

Semi-Automatic Identification of Counterfeit Offers in Online Shopping Platforms

Patrick Arnold^a, Christian Wartner^b, and Erhard Rahm^b

^aDepartment of Computer Science, Leipzig University, Leipzig, Germany; ^bDepartment of Computer Science, Faculty of Mathematics and Computer Science, Leipzig University, Germany

ABSTRACT



Product counterfeiting in online platforms is an increasingly serious problem causing estimated losses of billions of dollars every year. The huge number of online shops and offered products call for largely automated approaches to identify likely counterfeits, although identifying counterfeits is very difficult even for humans. The authors propose the adoption of a semi-automatic workflow to inspect product offers in online platforms and to determine likely counterfeit offers based on different criteria. Such suspicious offers are to be presented to a domain expert for manual verification. The workflow includes steps to match and cluster similar product offers, and to assess the counterfeit suspiciousness based on different criteria. The goal is to support the periodic identification of many counterfeit offers with a limited amount of manual effort. The authors also present a preliminary evaluation of the proposed approach on a case study using the eBay platform.

KEYWORDS

Counterfeits; product clustering; product piracy

Introduction

From a legal point of view, counterfeiting or product piracy refers to infringements against intellectual property rights, such as copyrights, trademarks, and design rights (Organization of Economic Development 2007). Goods that violate these rights or which are purposely produced to defraud potential customers are called counterfeits, copies, replicas, imitations, knockoffs, or fakes. The volume of product counterfeiting has substantially increased in the recent past. According to the European Commission (2013), the number of registered counterfeit cases at the European Customs almost increased by a factor of 18 since 2001, especially for small parcels in express and postal traffic that most likely result from Internet sales. The International Chamber of Commerce estimates that counterfeiting accounts for about 5%–7% of world trade (i.e., about \$600 billion a year). Almost all kinds of products are subject to counterfeiting, ranging from electronic devices and apparel to food and drugs. Fake products are offered and sold in numerous online shops and on auction sites, as well as on B2B marketplaces for wholesale trading.

CONTACT Patrick Arnold  arnold@informatik.uni-leipzig.de  Department of Computer Science, Leipzig University, Augustusplatz 10, 04109, Leipzig, Germany.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/wico.

Counterfeits not only cause an enormous economic loss, but can also damage the reputation and reduce the innovative strength of a company (Wilke and Zaichkowsky 1999). Buyers of faked products not only receive a low-quality product in many cases, but may even be exposed to serious safety and health risks (e.g., in the case of faked medication) (Spring 2006; Rahm 2014).

In this article, the authors focus on product imitations sold on the web without considering infringements of digital content such as software, music, or videos. The Internet allows easy access to markets all over the world and has become a major channel to sell faked products. Taking actions against counterfeiting on the web is challenging due to the huge number of involved traders, websites, and products. Even when counterfeit offers are detected and banned from a site, it takes little effort for an infringer to reappear on another site or under a new name (Roth 2011). Furthermore, even for humans, it is difficult to identify a likely counterfeit from an online product offer as the offer can use the description and images of the genuine product.

Manually monitoring a large number of websites and product offers is almost infeasible so that purely manual approaches are insufficient for fighting counterfeits on the web. Hence, there is a strong need for automated methods to identify likely counterfeits before they are purchased by unaware customers. While different platforms and some companies try to address this problem (see next section), there is not yet a published approach to automatically or semi-automatically discover counterfeits. The authors will thus present a first principles approach to analyze product offers in web shop toward their counterfeit likelihood. Suspicious offers must be verified by a human domain expert resulting in a semi-automatic approach.

There are various parties who could benefit from such a semi-automated monitoring and counterfeit detection. Their goals differ in the number and type of products and whether they refer only to one or to several online sales platforms:

1. *Manufacturers* (owners of the trademark resp. copyright) are only interested in detecting imitations of their own products. They usually have a specific list of their products and try to find counterfeits either within a specific online sales platform or across several such platforms.
2. *Public authorities* (e.g., customs, police) can be interested in all imitations of one or several kinds of product(s), either within a specific platform or across several platforms.
3. Owners of an *online sales platform* are interested in detecting all fake products of one or several product type(s) or manufacturer(s) distributed only on their site.
4. *Consumers* want to know if the offer they are interested in is genuine or not.

As an initial step, the authors propose and analyze a semi-automatic approach to identify likely counterfeits on a single platform (case 3), although an extension to several platforms is relatively straightforward. Their approach

is based upon clustering similar product offers referring to the same product (e.g., a specific purse of a specific brand). Subsequently, they analyze each cluster and seek suspicious offers within the clusters based on different criteria, such as unusually low prices or retailer ratings. While the counterfeit candidates still must be manually verified, the approach is expected to significantly reduce the time and cost for detecting counterfeits. There are different operative and legal measures against detected counterfeiters: removing the offers infringing someone's copyright or trademark, closing the account of the counterfeiter from the platform, or suing the counterfeiter. While legal actions are complex and time-consuming, the removal of counterfeits and counterfeiters are relatively easy as first steps and suitable for mass application.

The remainder of this article is structured as follows. In section 2, the authors discuss related work. In section 3, they explain the proposed workflow and its principle techniques to find and cluster relevant offers, and to score their trustworthiness. In section 4, they present a case study for applying the proposed approach on the auction platform eBay as well as the results of their experiments. Finally, they conclude with possible future work and research directions in section 5.

Background and related work

According to the EU customs,¹ the most frequently discovered imitations at the EU border are shoes (28%) and apparel (20%), as well as accessories and luxury goods such as watches, purses, or wallets (13%), which are items most frequently offered in online web shops such as eBay. In Berman (2008), different results from manual counterfeit evaluations have been published. Test purchases made on eBay showed that up to 90% of those articles turned out to be fake items.

Traditional efforts to fight counterfeiting solely concentrate on technical means for product authentication such as holograms, serial numbers, or RFID tags (Staake, Thiesse, and Fleisch 2005; Jordan and Kutter 2012). While often successful, these methods are not universally applicable or are too expensive to control (e.g., for consumer products such as cosmetics and drugs). To detect counterfeits in web shops, these techniques become even less effective because an online customer cannot verify the hologram or RFID tag of a product. Furthermore, counterfeiters often use the so-called bait-and-switch strategy where the original image and description of a product are displayed on the website, yet an imitation is delivered (Mavlanova and Benbunan-Fich 2011).

Several theoretic studies and surveys about counterfeiting show reasons why online product piracy is such a profitable business, and analyze the people who sell them and who (deliberately or unwittingly) purchase them

(e.g., Wilke and Zaichkowsky 1999; Sharma and Chan 2011; Radón 2012). Only a few studies propose solutions on how to actually deal with product counterfeiting (e.g., Berman 2008; Ludovica Cesareo 2013), but they focus on the consumers, not on the counterfeiters. As a result of their research, most companies try to educate people about the detriment effects of counterfeiting, about quality issues, health risks, and how they can verify the validity of their purchased products. They provide websites with lists of authorized distribution partners or how to validate serial numbers. However, such measures will likely reach only a portion of prospective customers for few products and will have little effect regarding the vast number of products sold online every day.

The need to automatically monitor web shops and other e-commerce platforms for counterfeits has already been observed some years ago (e.g., Pinsdorf and Ebinger 2005; MarkMonitor 2012). There are also some private agencies such as MarkMonitor, GenuOne, or Cyveillance that support companies to detect brand abuse and intellectual property rights infringements. Auction sites like eBay also try to identify counterfeits on their platform. Understandably, the approaches used in practice have not been described in the open literature so that their details and effectiveness are unknown. Hence, the development of generally applicable but customizable approaches for discovering online counterfeits and their evaluation are open research challenges (Rahm 2014).

The approach proposed in this article automatically finds and extracts product offers from websites and clusters similar offers before the counterfeit likelihood is analyzed. There is related research for some of these tasks, especially on extracting product information (Qiu et al. 2016), matching product offers (Kim and Ahn 2008; Li and Liu 2010; Köpcke et al. 2012), or monitoring product prices (Wartner and Kitschke 2011). The current approach will base on such previous work but make specific extensions to deal with finding likely counterfeits based on a combination of criteria. The authors also present an initial case study of the proposed approach to demonstrate its viability.

A semi-automatic approach to detect counterfeit offers

In this section, the authors describe a new semi-automatic workflow to find offers for products of interest on e-commerce platforms, and to analyze and score their counterfeit suspiciousness. They first provide an overview about the workflow and then outline its main steps (query generation, data extraction and transformation, clustering of product offers, scoring) in more detail. In the next section, they will apply and evaluate the approach for an eBay use case.

Overview

Figure 1 shows the main steps of the proposed workflow. One assumes that a user specifies in an input file or selects interactively the products of interest.

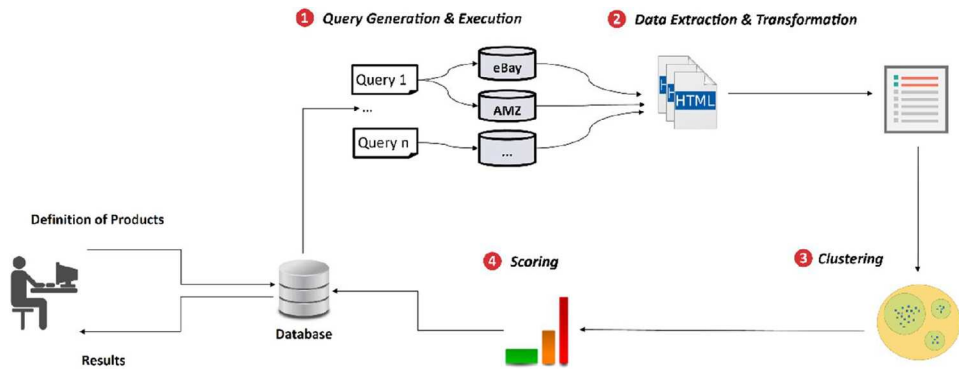


Figure 1. Workflow for detecting counterfeit offers.

Depending on her requirements, she may define specific products or broad product groups. She could also specify different web data sources (auction sites, web shops, etc.) to be examined, but this study will focus on a single site.

The workflow steps are executed once or periodically, and results can be stored in a database from where reports about potential counterfeits can be generated depending on the application needs. The main steps are the following:

1. Querying product offers: The first step is to query or search all product offers that correspond to the specified products of interest. This step entails the automatic generation of a potentially large number of suitable search queries tailored to the search capabilities of the sales platform.
2. Extraction and transformation of product offers: The search results must be processed to extract the individual product offers with their relevant attributes (e.g., product ID, title, retailer, price, date, or customer rating, depending on the data source). The product offers are preprocessed and cleaned for the subsequent steps.
3. Clustering of product offers: The retrieved product offers are clustered so that each cluster contains all offers of a specific product. This enables a comparison between different offers of the same product and can thus help to identify suspicious variations.
4. Counterfeit analysis and scoring: In the last step, the authors use different criteria to derive for each product offer a counterfeit score indicating its likelihood that it actually refers to a fake product. Suspicious offers are shown to the user for manual verification.

In the following, the authors discuss these steps in more detail.

Query generation

The basis for finding offers for counterfeit products is the search and extraction of product offers from online shops. Searching for counterfeits is based on the specification of products of interest. These may be specific products or all products of a specific type and/or from a specific manufacturer.

Manually finding all relevant offers is generally infeasible so that the authors use the automatic generation of suitable search queries based on the supported query capabilities of the considered online sales platform. Typically, the search interface supports different query predicates such as keyword search in the product title and description, as well as searching for specific attributes such as manufacturer or product type.

The challenge is to utilize these capabilities such that all relevant offers are found by using a minimum amount of queries. This goal is impaired by the typically highly heterogeneous descriptions and frequently missing information in product offers. Furthermore, there is an inherent tradeoff between recall and precision. Thus, specific product queries using manufacturer-specific product codes (e.g., “Gucci-3509/S”) or even global product identifiers such as the EAN or UPC² usually return precise results, but can lead to poor recall as this information is often not present in offers at platforms such as auction sites. On the other hand, more general queries for the product type or manufacturer (e.g., “Gucci” in the category sunglasses) have a better chance of returning all relevant offers, but will also return a large number of irrelevant offers that must be filtered out afterward. Hence, the precision for such queries is often rather low.

Putting these criteria together requires a system for a customized search for product offers in online sales platforms. According to the requirements of the user, the system must formulate search queries to retrieve the specified products by using the search interface or web API. At best, no relevant products are missing in the result set, and no irrelevant products are included. Also, the result should not contain duplicates, that is, the same offers occurring twice or more. Formulating such queries is called *query generation*. The implementation of this task is challenging but can build on recent research results (e.g., in the area of mashup applications that query information at runtime) (Barbosa and Freire 2004; Endrullis, Thor, and Rahm 2009). Some of these approaches have already been applied for finding product offers (e.g., to monitor and compare their prices) (Wartner and Kitschke 2011; Endrullis, Thor, and Rahm 2012). For a list of specific products, these approaches can either generate a single query per product (e.g., using information from the product title specified in the input), or it can be tried to find several products in a single query (e.g., if they share the same manufacturer or product type).

Data extraction and transformation

After queries are generated and executed, the results are stored and the relevant data must be extracted from the resulting web pages of the shop. Data extraction is easy if the website provides an API to programmatically submit queries and retrieve query results including relevant attributes as title, product ID, price, seller, user rating, and so on. Otherwise, web scrapers must be

employed to extract the relevant information from the HTML code of the query result pages. This process can be error-prone and implies a higher effort of creating and maintaining web wrappers.

The extracted product offers generally need further data transformation and cleaning to ensure sufficient quality for the further workflow steps of clustering and analysis. There are different techniques to remove irrelevant or invalid product offers, such as privately owned products in auction sites or offers that lack basic information like the price or title. Some product groups may require specific transformation steps to better support the comparison of their offers. For example, perfume items are typically sold in different sizes ranging from sample sizes of 5 ml up to large packages of 250 ml or more. It is thus important to extract the respective size and perhaps to also compute a normalized price (e.g., the price per ml) for easier price comparison. Other product type-specific preprocessing steps are the unification of clothing sizes and mapping synonyms to a single representation (e.g., “bag,” “evening bag,” “leather bag,” “hand bag” → “purse” in the accessories domain).

Clustering of product offers

The next step in the workflow is the assignment of the extracted offers to specific input products or the clustering of equivalent product offers. This is necessary since a counterfeit detection approach that considers differences between offers for the same product in addition to individual offer attributes is desired. Clustering or matching of offers is challenging due to their high heterogeneity and missing information. For example, product titles in offers for the same product may differ considerably. For a product officially named “Gucci Sun Dream MD-120b,” found offer titles may include “SunDream MD120b” or “Blue Gucci Sun Dream purse, model 120b.” Some offer titles may only say “Nice Sun Dream purse from Gucci,” thereby lacking significant product details.

Determining offers referring to the same product is a special case of object matching (or entity resolution) which aims at finding equivalent data objects in a dataset referring to the same real-world entity. This problem has been intensively studied already, and there are many available match approaches typically utilizing a combination of similarity scores for different attributes (Elmagarmid, Ipeirotis, and Verykios 2007). For example, one can evaluate the lexicographic similarity of the object names or other attributes. This way, an object matcher can easily recognize that two objects “Blue Gucci, MD120b” and “Gucci MD-120b (blue)” are highly similar so that they probably represent the same object. The use of dictionaries and thesauri helps to discover synonyms and to deal with homonyms. As pointed out in Köpcke and colleagues (2012), matching product offers is especially challenging and requires tailored approaches by considering information about the manufacturer and product type.

For the current authors' purposes, they not only must decide whether two offers refer to the same product, but they also want to group all matching offers for the same product in one cluster. Only then is it possible to compare all relevant offers with each other (e.g., with respect to their price). To cluster similar product offers, they propose the use of a hierarchical bottom-up clustering. The main advantage of this approach is that the number of elements in each cluster must not be known upfront, and the clusters do not overlap, which is the case with other state-of-the-art clustering approaches such as k-means (Everitt et al. 2011).

The proposed clustering algorithm is illustrated in Figure 2. Initially, each element (product offer) represents a separate cluster. The authors call the algorithm with the initial list of clusters and a minimal similarity threshold to be met by all pairs of elements in a cluster. The algorithm iteratively determines for each current cluster l the cluster j with the highest similarity. For this purpose, they define the similarity between two clusters as the smallest similarity between any two elements from the different clusters (function *CalculateSimilarity*). If the highest similarity between clusters l and j is above the minimal similarity threshold, the two clusters are merged since their offers likely refer to the same product. This process is continued as long as there are further pairs of clusters that can be merged.

Table 1 shows a simple scenario with two clusters X, Y, each containing two elements x_1 , x_2 resp. y_1 , y_2 , that are compared with each other. Thus, the authors must compare the similarity between each pair of cluster elements. The minimum value is 0.93 (similarity between x_2 and y_2), so one says that the similarity between these two clusters is 0.93. If this similarity exceeds the minimal similarity threshold and if no other cluster pair has a score above 0.93, one would combine the two clusters X and Y.

Scoring

To determine how suspicious an offer in a cluster is, one applies a scoring function to calculate a confidence score for each offer. This scoring function can consider several traits or indicators of "typical" counterfeits. One of the most significant traits of an imitation is the considerably lower price compared to the original product. Although there may be counterfeits having the same price as the original product, based on the authors' experience, it is the most important trait to find likely counterfeit offers on the web. As a low price is an important criterion for a consumer to buy fake products, analyzing the price possesses a key role in this approach (Schuchert-Güler and Eisend 2003). According to Schäfer and colleagues (2008) and Jordan and Kutter (2012), further indicators include the following:

- Retailer/seller rating, including user feedback;
- Dubious method of payments (e.g., Western Union, where a refund is not possible);

01	INPUT: Clusterlist L, minimal pairwise similarity
02	FOR EACH Cluster I in L
03	Determine cluster j in L with largest similarity s
04	IF s > minimal pairwise similarity
05	THEN <i>mergeClusters</i> (i, j)
06	ELSE quit
07	END IF
08	END FOR
09	FUNCTION <i>mergeClusters</i> (Cluster i, Cluster j)
10	Cluster k = cluster i U cluster j
11	Insert k in L
12	Remove i, j from L
13	END
14	FUNCTION <i>calculateSimilarity</i> (Cluster i, Cluster j)
15	minConfidence = 1
16	FOR EACH x in i and y in j
17	Caclulate confidence c of (x, y) using a match tool
18	IF c < minConfidence
19	THEN minConfidence = c
20	END IF
21	END FOR
22	RETURN minConfidence
	END

Figure 2. Pseudo code for hierarchical clustering.

- Country of origin of the product (since many counterfeits come from a few countries, including China);
- Missing seller or product information;

Table 1. Clustering example.

	X	
	x ₁	x ₂
	Y	
y ₁	1.0	0.98
y ₂	0.97	0.93

- Missing certifications;
- Unusual package sizes (e.g., in medication, where bulk packages suggest a counterfeit because they are usually not sold to consumers);
- Grammatical and orthographical mistakes in product description; and
- Type and volume of other products offered by a seller.

Figure 3 shows an example of a suspicious product offer with some of the mentioned indicators, such as unusually low price, payment method, and volume of available products.

By contrast, there exist traits that indicate a reliable offer (e.g., a retailer having a large product range, a generally good reputation, or whose account has existed for a long time). The current approach uses a scoring function which regards some of the criteria listed above and calculates the reliability (trustworthiness) of an offer. The scoring function is a linear combination of scores based on different indicators with each having a certain weight:

$$s = \frac{w_1*s_1 + w_2*s_2 + \dots + w_n*s_n}{n} \left(\sum_i w = 1, s_i = [0, 1] \right)$$

An indicator for the price can be a function like the following:

$$s_p = \frac{\text{price}}{\text{averageOfferPrice}}$$

This leads to a score that decreases with price deviations below the average price in a cluster of offers.

For the seller rating, one often has a direct value from the data source which can be normalized and used directly. For instance, eBay provides a rating of sellers in percentage. Other sites may use stars, which can be mapped to numerical values. Indicators like the payment method or origin can be mapped to numerical values just by categorizing origin countries or payment

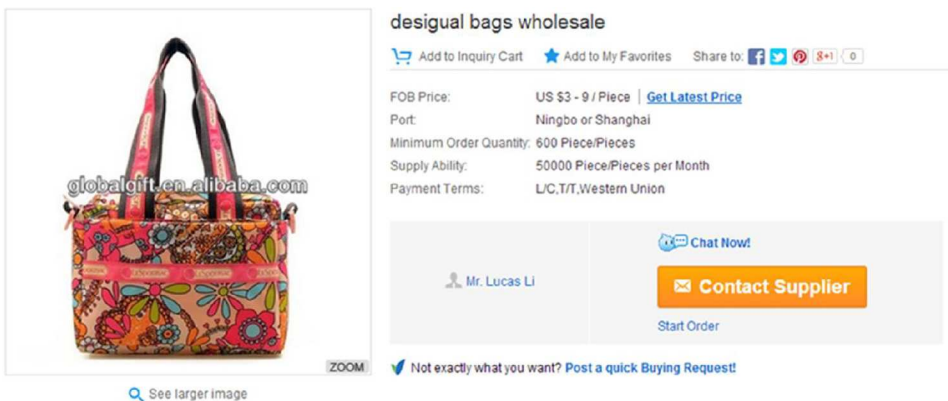


Figure 3. Suspicious product offer from the platform alibaba.com.

methods in groups and assigning each group a value between 0 and 1. The scoring function generally depends on the domain and sales platform.

Case study and evaluation

To test the applicability of the described workflow, the authors evaluated it regarding the two steps influencing the quality of counterfeit detection the most: the clustering step and the scoring step that allows the final categorization in reliable offers and potential fake offers. Without a correct clustering of similar offers, one cannot effectively use a scoring function that relies on comparisons between offers of the same type (e.g., one that puts heavy weight on price differences and not just absolute prices). If a satisfactory clustering can be achieved, the scoring function must be evaluated to give insight into the overall quality of the approach. Evaluation of query generation and data extraction is beyond the scope of this initial study. For this evaluation, the authors thus focus on relatively easy-to-find product offers. They also consider only offers from the eBay platform where they can use a comprehensive API³ to quickly gather all the information needed.

For the case study, the authors cannot provide definite answers on whether presumed counterfeit offers really refer to fake products as they do not have the real products and their assessment by experts. For evaluation purposes, they try to manually estimate the correctness of the final scoring step based on the content of the comments belonging to a seller (external and on the site), the overall offer description, as well as factors like if traders have been banned at some time after the data were extracted.

Test data

In order to judge the effectiveness of their approach, the authors generated a sample query to see how many suspicious offers were detected. They first defined a specific company and a general product. For the evaluation, they focused on several brand products that are often targeted by counterfeiters and offered on the web. They looked at multiple types of clothing products (brand shoes and shirts) and well known perfumes that fall into higher price segments. Overall, they extracted 1,341 offers (see Table 2). Offers refer to a product that can be easily found by querying the product name and manufacturer while applying a certain category as a filter for the output. The authors decided to use the product type *Eau de Toilette* for the evaluation of the final

Table 2. Test products and number of offers.

Product	Number of offers	Number of manually validated counterfeits
Brand Sport Shoes	321	—
Brand Shirt	233	—
Eau De Toilette	787	78

scoring of offers in regard to the likelihood of being offers for counterfeit goods. These offers and their respective vendors were manually flagged as counterfeits or non-counterfeits based on the user comments on the platform and external sites as well as the deletion of offers by the platform owner.

Clustering and scoring method

For clustering, the authors used the described approach on hierarchical clustering. They used a combination of lexicographic matchers to calculate the score for each element pair between clusters. For their tests, they used a minimal pairwise similarity of 0.7.

For scoring the trustworthiness of offers, they started with a scoring function that regards the price, the retailer rating, and the country of origin. Their formula originally looked as follows, with w_i being the weights and s_i being the scores of the several parameters they regard:

$$s = \frac{w_1 * \text{scorePrice} + w_2 * \text{scoreRating} + w_3 * \text{scoreOrigin}}{3}$$

$$\left(\sum_i w = 1, s_i = [0, 1] \right)$$

For price scoring, they use the following:

$$\text{scorePrice} = \frac{\text{price}}{\text{maxPrice} - (\text{maxPrice} * 0.25)}$$

Thus, the confidence of an offer decreases linearly with the deviation from the maximum price of the cluster. The authors declare a small scope below the maximum price as confidential (e.g., 25%), but products that are cheaper are gradually assigned lower scores. The scoring function for the retailer is the percentage of positive ratings retrieved for a seller, and the function for the origin assigns a score of 0.5 to the top countries frequently being involved in counterfeits and a score of 1 for other countries.

The authors planned on specifically weighting and testing multiple combinations. However, experiments quickly showed that rating and country of origin cannot be used as reliable scoring factors for the eBay platform using the chosen products. There was no correlation between rating and likelihood of counterfeit offers. Although the seller score provided by eBay appears to be a sensible criterion to judge the confidence of a product offer, it revealed to be rather misleading. First of all, the authors found possible counterfeit sellers having a perfect score of 100%, while they also found sellers who had a lower score even though their accounts belonged to well-known retailers. Note that 98% is already a very low score on eBay because most users give positive or neutral feedback.

Taking a closer look on the negative user comments, the authors found out that most users complained about long delivery times (or even no delivery of the ordered product at all), broken or damaged products, receiving the wrong product (or a wrong size, color, etc.), impolite or even abusive retailers, troubles when trying to return the product and to obtain the full refund, and so on. These are apparently the everyday problems users of eBay must deal with, while only very few complaints are about (possible) fake products. Although it might be assumed that an unreliable retailer might also tend to be a counterfeiter, such a conclusion is quite speculative. Additionally, the authors often found negative feedback referring to a disappointing product, which is beyond the responsibility of the retailer.

The scoring function generally depends on the domain and sales platform. The importance of indicators for counterfeits on the eBay platform for shoes might be different from indicators for all web shops and a different product. However, the price is an attribute that always exists and at the same time is the most important indicator. Because of that, the authors finally limited their tests to essentially using the *scorePrice* function alone.

Evaluation of clustering

Table 3 shows the effectiveness of the described clustering approach for the three different product types. Accuracy in this case refers to the percentage of clusters containing only offers for the same product (all t-shirts of type A in one cluster, t-shirts of type B in a different cluster, etc.). Offers for the same product that fall into different clusters are not inherently bad for this approach if the clusters are still large enough. The authors also specify the number of “superfluous clusters” that are not correct, but refer to a single product so that they could be merged with another cluster. On average, they obtained about 80% correct clusters and about 20% of clusters that are superfluous or contain different offers which can lead to a wrong scoring of offers.

Considering the high heterogeneity of product offers, the achieved result can be considered very satisfactory. The authors will also evaluate the quality impact for a corrected clustering where they merge the offers of superfluous clusters with their real clusters.

Table 3. Clustering quality.

Product	Number of clusters	Number of correct clusters	Number of superfluous clusters	Accuracy
Sport Shoes	37	30	4	0,81
Brand Shirt	34	29	2	0,85
EDT	54	41	5	0,76
Overall	125	100	11	0,80

Evaluation of counterfeit scoring

To gain insight in the overall quality of the scoring step and the influence of the clustering step, the authors test the counterfeit scoring with both the results of the fully automatic clustering and with a manually corrected clustering.

Table 4 shows the result of the approach with the fully automatic clustering. The scoring function returns a score between 0 and 1 that denotes how likely it is that an offer is a counterfeit product. The authors chose the thresholds 0.6, 0.7, 0.8, and 0.9 for classifying offers into reliable offers if they score above the threshold and counterfeit offers if they score below. Table 4 shows the precision, recall, and F-measure values for each of the threshold values.

The precision is the percentage of offers that were correctly classified as suspicious by the approach:

$$\text{precision} = \frac{|\text{offers automatically classified as counterfeits} \cap \text{offers manually classified as counterfeits}|}{|\text{offers automatically classified as counterfeits}|}.$$

By contrast, recall is the ratio of the manually flagged suspicious offers that could be found automatically:

$$\text{recall} = \frac{|\text{offers automatically classified as counterfeits} \cap \text{offers manually classified as counterfeits}|}{|\text{offers manually classified as counterfeits}|}.$$

The F-measure is the harmonic mean of precision and recall.

It can be seen that a high (reliability) threshold leads to more offers falling below it and it being classified as not reliable. The result then includes a high amount of the manually flagged offers, thereby improving recall. By contrast, precision improves for lower thresholds as these reduce the likelihood of false counterfeit candidates. The threshold of 0.7 produces the best balance resulting in the highest F-measure value. At this point, the recall stays stable compared with higher thresholds while the precision is already at a good level.

Table 5 shows the results when the clustering is manually corrected. In this case, precision is perfect for a threshold of 0.6 (i.e., there are no false positives in the offers classified as counterfeits). This is made possible by the correct clustering so that only the prices for offers of the same product are compared

Table 4. Results after automatic clustering step.

Threshold	Recall	Precision	F-Measure
0,6	0,064	0,833	0,119
0,7	0,564	0,543	0,553
0,8	0,564	0,283	0,378
0,9	0,564	0,22	0,317

Table 5. Results for manually corrected clustering.

Threshold	Recall	Precision	F-Measure
0,6	0,051	1	0,098
0,7	0,564	0,628	0,595
0,8	0,564	0,4	0,468
0,9	0,564	0,244	0,341

with each other, and a threshold of 0.6 refers to unrealistically large price differences. The best F-measure is again achieved for a threshold of 0.7. Compared to the automatic clustering case (Table 4), one observes the same recall but an improved precision due to the corrected clustering. The overall F-measure improved from 55% to about 60%.

Altogether, 10.3% of the offers (81 of 787) were flagged as suspicious when using a threshold of 0.7. In this case study, the authors found out that 54.3% of these offers (44) are most likely real counterfeit offers. While the evaluation results are not yet perfect, they show the viability of the proposed approach as it is possible to let the user only verify a smaller subset of the offers (81 instead of 781 offers), which can mean significant time savings for counterfeit detection.

One interesting result was that suspicious retailers who appeared in more than one cluster had offers with very bad scores in every cluster. For instance, there was one highly suspicious retailer having offers in nine clusters. The scores of the offers were very low in every cluster. This indicates it may be a good idea to use such collected evidence about suspicious retailers in future scorings.

Conclusion

The authors proposed a new approach to semi-automatically identify offers of counterfeits in online sales platforms such as large auction sites or web shops. The approach is based on a workflow with automatic search query generation for the products of interest, as well as clustering and scoring the trustworthiness of product offers. The initial evaluation showed the viability of the proposed approach but also the need for further improvements. The proposed clustering scheme worked relatively well but can be manually corrected by merging several clusters with offers for the same product. The manual verification of likely counterfeits can be restricted to smaller subsets of the product offers, thereby limiting the effort for counterfeit identification.

In future work, the authors see the need to evaluate and fine tune the proposed approach for additional sales platforms and product types. The used scoring based on price is only a first step and should be extended with additional criteria based on product type, sales platform, and perhaps insights from initial evaluations (e.g., about suspicious retailers). An inherent problem remains the difficult manual decision about whether a suspicious offer is really about a faked product. Hence, the manual verification and fine tuning

should ideally be performed in collaboration with experts affected by counterfeits such as the manufacturers of frequently faked products or platform owners.

Notes

1. http://ec.europa.eu/taxation_customs/customs/customs_controls/counterfeit_piracy/statistics/.
2. EAN = European Article Number; UPC = Universal Product Code.
3. eBay Developers Program, go.developer.ebay.com.

References

- Barbosa, L., and J. Freire. 2004. Siphoning hidden-web data through keyword-based interfaces. *Journal of Information and Data Management* 1 (1):133–44.
- Berman, B. 2008. Strategies to detect and reduce counterfeit activity. *Business Horizons* 51 (3): 191–99. doi:10.1016/j.bushor.2008.01.002.
- Elmagarmid, A., P. Ipeirotis, and V. Verykios. 2007. Duplicate record detection: A survey. *IEEE Transactions on Knowledge and Data Engineering* 19: 1–16. doi:10.1109/tkde.2007.250581.
- Endrullis, S., A. Thor, and E. Rahm. 2009. Evaluation of query generators for entity search engines. Proceedings of the International Workshop on Using Search Engine Technology for Information Management (USETIM), Lyon, France, August 24.
- Endrullis, S., A. Thor, and E. Rahm. 2012. Entity search strategies for mashup applications. Proceedings of the 28th International Conference on Data Engineering (ICDE), Arlington, April, 1–5.
- European Commission. 2013. Report on EU customs enforcement of intellectual property rights. http://ec.europa.eu/taxation_customs/customs/customs_controls/counterfeit_piracy/statistics/
- Everitt, B. S., S. Landau, M. Leese, and D. Stahl. 2011. *Cluster analysis*. 5th ed. Chichester, West Sussex, UK: John Wiley & Sons, Ltd.
- Jordan, F., and M. Kutter. 2012. Identifying counterfeit medicines with industry-suitable technologies. *Pharmaceutical Engineering, The Official Magazine of ISPE* 32 (3).
- Kim, K.-J., and H. Ahn. 2008. A recommender system using GA K-means clustering in an online shopping market. *Expert Systems with Applications* 34: 1200–09. doi:10.1016/j.eswa.2006.12.025.
- Köpcke, H., A. Thor, S. Thomas, and E. Rahm. 2012. Tailoring entity resolution for matching product offers. Proceedings of the 15th International Conference on Extending Database Technology (EDBT), Berlin, Germany, March 26–30.
- Li, G., and F. Liu. 2010. A clustering-based approach on sentiment analysis. International Conference on Intelligent Systems and Knowledge Engineering (ISKE), Hangzhou, China, November 15–16.
- Ludovica Cesareo, B. S. 2013. How to fight the online purchase of luxury counterfeit products: Manufacturer insights and strategies. *Società Italiana di Marketing* 3(4).
- MarkMonitor. 2012. Seven best practices for fighting counterfeit sales online. (W. Paper Ed.). https://www.markmonitor.com/download/wp/wp-Fighting_Counterfeit_Sales.pdf
- Mavlanova, T., and R. Benbunan-Fich. 2011. Counterfeit products on the Internet: The role of seller-level and product-level information. *International Journal of Electronic Commerce* 15(2): 79–104. doi:10.2753/jec1086-4415150203.

- Organization of Economic Development. 2007. Executive summary: The economic impact of counterfeiting and piracy. <http://www.oecd.org/sti/38707619.pdf>
- Pinsdorf, U., and P. Ebinger. 2005. Automated discovery of brand piracy on the Internet. *International Conference on Parallel and Distributed Systems 2*: 550–54. doi:10.1109/icpads.2005.99.
- Qiu, D., L. Barbosa, L. Xin Dong, Y. Shen, and D. Srivastava. 2016. DEXTER: Large-scale discovery and extraction of product specifications on the web. Proceedings of the VLDB, New Delhi, India, September 5–9.
- Radón, A. 2012. Counterfeit luxury good online: An investigation of consumer perception. *International Journal of Marketing Studies* 4(2): 74–79. doi:10.5539/ijms.v4n2p74.
- Rahm, E. August 2014. Discovering product counterfeits in online shops: A big data integration challenge. *ACM Journal Data and Information Quality* 5: 1–3. doi:10.1145/2629605.
- Roth, G. 2011. Counterfeiting in an online world. *Intellectual Asset Management (IAM) Magazine* 70–73.
- Schäfer, A., K.-H. Lang, J. Kühnert, R. Pieper, and P. Wanders. 2008. *Verbraucherleitfaden: Schutz vor Produkt- und Markenpiraterie*. Wuppertal: Institut für Arbeitsmedizin, Sicherheitstechnik und Ergonomie e.V. (ASER) an der Bergischen Universität Wuppertal.
- Schuchert-Güler, P., and M. Eisend. 2003. Non-price determinants of German consumers' inclination to purchase counterfeit products. In *Diskussionsbeiträge des Fachbereichs Wirtschaftswissenschaft der Freien Universität Berlin / Betriebswirtschaftliche Reihe*, ed. F. U. Wirtschaftswissenschaften.
- Sharma, P., and R. Chan. 2011. Counterfeit proneness: Conceptualization and scale development. *Journal of Marketing Management* 27(5–6): 602–26. doi:10.1080/0267257x.2010.489829.
- Spring, T. 2006. Fakes! *PC World* 24(2): 105–10.
- Staake, T., F. Thiesse, and E. Fleisch. 2005. Extending the EPC network: The potential of RFID in anti-counterfeiting. Proceedings of the 2005 ACM Symposium on Applied Computing, Santa Fe, New Mexico, March 13–15.
- Wartner, C., and S. Kitschke. 2011. PROOF: Produktmonitoring im Web. BTW (Datenbanksysteme für Business, Technologie und Web). Kaiserslautern.
- Wilke, R., and J. L. Zaichkowsky. 1999. Brand imitation and its effects on innovation, competition, and brand equity. *Business Horizons* 42: 9–18. doi:10.1016/s0007-6813(99)80033-0.