (center), bags (upper-left), and shoes (bottom-right). Note that the class information (clickbait vs normal) was NOT used to select and place the examples (i.e., purely visual-based mapping), but they are only provided to help identify their correct classes.

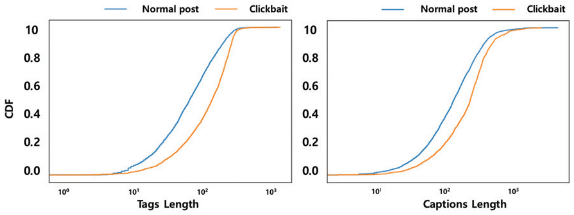


Figure 6: Difference in hashtag and caption length found in data labeled with clickbait and normal post

dataset, we visualized image samples using t-SNE (Maaten and Hinton 2008) embedding, as shown in Figure 5. For better visualization, we filled every grid with an image that had the nearest t-SNE embedding to the center of the grid. In Figure 5, images with similar content are grouped together, which indicates that the general image feature represents fashion images well.

Text Features Clickbait on visual media tends to be exposed upon search via accompanying popular hashtags that are not relevant to the image. Such hashtags include highly recognizable brand or product names and words that are popular (e.g., #love and #followme) on a platform. These hashtags are frequently included in content searches and are used to increase click rates. Figure 6 shows that there are differences in the lengths of the captions and hashtags between clickbait and normal posts in the labeled data. Thus, we examined the pattern of hashtag use by examining the lengths of hashtags.

Moreover, the numbers of the top-100 most popular hashtags on Instagram and world-famous fashion brand names are used as features. We also used the number of top-100 hashtags within our fashion dataset and the number of top-100 hashtag pairs that belong to the co-mentioned hashtag pairs as features. Moreover, because clickbait is known to utilize emotionally charged words in its headlines to attract clicks, we used emoji-related indexes, such as the presence and number of emoji, as well as their ratios, as features. Finally, the bag of words is also used as a text feature.

Meta Features To capture the characteristics of clickbait that are not directly related to either images or text, we computed the numbers of followings, followers, and media counts of the posting users and use them as features. The audience engagement metrics, such as the numbers of likes and comments, are also crucial features in classifying clickbait. Finally, the presence or absence of location tags can also have a substantial impact on certain types of content.

Classification Model

We used a random forests algorithm, which chooses a set of features randomly and creates a classifier using a bootstrapped sample of the training data (Pal 2005). We split the data, which represent 7,769 posts in total, into training and testing sets (8:2) and conduct 5-fold cross validation to optimize the model. 5,000 estimators are used in our model. We adjusted the parameters to use 5% of the features of the entire input data for classification.

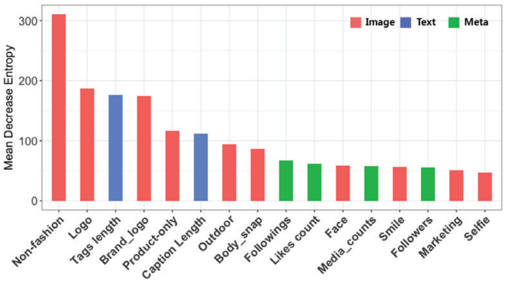


Figure 7: The feature importance for clickbait classification measured by Shannon's entropy. Text and image features are of high importance.

Model Performance Table 2 presents the results of our clickbait classification model. We report the accuracy, the ROC AUC, the average precision, and F1 score to evaluate our classification model. Our model achieves strong classification performance (with an accuracy exceeding 0.8) when both image and text features are used. The image features, in particular, yield the highest accuracy (0.837), which is slightly greater than the accuracy associated with the text features (0.83). Meta features alone yield the lowest performance, and a model combining the image and text features achieves the highest performance.

Table 2: Clickbait classification result. The combination of image and text features has strong performance in the clickbaits classification.

Feature	Accuracy	ROC AUC	Precision	F1
Meta (M)	0.622	0.691	0.646	0.542
Image (I)	0.837	0.904	0.877	0.821
Text (T)	0.830	0.903	0.872	0.820
M+T	0.827	0.905	0.882	0.818
I+M	0.834	0.905	0.880	0.819
I+T	**0.864**	**0.938**	**0.934**	**0.853**
I+M+T	0.863	**0.938**	**0.934**	**0.853**

We also note that the meta features, when added on top of the image and text features, do not lead to any performance gain, suggesting that the utility of meta features is already included in other cues, such as the image or text feature groups.

Feature Importance Although the results of the classification model demonstrate the type of feature that influences the classification of clickbait, they do not reveal the individual features that are important for classifying clickbait. Thus, we utilized the Random Forest model, a method constructed with a number of decision trees, and measure the importance of features to classify clickbait on visual social media. To estimate the importance of each feature for classification, we measured the mean decrease in Shannon's entropy. A larger mean decrease in entropy value indicates greater information gain, thus implying the importance of data classification. Figure 7 shows the feature importance of 16 effective features. The majority features in the image and text categories have the highest predictive score in our classification, whereas the meta features rank low.

The results imply that clickbait classification on visual social media requires text features that are more specific to the platform (e.g., hashtag usage) than are traditional linguistic patterns that had been used for the general problem of clickbait detection. Above all, better results are obtained in clickbait classification when image features are combined with text features rather than meta features. Thus, certain features that represent image themes and the usage patterns of certain hashtags are powerful elements for clickbait classification.

Table 3: The list of top 10 brands classified as clickbait, many of which are couture brands, but a few high-street brands are also ranked.

Type	Brand	Ratio	Popularity*
Couture	Cartier	0.218	5,734,665
Couture	Hermes	0.171	5,551,142
Couture	Chloe	0.167	5,167,322
High-street	American Eagle	0.161	2,598,674
Couture	Jilsander	0.160	178,259
Couture	Gucci	0.160	17,676,731
Couture	Alaia	0.159	500,108
Couture	Loropiana	0.157	115,546
High-street	American Apparel	0.156	2,116,050
Couture	Goyard	0.156	210,436

*Popularity measured as the number of followers of the brands' official Instagram accounts as of January 15, 2018

Clickbait Prevalence To examine the prevalence of clickbaits related to fashion brands, we applied our classifier to the entire data set and calculated the clickbait ratio. Overall, our model classified 11% of the posts as clickbait. The optimal decision threshold of 0.55 was empirically chosen based on the classification accuracy.

Clickbait features not only high-end couture groups such as Cartier, Goyard, Chloe, Gucci, and Hermes but also high-street brands such as American Apparel and American Eagle (Table 3). Clickbait does not adopt only expensive and luxurious brand names, which indicates that clickbait could have a strong bearing on brand awareness and popularity rather than brand value. In this regard, we verified that the official accounts of the top-10 brands have at least 100,000 followers. Thus, choosing metrics related to popularity, such as the number of followers of a branded account and the total number of brands mentioned on a social media platform, might be helpful in classifying clickbaits on visual social media platforms.

Feature Correlation Next, given that image and text features are equally well performing, we examined whether these features are redundant or whether they identify different types of clickbait. Figure 8 includes a scatter plot that demonstrates the clickbait classification score based on the image and text prediction as well as a few examples of classified posts. A higher score indicates that the corresponding

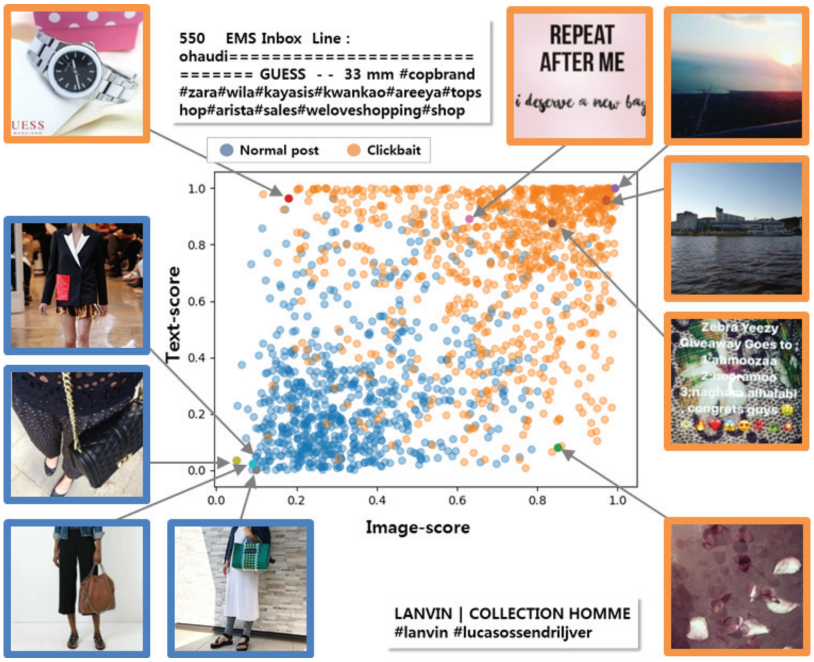


Figure 8: Clickbait classification results and sample posts based on image and text prediction scores. The higher scores of image and text features correspond to the successful classification results, and the pixels are color-coded according to the classification results.

image is more likely to be clickbait. The clickbaits and normal posts fall in the upper right and lower left corners of the scatter plot, respectively. By examining these regions, it is evident that clickbait images are of the landscape or graphic message type. Additionally, normal posts have been classified into images containing fashion-related products or branded logos. The results obtained using the captions of well-classified posts show that the clickbait posts contain a relatively large number of hashtags and that the subjects of these hashtags are not limited to fashion (e.g., *"California Hot Guys and Blue Skys #California #fun #Beautiful #Beach #Love #kush #follow #pics ..."*).

Although most of the posts receive consistent clickbait scores using image and text features, some posts exhibit unbalanced prediction scores; examples are shown in the lower right (where the image score is high, but the text score is low) and the upper left (where the image score is low, but the text score is high) corners. The cases that are accurately classified as clickbait have lengthy captions and hashtags. However, the lower right corner includes non-fashion content with short text. Overall, the unbalanced prediction scores indicate that it is difficult to classify clickbait when judged only with text or images. However, our model provides a novel approach to classify these vague posts as clickbaits by combining image, text, and meta features.

Discussion

Research Implications

Clickbait is a growing concern on visual social media; it is detrimental to marketers and consumers who seek to exchange relevant product information through the media. Our analysis suggested that a multi-modal, machine learning-based approach can successfully classify clickbait to some extent, but we also found that some clickbaits deliberately exhibit common patterns found in normal posts and are thus very hard to distinguish from legitimate posts. Moreover, our results indicate that conventional network and user engagement features (i.e., meta features), such as the number of likes, are not very informative for clickbait identification. Although service providers, such as Instagram, may attempt to develop and operate a screening algorithm to filter such content, clickbait identification will remain a challenge as clickbaiters adopt more effective and complex approaches to elude filtering.

Our large-scale analysis demonstrated that such clickbait is prevalent, irrespective of brand type or popularity. Considering that content in social media is generated by general users, not by a few media outlets as in the traditional news industry, our results may also suggest that the use of clickbait has already been established as a common tactic and practice adopted by ordinary users.

Clickbait also threatens the utility of social media as a market analysis tool because it distorts and exaggerates the actual volume of relevant content that could be searched. Such fallacious reports can mislead brands and their consumers. In a similar context, a recent study has also reported that a considerable number of online social media accounts are in fact bots and that their activities are not real (Varol et al. 2017). These findings reinforce that opinions and preferences estimated from social media can be easily manipulated

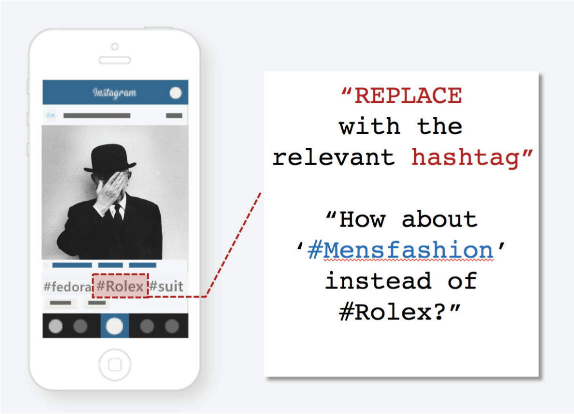

Instagram
#fedora #Rolex #suit
“REPLACE with the relevant hashtag”
“How about ‘#Mensfashion’ instead of #Rolex?”

Figure 9: An example of a posting guidance application for users that avoids the characteristics of clickbait on Instagram

and thus must be interpreted carefully.

Limitations and Future Work

There are several limitations of our current study. First, the scope of our analysis was limited to the domain of fashion. Therefore, our trained model can not be applied to other domains, such as sports or food. We believe that the text and meta features in our model are generally applicable to many topics, but the image features are domain specific and need to be articulated accordingly when applying the model to other domains. For future work, we would like to explore clickbait patterns in other domains to identify commonalities of clickbaits across domains. We will also develop an unsupervised approach that can automatically discover clickbait patterns and that can be applied to general posts about any topic.

Second, our study mainly focuses on detecting clickbait posts and measuring how prevalent they are on Instagram, whereas a further investigation is necessary to understand the impacts of such clickbait on the actual user experience. For example, frequent exposure to clickbait can undermine the credibility and perceived utility of the medium, ultimately leading people to migrate to an alternative channel. Therefore, our future work will further examine the perceptual effects of clickbait on users.

Third, our analyses in this paper mainly focus on clickbait posts and their characteristics, whereas it is also important to understand *who* posts clickbaits, i.e., clickbaiters. Unfortunately, this analysis will require a different, user-centric dataset and in-depth analyses, which are beyond the scope of the present study. We manually examined a few users in our dataset who are the most frequent posters of clickbaits. Some of these users were spammers (e.g., uploading duplicated texts multiple times), whereas some accounts were owned by online shopping malls, and others appeared to be regular users. A more comprehensive user analysis is therefore necessary to categorize user types and their purposes in using clickbait.

Many social media users choose to use popular hashtags to improve the visibility and popularity of their posts, even though these hashtags may be irrelevant to the posts. Our classifier can also be used to alert users when they attempt to use irrelevant hashtags based on the consistency between the image and text. Furthermore, the service provider can recommend new hashtags that are popular but are not considered to be clickbait (see Figure 9).

Conclusion

Filtering clickbait on social networking services is the primary solution for both strengthening brand awareness and increasing the efficiency of user information acquisition. Instagram, an online community that includes more than 400 million users (Deeb-Swihart et al. 2017), is currently the most popular image-based information sharing platform, and numerous businesses, especially fashion companies, are attempting to increase their brand awareness and sales opportunities to customers through their activity on Instagram. Many brand marketers use social networks to accurately measure their market share and product exposure. However, because clickbait has been studied primarily in journalism, many challenges remain regarding the classification of clickbaits on visual social media platforms.

In this paper, we defined and classified clickbait on Instagram, focusing on posts that exhibit incongruities between the images and the hashtags that they contain. The primary goal of this research was to characterize and classify clickbait on Instagram in terms of three different types of features, namely, text, meta and image features. Through a focus group study and crowdsourcing, we found that people regard posts that do not explicitly show the feature mentioned in a hashtag within an image as clickbaits. The image and text features outperform the meta features in clickbait classification, and, when combined, the image and text features provide the best performance. In particular, a post that combines a non-fashion image with a caption that includes a long list of hashtags regarding various topics is highly likely to be clickbait.

Our clickbait classification model is expected to be applicable to various topics and cultures. Furthermore, we can provide service developers and app designers with insight useful for application development, such as creating captions that can increase visibility while not being considered clickbaits, and help users increase the attractiveness of their posts.

Acknowledgement

This research was partly supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT (No. NRF-2017M3C4A7063570) and partially supported by Barun ICT Research Center at Yonsei University.

References

Anand, A.; Chakraborty, T.; and Park, N. 2017. We used neural networks to detect clickbaits: You wont believe what happened next! In *Proc. of the ECIR*.

Beleslin, I.; Njegovan, B. R.; and Vukadinović, M. S. 2017. Clickbait titles: Risky formula for attracting readers and advertisers. In *Proc. of the IS*.

Biyani, P.; Tsioutsiouliklis, K.; and Blackmer, J. 2016. 8 Amazing Secrets for Getting More Clicks: Detecting Clickbaits in News Streams Using Article Informality. In *Proc. of the AAAI*.

Blom, J. N., and Hansen, K. R. 2015. Click bait: Forward-reference as lure in online news headlines. *Elsevier Journal of Pragmatics* 76:87–100.

Chakraborty, A.; Paranjape, B.; Kakarla, S.; and Ganguly, N. 2016. Stop clickbait: Detecting and preventing clickbaits in online news media. In *Proc. of the ASONAM*.

Chakraborty, A.; Sarkar, R.; Mrigen, A.; and Ganguly, N. 2017. Tabloids in the era of social media? understanding the production and consumption of clickbaits in twitter. In *Proc. of the CSCW*.

Chen, Y., and Rubin, V. L. 2017. Perceptions of Clickbait: A Q-Methodology Approach. In *Proc. of the ACL*.

Chesney, S.; Liakata, M.; Poesio, M.; and Purver, M. 2017. Incongruent headlines: Yet another way to mislead your readers. In *Proc. of the EMNLP Workshop*.

Daer, A. R.; Hoffman, R.; and Goodman, S. 2014. Rhetorical functions of hashtag forms across social media applications. In *Proc. of the SICDOC*.

Deeb-Swihart, J.; Polack, C.; Gilbert, E.; and Essa, I. A. 2017. Selfie-presentation in everyday life: A large-scale characterization of selfie contexts on instagram. In *Proc. of the ICWSM*.

Glenski, M.; Ayton, E.; Arendt, D.; and Volkova, S. 2017. Fishing for clickbaits in social images and texts with linguistically-infused neural network models. *CoRR* abs/1710.06390.

Ha, Y.-I.; Kwon, S.; Cha, M.; and Joo, J. 2017. Fashion Conversation Data on Instagram. In *Proc. of the ICWSM*.

Han, J.; Choi, D.; Joo, J.; and Chuah, C.-N. 2017. Predicting popular and viral image cascades in pinterest. In *Proc. of the ICWSM*.

Hayati, P.; Potdar, V.; Talevski, A.; Firoozeh, N.; Sarenche, S.; and Yeganeh, E. A. 2010. Definition of spam 2.0: New spamming boom. In *Proc. of the DEST*.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proc. of the CVPR*.

Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. Imagenet classification with deep convolutional neural networks. In *Proc. of the NIPS*.

Maaten, L. v. d., and Hinton, G. 2008. Visualizing data using t-sne. *Journal of Machine Learning Research* 9:2579–2605.

Pal, M. 2005. Random forest classifier for remote sensing classification. *International Journal of Remote Sensing* 26(1):217–222.

Patidar, V., and Singh, D. 2013. A survey on machine learning methods in spam filtering. *International Journal* 3(10):964–972.

Potthast, M.; Köpsel, S.; Stein, B.; and Hagen, M. 2016. Clickbait detection. In *Proc. of the ECIR*.

Rony, M. M. U.; Hassan, N.; and Yousuf, M. 2017. Diving deep into clickbaits: Who use them to what extents in which topics with what effects? In *Proc. of the ASONAM*.

Varol, O.; Ferrara, E.; Davis, C. A.; Menczer, F.; and Flammini, A. 2017. Online human-bot interactions: Detection, estimation, and characterization. In *Proc. of the ICWSM*.

Verma, J.; Dhawan, S.; Verma, J.; and Dhawan, S. 2015. Clustered k-nearest neighbor process for spam detection in social networks. *International Journal for Innovative Research in Science and Technology* 2:90–93.

Won, D.; Steinert-Threlkeld, Z. C.; and Joo, J. 2017. Protest activity detection and perceived violence estimation from social media images. In *Proc. of the ACM MM*.

You, Q.; García-García, D.; Paluri, M.; Luo, J.; and Joo, J. 2017. Cultural diffusion and trends in facebook photographs. In *Proc. of the ICWSM*.

Zheng, H.-T.; Yao, X.; Jiang, Y.; Xia, S.-T.; and Xiao, X. 2017. Boost clickbait detection based on user behavior analysis. In *Proc. of the APWeb-WAIM*.