Generating Tips from Product Reviews

Sharon Hirsch
Ben-Gurion University of the Negev, Israel
hirschsh@post.bgu.ac.il

Slava Novgorodov
eBay Research
Israel
snovgorodov@ebay.com

Ido Guy
eBay Research
Israel
idoguy@acm.org

Alexander Nus
eBay Research
Israel
alnus@ebay.com

ABSTRACT

Product reviews play a key role in e-commerce platforms. Studies show that many users read product reviews before purchase and trust them as much as personal recommendations. However, in many cases, the number of reviews per product is large and finding useful information becomes a challenging task. A few websites have recently added an option to post *tips* – short, concise, practical, and self-contained pieces of advice about products. These tips are complementary to the reviews and usually add a new non-trivial insight about the product, beyond its title, attributes, and description. Yet, most if not all major e-commerce platforms lack the notion of a tip as a first class citizen and customers typically express their advice through other means, such as reviews.

In this work, we propose an extractive method for tip generation from product reviews. We focus on five popular e-commerce domains whose reviews tend to contain useful non-trivial tips that are beneficial for potential customers. We formally define the task of tip extraction in e-commerce by providing the list of tip types, tip timing (before and/or after the purchase), and connection to the surrounding context sentences. To extract the tips, we propose a supervised approach and provide a labeled dataset, annotated by human editors, over 14,000 product reviews using a dedicated tool. To demonstrate the potential of our approach, we compare different tip generation methods and evaluate them both manually and over the labeled set. Our approach demonstrates especially high performance for popular products in the Baby, Home Improvement and Sports & Outdoors domains, with precision of over 95% for the top 3 tips per product.

ACM Reference Format:
Sharon Hirsch, Slava Novgorodov, Ido Guy, and Alexander Nus. 2021. Generating Tips from Product Reviews. In *Proceedings of the Fourteenth ACM International Conference on Web Search and Data Mining (WSDM '21), March 8–12, 2021, Virtual Event, Israel.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3437963.3441755

1 INTRODUCTION

The importance of product reviews for many e-commerce platforms has been proven empirically across different shopping domains [8, 11, 25, 42]. Recent studies have demonstrated that over 85% of the customers often read product reviews before making a purchase and trust them as much as personal recommendations [5]. Online shoppers read reviews for various reasons, such as seeking for other customers' opinions, looking to read about personal experiences, or obtaining buyers' point of view on product characteristics. In some cases, customers also read reviews to find *tips* - short, concise, practical and self-contained pieces of advice. Tips can provide complementary insights on top of the existing product information, such as title, attributes, and description. They can be useful both before the purchase, to learn more about the product, and after the purchase, when the product is already at hand. Each of these use cases holds its own value for e-commerce platforms: before the purchase, tips help customers make a more informed purchase decision, whereas after the purchase, tips can motivate customers to return to the site and increase their engagement.

The large number of reviews on e-commerce platforms, especially for popular products[1], makes the process of finding useful information challenging. In order to find relevant pieces of information, customers usually sort and filter the reviews by different parameters, such as date or review score, but eventually they often consume few of the reviews, and might therefore overlook many helpful pieces of advice. Tips are not typically enabled as first-class user-generated content type on major e-commerce websites. Therefore, customers have to provide their tips and advice (if they wish to) through other user-generated content options, such as reviews.

In this work, we propose an automatic method for generating such short and concise tips from customer reviews. We propose an extractive approach, where we are aiming to find several tips out of hundreds of review sentences per product. Extracting only few, yet informative and helpful sentences from a large number of reviews, can save a lot of effort to customers and can come in especially handy for mobile device users, who often seek for concise content.

Tip extraction methods have been previously studied in other domains, especially travel [15, 37, 43]. However, these tips are different in their applicability. For example, travel tips mainly focus on logistics, opening hours, discounts, or special attractions to notice, while our tips focus on product aspects, such as usage and workarounds. To the best of our knowledge, we are the first to study product tips. As part of our study, we provide both analysis of tips "hidden" in reviews and an evaluation of methods for extracting them that attain high precision for popular products.

Our work uses a publicly available dataset of product reviews from one of the world's largest e-commerce platforms [16]. As the proposed approach is supervised, we extended the existing dataset with labels. First, we define a tip as a short, concise and self-contained piece of advice, in line with previous work in other domains [15, 37]. The data labeling is performed manually by human annotators via a dedicated tool designed for this task. We identify

WSDM '21, March 8–12, 2021, Virtual Event, Israel
© 2021 Association for Computing Machinery.
ACM ISBN 978-1-4503-8297-7/21/03…$15.00
https://doi.org/10.1145/3437963.3441755

[1] For instance, SENSO Bluetooth Headphones has over 36,000 reviews on Amazon.com

10 main types of product tips that commonly appear in reviews and list them in the annotation tool. Additionally, the annotator has to select the tip's timing (i.e., if the tip is useful before and/or after the purchase), and connection to the surrounding context sentences (i.e., whether they could or need to be used as part of the tip). The additional labeled set contains 14,000 product reviews and is released for public use as part of this work. Specifically, we focus on five popular e-commerce domains: Baby, Home Improvement, Musical Instruments, Sports & Outdoors, and Toys.

As mentioned above, we apply a supervised approach and experiment with a wide range of well-known classifiers, from baselines such as Naïve Bayes and basic LSTM [17], to state-of-the-art approaches, such as BERT [10]. In addition, we use a baseline method of taking sentences starting with a verb, which was applied in previous work and found useful for different types of tips [37]. We aim at extracting a small number of high-quality tips per product, since the presentation area on product pages is usually very limited, especially on mobile devices [30].

For our evaluation, we use two main methods. First, we perform a standard multiple train/test evaluation via random re-sampling on the collected labeled data and report the precision/recall for each of the classes. In addition, in order to simulate the practical use-case of extracting the tips from a large set of reviews, we run our method on previously-unseen products from these domains. We then take the top-k tip sentences identified by the model (ordered by classification score), and ask our annotators to manually evaluate the generated tips. The second evaluation method allows us to estimate the quality of our algorithm in a real-life scenario, and gain initial insights about the number of reviews needed to produce high-quality tips for a product. The results of the second evaluation demonstrated high precision, especially for the Baby, Home Improvement and Sports & Outdoors domains, with over 90% precision for the top 5 tips.

Our contributions can be summarized as follows:

- To the best of our knowledge, we are the first to introduce and study the tip extraction task in electronic commerce.
- We provide an extensive analysis of tips, their types, and distribution in reviews across different e-commerce domains.
- We present several supervised methods for detecting the tips and perform an extensive evaluation.
- We release our annotated data for public use as an extension to a popular e-commerce dataset [16]. The data includes 14,000 product reviews with over 85,000 labeled sentences.

2 RELATED WORK

Previous work has shown that online reviews from customers have a strong effect on other customers' purchase decision process in e-commerce [8, 11, 25, 42]. The sharp increase in the number and variety of reviews brings new challenges to the table, such as review quality estimation [7, 20] and fabricated review detection [1, 19]. A number of studies have shown that information overload, due to the immense number of reviews, leads to an increase in the time required to make a decision and degrades decision quality [34, 35]. There are several proposed approaches to deal with this challenge. Some focus on selecting a compact and representative subset of reviews (e.g., [14, 21, 22, 29]), while others apply review summarization techniques and generate an aggregate statistics of negative and positive feedback about different product features (e.g., [9, 18, 33]). Another related research direction deals with ranking the reviews according to different properties (e.g., helpfulness votes) [2, 36]. Finally, a recent work [30] used product reviews as a source of generating short product descriptions. In contrast to most of the approaches mentioned above, our method's building block is a review sentence rather than the entire review. Previous approaches that worked on a sentence level fundamentally differ from our approach. On the one hand, we do not aim to cover all possible aspects contained in the reviews. On the other hand, the summaries and/or the descriptions generated by the above-mentioned methods do not necessarily contain any tips.

A few past studies examined the identification of tips in domains other than e-commerce. Weber et al. [37] aimed at extracting tips from Yahoo Answers to address specific search queries. Similar to this work, they defined a tip as a "short, concrete and self-contained bits of non-obvious advice". Their proposed extraction mechanism mainly used the question-answer structure. Specifically, they considered only "how-to" questions and collected short answers that start with a verb. The final tips were always of the form "X:Y", where X is the tip's goal, and Y is the suggestion. This approach is not applicable in our setting, since we are working with product reviews rather than question-answer pairs and search queries. Nonetheless, we consider sentences that start with a verb as a baseline to all other methods.

Li et al. [23, 24] study abstractive tip generation. While the task seems to resemble our work, their definition of tips is very different from ours. They used two datasets from the e-commerce and restaurants domains. The first dataset was from Amazon, where the extracted tips originated from the "summary" part of the review, for example "*One of our favorite games!*" or "*My son really loves this simple toy*". The second dataset was from the Yelp Challenge and included restaurant tips and reviews, for example "*Love their soup!*" or "*Pretty good local service*". These "tips" do not provide much non-trivial information or insights, but rather reflect an opinion summary.

Travel is the most popular domain for tip-related research, mostly because there are many available datasets (e.g., forums, blogs, questions and answers) and since the user is typically visiting an unfamiliar environment, where advice from knowledgeable individuals can be valuable. In contrast to tips in e-commerce, which are mostly about different usages of the products, travel tips focus mainly on logistics, opening hours, discounts, special attractions to be noticed, and so forth. Closest to our work is the research by Guy et al. [15] and by Zhu et. al [43]. The work by Guy et al. relies on 150 human-generated templates for travel tips. Examples of such templates are "make sure to *", "check the * for" and "the * is closed on mondays", where the asterisk can represent any word. The work by Zhu et al. extends the work by Guy et al. and introduces an unsupervised approach that solves a similar task without relying on training data. The key difference from our work is the applicability of the proposed methods to the e-commerce domain. In this paper, we define the tip extraction task along with e-commerce specific tip types and their context, while in their works they focus on travel-specific language. Moreover, a template-based approach is not applicable in

Table 1: Characteristics of the original datasets.

	Baby				Home Improvement				Musical Instruments				Sports & Outdoors				Toys			
	Avg	Std	Median	Max	Avg	Std	Median	Max	Avg	Std	Median	Max	Avg	Std	Median	Max	Avg	Std	Median	Max
Reviews per product	22.81	36.76	11	780	13.16	16.22	8	504	11.40	12.93	8	163	16.14	25.67	9	1042	14.06	15.85	9	309
Sentences per review	6.14	5.62	5	213	6.89	7.21	5	198	5.81	5.93	4	116	5.71	5.94	4	283	6.40	6.22	5	222
Words per sentence	16.19	10.70	14	425	16.08	10.54	14	573	15.66	10.99	14	230	15.39	10.53	13	829	15.74	10.25	14	586
Number of products	7,050				10,217				900				18,357				11,924			
Number of reviews	160,792				134,476				10,261				296,337				167,597			

our setting, since we did not find any dominant repetitive n-gram patterns in our annotated tips.

Another closely connected field of research focuses on detecting text units that include pieces of advice. Wicaksono and Myaeng [39] proposed to use conditional random fields, to extract advice sentences from travel forum entries. An earlier work by the same authors [38] focused on finding advice sentences in travel blogs. They proposed several linguistic features, mostly defined by handcrafted rules that were looking for the appearance of terms such as “I suggest”, “I strongly recommend”, or “advice”, with an associated proper noun, representing a travel entity, such as a hotel. Our approach applies a preliminary rule-based step, to filter out sentences with very low likelihood of being tips. However, rules do not suffice in our case due to the scarcity of repetitive patterns in e-commerce tips. We therefore propose a supervised model as our main method for tip extraction.

3 DATASETS AND CHARACTERISTICS

In this section, we describe the datasets used for our analysis and experimental evaluation, their characteristics, and the annotation process we used in order to produce labeled data.

3.1 Datasets

Our research was conducted over five publicly available product datasets [16] from five e-commerce domains: Baby (baby clothing and supplementary products), Home Improvement (tools for home improvement), Musical Instruments (musical instruments, parts, and related accessories), Sports & Outdoors (equipment for sports and outdoor activities), and Toys (children’s toys and games). The datasets contain, per each product, its metadata (title, image, etc.) and all its associated user reviews. Table 1 depicts the main characteristics of the five datasets. The largest dataset is Sports & Outdoors containing 18,357 products with nearly 300K reviews in total, while the smallest is Musical Instruments, with 900 products and a little over 10K reviews in total. The median number of reviews per product ranges from 8 to 11, while the median number of sentences per review is between 4 and 5.

3.2 Tip Definition

We define a *tip* as a short, concise, practical, and self-contained piece of advice. In general, tips can be useful both before and after the purchase. Before-purchase tips are useful to learn more about the product, and after-purchase tips are helpful when the product is already in hand. Both of these use cases are important for e-commerce platforms: tips before the purchase help with the purchase decision by providing more information. Useful tips after purchases can increase customer satisfaction and motivate return to the site for additional shopping.

Despite the straightforward definition of a tip, there are some borderline cases that should be discussed. First, many review sentences may look like a tip, while they are very subjective and contain a personal experience (e.g., “*In addition, we live in a colder climate and do not heat the house above 62 degrees at night, so combined with a little heater, this sleep sack does the trick.*”) Such sentences are not considered as tips. Another type of a borderline case not considered as tip is a descriptive sentence (e.g., “*The cards are high gloss with full color pictures on them.*”). Other non-tip sentences are obvious or trivial remarks (e.g., “*Just make sure you have a bunch of batteries to get started.*”) or those that repeat details already provided as part of the product description, without adding any new information (e.g., “*Perfect for tuning electric guitars.*” for a BROTOU Guitar Tuner that includes the following sentence in its description: “Fits most electric guitars”).

3.3 Data Annotation

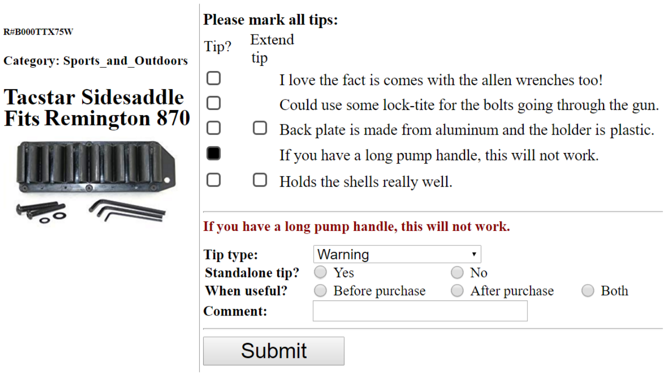
Labeling for training and evaluation in this work was performed by in-house annotators, after three hours of training and qualification tests. The pool included a total of 10 annotators, who were assigned tasks from each of the five domains randomly.[2] Unless otherwise stated, each evaluation was performed by a single annotator. Labeling was performed using a dedicated tool developed for this task. The tool’s user interface is depicted in Figure 1. As shown in the figure, each review was split into sentences. For each sentence, the annotator was asked to select whether it is a tip (as defined in Section 3.2) or not. For each selected tip, the annotator was asked to select the tip type (out of 10 pre-defined types, presented at the first column of Table 3). Afterwards, annotators had to indicate if the sentence is a standalone tip or needs additional context. Moreover, annotators selected if this tip is useful before the purchase, after the purchase, or both. They could also (optionally) choose the previous and/or next sentence as useful information for extending the selected tip, regardless if it was marked as a standalone or not (see “extend tip” checkbox in the second column in the annotation tool screenshot). To measure the agreement between the annotators, we asked four of them to annotate the same 100 review sentences as tips or not. The Fleiss' Kappa [12] among them was 0.815, indicating a high agreement level. All data annotated using the tool will be publicly released as an extension to the original public dataset.[3]

[2] Annotators were granted monetary compensation for their work.

[3] The dataset is available at http://proj.ise.bgu.ac.il/public/gen_tips.zip

Table 2: Characteristics of labeled tips across the five domains.

	Baby	Home Improvement	Musical Instruments	Sports & Outdoors	Toys
# of products	2,612	2,711	736	2,722	2,736
# of reviews	2,800	2,800	2,800	2,800	2,800
# of sentences	17,560	19,436	15,460	15,957	16,758
# of tips (% of sentences)	954 (5.43%)	880 (4.53%)	537 (3.47%)	805 (4.71%)	670 (4.00%)
Avg (median) words per tip sentence	21.33 (19)	21.11 (19)	21.98 (19)	20.42 (19)	20.38 (18)
Before purchase	49.37%	35.34%	39.66%	39.13%	47.01%
After purchase	21.07%	37.27%	27.75%	27.83%	23.88%
Both	29.56%	27.39%	32.59%	33.04%	29.10%
Standalone	81.97%	79.20%	80.07%	84.10%	84.48%
Extend before tip	9.12%	9.31%	9.57%	11.25%	9.55%
Extend after tip	17.92%	16.20%	15.16%	19.43%	15.97%
Most common types	Warning (38.05%) Usage (23.69%) Size (8.07%)	Usage (40.23%) Warning (29.55%) Workaround (7.05%)	Usage (40.41%) Warning (30.17%) Workaround (8.38%)	Warning (29.94%) Usage (29.81%) Size (9.81%)	Warning (34.63%) Usage (23.88%) Population segment (13.13%)



R#B000TTX75W

Category: Sports_and_Outdoors

Tacstar Sidesaddle Fits Remington 870

Please mark all tips:

Tip?	Extend tip	
□		I love the fact is comes with the allen wrenches too!
□		Could use some lock-tite for the bolts going through the gun.
□	□	Back plate is made from aluminum and the holder is plastic.
■		If you have a long pump handle, this will not work.
□	□	Holds the shells really well.

If you have a long pump handle, this will not work.

Tip type: Warning ▾
Standalone tip? ○ Yes ○ No
When useful? ○ Before purchase ○ After purchase ○ Both
Comment:

Figure 1: Annotation interface.

Table 3: Types of tips, their distribution (portion of all sentences marked as tips), and examples across all domains.

Warning	32.71%	The metal mounting clips scratched the edges of my trunk lid.
Usage	31.12%	Best used when replacing strings, so you can apply while they are off.
Workaround	6.66%	I needed a 5/8 female to 3/8 male adapter to get my mic to mount.
Complementary product	5.54%	You will need to buy fasteners for it, since the box only contains the vice.
Size	5.49%	But definitely order at least one size bigger than you wear.
Maintenance	4.60%	The slipcovers come off easily to be machine-washed.
Population segment	4.24%	Recommend for a 4-5 year old that likes cars and trucks.
First time use	3.98%	The wheels do need to be pumped with a bike pump prior to use.
Alternative use	2.99%	My baby doesn't need it anymore so now I use it as my neck pillow.
Other	2.68%	Her hair is much brighter blue than it appears in the photo.

3.4 Tip Characteristics

We sampled uniformly at random 14,000 product reviews across the five domains (2,800 per domain) from the original datasets (Table 1). The annotators labeled these 14,000 reviews, which included 85,171 sentences in total. Table 2 depicts the full statistics of the annotated dataset, including the number of labeled sentences and the collected tips across the five domains, along with the most common tip types. Overall, 3,846 sentences were annotated as tips, accounting for only 4.52% of all labeled sentences. This is a substantially lower percentage than the 23.3% reported for reviews of tourist attractions [15], indicating that tips are scarcer in product reviews than in travel reviews. Table 3 depicts the distribution of the 10 different tip types in our labeled tips. The most popular tip types were 'Warning' and 'Usage', accounting each for slightly over 30% of all tips. The third most popular type was 'Workaround', followed by 'Complementary product' and 'Size'. The least popular tip type was 'Other', while about half of these tips related to differences between the actual product received and the seller provided information (product image, title, or description; see example in Table 3). As depicted at the bottom of Table 2, some variance can be observed for the distribution of top tip types across the five domains. 'Usage' and 'Warning' are at the top of the list in each of the five domains, with 'Usage' the most common for Home Improvement and Musical Instruments and 'Warning' for Baby, Sports & Outdoors, and Toys.

Another interesting characteristic is the connection to the surrounding context sentences. Most of the tips (81.96%) were standalone, while only 18.04% were labeled as non-standalone sentences. As can be seen in Table 2, these results are rather consistent across the five domains. Overall, 26.91% of the tips could be extended to the adjacent sentence, with about two thirds of these to the succeeding sentence and a third to the preceding sentence. For example, for the standalone tip sentence: "*One note of caution, this is a very heavy router because it is a large plunge router*", the succeeding sentence was marked as an extension: "*I mounted it on a Rockler X-Large router plate which is 1/4 inch thick aluminum, but it has a very slight bow in the middle.*" For the non-standalone tip sentence: "*You need a pipe cleaner to really get it*", the preceding sentence was annotated as an extension: "*My one complaint is that there is an area that is hard to clean in the 4 piece nipple apparatus.*" Overall, however, the low portions of non-standalone tips indicates that our choice to focus on single-sentence tips covers the majority of the cases. We leave the expansion to multi-sentence tips for future work.

Table 4: Tip timing (before and/or after purchase) distribution by type.

Type	Before purchase	After Purchase	Both
Warning	76.71%	13.04%	10.25%
Usage	17.88%	45.53%	36.59%
Workaround	3.91%	54.30%	41.80%
Complementary product	38.97%	11.74%	49.30%
Size	62.56%	1.42%	36.02%
Maintenance	3.95%	45.20%	50.85%
Population segment	78.53%	1.23%	20.25%
First time use	22.22%	49.67%	28.10%
Alternative use	6.09%	20.87%	73.04%
Other	43.69%	3.88%	52.43%

The number of tips marked as being useful before the purchase was somewhat higher than those marked as useful after the purchase: 42.25% versus 27.61%, respectively. The rest, nearly a third (30.14%), were annotated as useful both before and after the purchase. These portions varied substantially across the different tip types, as depicted in Table 4. While 'Population segment', 'Warning', and 'Size' are typically useful before the purchase, 'Workaround', 'First time use', 'Usage', and 'Maintenance' tips are more often useful after the purchase. 'Alternative use' tips are prominently useful both before and after the purchase. We conjecture that alternate-use tips can both influence the purchase decision, as they reveal additional functionalities, and are also handy when the product is in possession, extending its potential use. The substantial differences across tip types are also reflected in differences across the domains, as can be seen in Table 2. For instance, domains with higher portion of 'Warning' tips, exhibit higher portions of tips that are useful before the purchase.

Tip Analysis. As a first step after obtaining the labeled data, we analyzed additional tip features. Table 5 presents the portion of tips according to different characteristics of the sentence in question, the originating review, and its authoring reviewer. It can be seen that longer sentences have higher likelihood of being tips: of the sentences consisting of 30 words or more, 8.6% were marked as tips, accounting for 23.1% of the tips in our dataset. At the other extreme, only 1.9% of the sentences composed of 6 to 9 words were labeled as tips. Inspecting review characteristics, it can be seen that sentences that originate from short (1-2 sentences) and especially long (over 15 sentences) reviews have somewhat lower likelihood of being tips. As can also be observed from the table, sentences that originate from reviews with two or more 'helpful' votes were more likely to be considered as tips, whereas opening sentences (positioned first) and those originating from reviews with no 'helpful' votes had lower likelihood to be tips. Finally, we inspected characteristics of the reviewer who wrote the originating review. Interestingly, reviewers with many reviews on the site (130 and more) tend to include fewer tip sentences in their review. It could be that such "heavy" reviewers focus on other aspects in their reviews, such as personal experiences and opinions. Additionally, reviewers who had especially lower portion of past reviews with at least one helpful vote, produced a lower portion of tips, as can be seen in the last section of Table 5. We note that the results for each of the five domains demonstrated similar trends to those shown in Table 5.

Table 5: Tip characteristics (binned) within the complete dataset (all five domains). '%' and '%T' denote, per bin, the portion of tips out of all tips and the portion of sentences marked as tips, respectively.

	Sentence Length (Number of Words)						
0	6-9	10-13	14-17	18-21	22-25	26-29	30+
%	6.6	12.4	16.1	16.8	13.9	11.1	23.1
%T	1.9	3.0	4.3	5.5	6.4	7.5	8.6
	Review Length (Number of Sentences)						
	1-2	3	4	5-6	7-9	10-15	16+
%	5.8	8.8	10.6	17.6	17.6	20.0	19.8
%T	4.5	4.3	4.9	4.9	4.6	4.7	3.8
	Tip Position within Review						
	First	Middle	Last				
%	10.3	74.6	15.1				
%T	2.9	5.0	4.2				
	Review's Number of Helpful Votes						
	0	1	2+				
%	46.2	16.1	37.7				
%T	4.2	4.4	5.1				
	Reviewer's Number of Past Reviews						
	0-19	20-39	40-79	80-129	130-199	200+	
%	13.3	21.0	24.1	12.7	8.6	20.3	
%T	4.8	4.8	4.5	4.2	4.4	4.1	
	Reviewer's Portion of Past Reviews with Helpful Votes						
	0-20	20-30	30-40	40-50	50-60	60-100	
%	9.8	15.5	18.2	17.0	16.4	23.0	
%T	3.9	4.3	4.7	4.7	4.8	4.5	

Tip vs. Non-tip Language. We also set out to examine the most prominent language differences between the two classes of sentences – tips versus non-tips. To this end, we used Kullback-Leibler (KL) divergence, which is a non-symmetric distance measure between two given distributions [3]. Specifically, we calculated the terms that contribute the most to the KL divergence between the language model of the tip sentences versus the language model of the non-tip sentences and vice versa [6].

Table 6 presents the most distinctive unigrams and bigrams. It is noticeable that the most characterizing unigram of tip sentences compared to non-tip sentences is the second-person pronoun `you`, while the first-person pronoun `i` tops the non-tip list. This indicates that tip sentences are usually phrased in a second-person language rather than first. The unigram `my`, which also reflects a first-person language is also high on the non-tip list, as well as the plural `we`, while `your` is on the tip list. Other unigrams on the tip list include different prepositions and verbs, such as `use` and `need`, as well as the explicit `suggest` and `note`. The non-tip list includes `was` and `had`, which are often used to describe a past experience, the verbs `like` and `love`, which reflect subjective opinions, as well adjectives and adverbs that reflect positive impressions, such as `great`, `very`, and `nice` and the noun `quality`, which is often associated with the reviewer's opinion on a certain feature of the product.

Inspecting the bigram lists, the tip list includes several expressions in second-person languages, such as `if you`, `you can`, `you have`, `you need` and `you use` and expressions typical to advice giving, such as `make sure`, `will need`, `be careful`. The non-tip bigram list, on the other hand, includes many first-person expressions, such as `i have`,

Table 6: Most distinctive unigrams and bigrams for tips vs. non-tip sentences.

Unigrams		Bigrams	
Tips	Non-tips	Tips	Non-tips
you	i	if you	i have
the	this	on the	the price
to	my	you have	i am
if	price	need to	i was
your	love	make sure	a great
or	our	you can	my son
use	his	sure you	a very
need	we	when you	i had
can	great	use the	in my
suggest	he	will need	i love
not	nice	have to	i bought
head	bought	that you	for my
put	like	you use	i got
note	very	you need	bought this
make	quality	be careful	i will

and `i`, `i am`, `i got`, and `i was`, in addition to expressions that reflect personal experiences, such as `my son` or `i bought`. It also includes `the price`, referring to a listing-specific characteristic, which may change from one seller to another, and phrases that reflect subjective feelings, such as `a great`, `a very`, and `i love`.

4 TIP EXTRACTION

In this section, we describe the key components of our tip generation method. Given the set of all user reviews for a product, we go through the following steps to extract the tip sentences. We first apply rule-based filtering crafted based on analysis of the datasets. Following, we use a supervised approach that learns to identify tip sentences. To this end, we experiment with different types of classifiers, including state-of-the-art methods in language modeling, and compare their performance on a labeled set.

4.1 Rule-based Filtering

Since the data is very skewed, before applying any advanced classification approaches, we look for methods that would easily filter out non-tip sentences. After performing an analysis of basic features (e.g., sentence length) and words (KL) we could not observe simple rules that can massively be used for filtering, as done for the task of description generation [30]. Nevertheless, we identify a few rules that could save up to 39% of the labeling task. The proposed rules decrease the total number of sentences from 85,171 to 51,929 and increase the total percentage of tips from 4.52% (as reported in Section 3.4) to 5.89%. To derive many of these rules, we observed the top KL n-grams for $n \in \{1, 2, 3\}$ and considered those that hardly appear in tips, i.e., in fewer than 5 sentences in our training set, but do appear frequently in non-tip sentences. Our rule list includes the following:

(1) **Short**: sentences of 5 words or fewer generally contain little information and rarely reflect any useful piece of advice. For example, "*Recommended*", "*Very good quality*", "*Useful*", or "*Will buy again*" were among the most common short review sentences in our datasets. Analogously, in the travel domain it has been previously demonstrated that short sentences cannot serve as high-quality tips [15]. Overall, short sentences accounted for 11.2% of all review sentences across all domains in our dataset.

(2) **Enthusiastic**: some reviewers tend to describe the product using strong-sentiment positive adjectives, such as 'wonderful', 'adorable', 'amazing', 'fantastic' and verbs such as 'love' and 'like'. Examples include "*I love this product and recommend it for everyone*" and "*Amazing quality, very useful for outdoor activities*". Overall, 18.6% of the review sentences matched this filtering criterion.

(3) **Listing-specific**: review sentences that focus on listing-specific aspects, which may vary across different sellers of the product, such as price, shipping and return policy, warranty, and similar. Tokens used for filtering included 'price', 'money', 'cheap', 'expensive', 'shipping', 'return', 'warranty' and also the dollar sign '$'. Examples of such sentences include "*Very useful product for just 20$*" and "*The shipping was almost as much as the panel itself*". Overall, 14.7% of the review sentences matched this filtering criterion.

(4) **Personal**: sentences with a first-person pronoun, such as 'i'm', 'i'll' and 'i've'. As demonstrated in the previous section, such pronouns hardly ever occur on a product tip. Overall, 8.4% of the review sentences matched this filtering criterion.

We do not automatically filter out sentences from reviews with low ratings and do not run any sentiment analysis to filter out negative sentences, since these may hide useful tips, such as warnings, workarounds, or alternate use. For example, the sentence "*Tried using it lighting with a match which worked, but the off switch would not shut the flow of fuel off completely*" is a workaround tip for a Mini Jet Pencil Lighter and *"They fog up almost as quickly as my old Speedo goggles (several years old) which is after about one lap of the pool"* is a warning tip for a Speedo Baja Swim Goggle. Both tips were extracted from negative (one and two stars) reviews.

Our rules are designed to filter out sentences that are very likely not to contain a tip, hence we prefer to apply only high-precision rules. As already mentioned, our rules filtered out 39% of the review sentences, leading to a tip portion of 5.89% in the remaining set.[4]

4.2 Automatic Classification

After applying the initial rule-based filtering, we explore a supervised approach, by training a classifier to predict whether a product review sentence contains a tip. We use the labeled dataset described in Table 2 and experiment with various classifiers:

Naïve Bayes. We examine a common model for text classification - Naïve Bayes [32]. Our features include textual features, specifically the unigrams, bigrams, and trigrams of the review sentence. We also experiment with a variant that includes additional features, based on the characteristics described in Table 5.

LSTM. A recurrent neural network based on a long short-term memory (LSTM) [17] architecture, with Global Vectors for word representation (GloVe) [31] pre-trained on the Wikipedia 2014 and Gigaword 5 corpora.

LSTM with Attention. The attention mechanism enables the network to focus on relevant parts of the input [41]. The overall architecture of the "attention network" consists of two components: an LSTM-based word sequence encoder and a word-level attention layer. Given a review sentence split by words, we first embed each word using pre-trained GloVe embeddings, as previously described, and then use the LSTM network to produce the hidden states. The

[4] The portions of all four rules do not sum up to the total portion of filtered sentences, since some sentences match more than one rule.

Table 7: Recall results for classifying review sentences as tips at four different precision levels: 75%, 80%, 85%, and 90%.

Classifier	Recall@Precision=			
	75%	80%	85%	90%
Naïve Bayes	43.42%	26.71%	11.37%	6.18%
Naïve Bayes w/Features	40.71%	26.17%	12.07%	6.78%
LSTM	30.03%	25.25%	16.79%	15.30%
LSTM w/Attention	30.41%	22.40%	15.62%	12.38%
FastText	41.92%	29.01%	15.49%	8.88%
BERT	**70.47%**	**58.05%**	**36.05%**	**19.33%**

Table 8: Reviews per product by percentile and average.

Domain	10th	30th	50th	70th	90th	Average
Baby	5	6	9	13	29	22.81
Home Improvement	5	7	11	19	49	13.16
Musical Instruments	5	6	9	14	31	11.40
Sports & Outdoors	5	6	8	12	25	16.14
Toys	5	6	8	10	20	14.06

attention mechanism is often used to put more focus on certain words in the review sentence. For example, in the sentence "*Make sure to switch off the guitar*" the words "*make sure*" receive higher weight, and in the sentence "*Just be careful when opening the hood*", the higher weight is given to "*just be careful*". We feed the word annotations through a single-layer perceptron network to receive a latent representation. Then, we calculate the similarity of the latent representation with a word-level context vector, normalized by a softmax function, to produce the word's importance weight. We then construct the sentence vector as a weighted sum of the word annotations based on each word's weight.

For both LSTM methods, we performed hyper-parameter tuning, which included the batch size, number of epochs, learning rate, and the number of hidden units in the layers.

FastText. A library for learning word embeddings and text classification created by Facebook's AI Research called FastText [4]. Each word is represented as a bag of character n-grams, and the final word embedding is the sum of character n-grams. This is useful for generalizing words with similar roots that appear in different forms (e.g., `build` and `building`). Hyper-parameter tuning was performed on the n-grams length, learning rate, and number of epochs.

BERT. A state-of-the-art technique for NLP pre-training called Bidirectional Encoder Representations from Transformers (BERT) [10]. This is a deep bidirectional, unsupervised language representation, pre-trained using only a plain text corpus. In contrast to context-free models, such as word2vec [28] or GloVe [31], which generate a single word embedding representation for each word in the vocabulary, BERT generates a representation of each word based on the other words in the sentence. BERT is useful for extracting high-quality language features from text data. In addition, it is useful for fine-tuning a model on a specific task, as we did in our experiments. We used the pre-trained 'BERT-Base, Uncased'[5] model to train a binary classifier for our task. Our hyper-parameter tuning included the batch size, number of epochs, and learning rate.

Table 9: Precision of top-k predicted sentences by BERT.

Domain	Top 1	Top 3	Top 5
Baby	98.00%	96.00%	97.20%
Home Improvement	100.00%	96.00%	92.80%
Musical Instruments	90.00%	84.33%	80.40%
Sports & Outdoors	98.00%	95.33%	94.00%
Toys	94.00%	90.67%	89.20%

5 EVALUATION

In the following section, we present two evaluation methods for the extraction models. The first method is a standard train/test evaluation repeated 50 times, where each time the training and the test sets are randomly re-sampled from the labeled data [40]. The second method simulates the practical scenario of extracting tips from a large set of reviews of previously-unseen products. We use the best performing model from the first evaluation method and apply post-processing human annotations for evaluation.

We compare the classifiers presented in Section 4.2, namely Naïve Bayes (including a variant with additional features), LSTM (including a variant with attention), FastText, and BERT. We use our labeled dataset containing a total of 3,059 tips and 48,870 non-tip sentences, obtained after the initial rule-based filtering described in Section 4.1. In addition, we use a simple baseline method that labels all sentences starting with verbs (VB or VBP part-of-speech tags [27]) as tips. This method has been used in key tip studies in domains other than e-commerce, either as the main approach [37] or as a baseline [15]. It yields low precision of 17.72% and low recall of 4.48% on our dataset, indicating that the task is far from trivial. This observation is aligned with previous findings in the travel domain, where 9.8% of the tips were reported to start with a verb [15]. As previously mentioned, the template-based technique proposed for travel tip extraction [15] is not applicable for our setting, since we could not detect dominant repetitive n-gram patterns in our data. The approach proposed by [37] is also not applicable, as we are working with product reviews rather than question-answer pairs.

We evaluate the algorithms over 6,118 sentences (a balanced set of 3,059 tips and 3,059 random non-tip sentences) and divide the labeled data to a 80% train and 20% test. We perform this evaluation 50 times, each with another random set of $3,059$ sentences from the 48,870 labeled non-tips and randomly split the 6,118 sentences to train and test sets. As previously mentioned, our goal is to produce a small number of high-quality tips, especially suitable for presentation on mobile devices, with limited screen space. Specifically, we aim to produce up to 5 tips per product with high precision, even at the cost of compromising some of the recall. Table 7 depicts the average recall results of the different classifiers across the 50 train-test samples, while we set the desired precision threshold to 75%, 80%, 85%, and 90%. Note that when presenting 5 tips, 80% precision yields an average of one wrong tip per product.

As can be seen, BERT outperforms its competitors at all precision levels and achieves the best results, with 58.05% recall for 80% precision. It can also be observed that the addition of the non-textual features described in Table 5 to the Naïve Bayes classifier does not show a substantial performance gain; we therefore did not work with these features in the remainder of our experiments. The

[5] https://github.com/google-research/bert

Table 10: Examples of automatically-extracted tips.

Domain	Product	Tip	Tip Type
Baby	SoundSpa On-The-Go White Noise Machine	Be prepared to purchase a battery charger for AAA NiMH batteries if you want to run this thing every night.	Complementary product
Baby	HALO Early Walker Sleepsack Wearable Blanket	However, you need to keep in mind that the feet at the bottom are not designed for running around playing really but more for sleeping in.	Usage
Home Improvement	PORTER-CABLE 7-Amp Plate Joiner Kit	If you keep your fingers above and on the opposite side of the board, there is no danger.	Warning
Home Improvement	8-LED Motion-sensing Night Light	If installing flat (like on a ceiling or under a shelf) the adhesive tape won't hold the weight of the unit with batteries .. you will have to use the mounting screws.	Workaround
Musical Instruments	Dunlop Acoustic Trigger Gold Guitar Capo	Be careful about using capos like this, because if you're using a fine guitar, it may damage the finish.	Maintenance
Musical Instruments	String Swing Metal Home & Studio Wide Guitar Hanger	The yoke width is adjustable, and a combination of slope and two keeper rings prevent the instrument from coming off the holder.	Usage
Sports & Outdoors	Invicta Men's Pro Diver Stainless Steel Watch	The directions that come with the watch are not very helpful and do not indicate that you have to unscrew the stem in a counterclockwise direction to pull it out to set the time and date.	First time use
Sports & Outdoors	Park Tool CT-5 Mini Chain Brute Bicycle Chain Tool	Be sure to read the directions as placing the chain in the wrong slot to break it or re-unite the chain can bend the links out of shape.	Warning
Toys	Pretend & Play Teaching Cash Register	Take out one or two of the screws that hold that transparent piece of plastic to the top of the cash drawer.	First time use
Toys	Melissa & Doug Shapes Chunky Puzzle	They could also be used for tracing with paper and a crayon/marker.	Alternative use

best performing configuration for the BERT classifier was with a batch size of 8, 5 epochs, and learning rate of 0.0002.

5.1 Evaluation over Unseen Products

In order to simulate the practical use case of extracting the tips from large sets of reviews, we use the following evaluation method. We run the BERT model on previously-unseen products from the five domains, rank the tips by the model's score, and select the top 5 tips. Then, we ask our in-house annotators to manually evaluate the generated tips. We specifically focus on popular products with many reviews, as for these we believe our method can work well even with low recall (as we mentioned before, we do not want to compromise precision). We check the quality for the scenario where we present to the user only a small number of tips: one, three, and five. This reflects the business need of extracting only few, but high-quality tips, which can fit within a limited user interface space on the product page. This evaluation method also allows us to gain insights about the number of reviews required to produce high-quality tips per each product. We start by considering all products above the 90th percentile according to their number of reviews in each of the five domains (Table 8 presents the statistics of review number per product). Then, we randomly sample 50 products from each domain and apply the BERT classifier on all review sentences of these products. Finally, we present the top sentences (ordered by classification score) and ask the annotators to evaluate if they are tips. Each sentence is reviewed by two annotators and considered as a tip only if both agree on it. The results of the top 1, 3, and 5 sentences are depicted in Table 9. Inspecting precision@1 (i.e., the portion of sentences with highest classification score that are deemed as tips), Home Improvement demonstrates the highest performance with a perfect 100%, followed by Baby and Sports & Outdoors domains that attain a high 98%, and Toys with 94%. The lowest precision@1 is yielded for Musical Instruments, with 90%. Precision remains high in the Baby, Home Improvement and Sports & Outdoors domains for the top 3 and top 5. For Toys, it is down to around 90% for the top 3 and top 5. For Musical Instruments, the sharpest drop is recorded, down to around 84% for the top 3 and 80% for the top 5. These results show that our method can produce top tips at high precision: the precision@1, precision@3, and precision@5 across the five domains are 96%, 92.47%, and 90.75%, respectively. The Baby, Home Improvement, and Sports & Outdoors domains include relatively higher portion of tips within their reviews (Table 2) and yield the highest tip precision (Table 9), implying they are especially suitable for tip extraction. Table 10 presents examples of extracted tips, including the product's domain, title, and tip's type.

5.2 Limitations

The main limitation in the described approach is the low recall. In addition, tip sentences are scarce. Particularly, every 22 review sentences contain only one tip sentence, on average. Even after applying the rule-based filtering, a tip appears every 17 review sentences, or every 4-5 reviews on average. These two limitations make our approach applicable to products that have accumulated a large number of reviews. Having said that, popular products are broadly exposed, and so their effect on a large number of users can be high and henceforth pave the way for tips receiving more prominence on e-commerce platforms.

Another limitation is the potential repetition of tips extracted for a given product. As the number of desired tips per product grows, diversification should be applied to avoid redundant tips. For example, the following usage tips, "*To remove the marker designs from the screen you need to use a damp cloth then allow to completely dry - just a bit of a pain*" and "*To clean the screen off, you just run a damp paper towel or cloth over the screen*", were generated for the same Widescreen Light Designer. Semantic similarity can be used to cluster similar tips, so that the final list is diverse. The clusters' size can also assist in selecting the top tips to be presented. Such an approach was found productive in previous work on product description generation from reviews [30] and can be similarly applied in our case.

6 CONCLUSION AND FUTURE WORK

In this work, we propose a tip extraction method from product reviews. We focus on five domains that naturally contain useful and non-trivial tips across the reviews and are likely to be beneficial for potential customers. We formally define the task of tip extraction in e-commerce by providing the list of tip types, tip timing (before and/or after the purchase), and connection to the surrounding context sentences. We evaluate different approaches of supervised tip extraction that are trained on labeled data from 14,000 product reviews. Tips are labeled using a dedicated tool and released for public use, as part of a dataset's extension. The best performing method, BERT, achieves recall of 58.05% at 80% precision on a balanced test set. Moreover, when the method is applied to unseen products, the precision@1 is 90% for the lowest domain (Musical Instruments) and 100% for the highest (Home Improvement). Precision@5 is 80.4% for the lowest domain (Musical Instruments) and 97.2% for the highest (Baby). Our method is not specific to any of the five domains and can therefore be potentially applicable to other e-commerce areas.

For future work, we plan to focus on five main directions. Currently, we focus on extracting single-sentence tips, but as discussed in Section 3, over 25% of the tips can be extended to include adjacent sentences; hence, extending our approach to support multi-sentence tips is an intriguing direction. Second, we plan to explore abstractive approaches that combine content from different sentences and adapt them [13, 26]. This can help increase the number of extracted tips and also deal with sentences that contain irrelevant information in addition to the tips. Third, tip diversification is an important step in providing multiple useful and non-repetitive tips. Fourth, we plan to investigate how to elevate characteristics specific to different tip types in order to improve the overall tip extraction quality. Finally, presenting product tips on e-commerce platforms both before and after product purchases can serve to study their actual impact on user behavior. In-vivo experimentation can reveal the actual effect of tips on user interaction such as clicks, purchases, and long-term activity on the site.

REFERENCES

[1] Leman Akoglu, Rishi Chandy, and Christos Faloutsos. 2013. Opinion fraud detection in online reviews by network effects. In *Proc. of ICWSM*.

[2] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2009. Multi-facet rating of product reviews. In *ECIR*. Springer, 461–472.

[3] Adam Berger and John Lafferty. 1999. Information retrieval as statistical translation. In *Proc. of SIGIR*. 222–229.

[4] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *TACL* 5 (2017), 135–146.

[5] BrightLocal. 2018. Local Consumer Review Survey. (2018). https://www.brightlocal.com/research/local-consumer-review-survey/

[6] David Carmel, Erel Uziel, Ido Guy, Yosi Mass, and Haggai Roitman. 2012. Folksonomy-Based Term Extraction for Word Cloud Generation. *ACM Trans. Intell. Syst. Technol.* 3, 4, Article 60 (Sept. 2012), 20 pages.

[7] Chien Chin Chen and You-De Tseng. 2011. Quality evaluation of product reviews using an information quality framework. *Decision Support Systems* 50, 4 (2011), 755–768.

[8] Judith A Chevalier and Dina Mayzlin. 2006. The effect of word of mouth on sales: Online book reviews. *Journal of marketing research* 43, 3 (2006), 345–354.

[9] Paolo Cremonesi, Raffaele Facendola, Franca Garzotto, Matteo Guarnerio, Mattia Natali, and Roberto Pagano. 2014. Polarized review summarization as decision making tool. In *Proc. of AVI*. 355–356.

[10] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[11] Wenjing Duan, Bin Gu, and Andrew B Whinston. 2008. Do online reviews matter? An empirical investigation of panel data. *Decision support systems* 45, 4 (2008), 1007–1016.

[12] Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin* 76, 5 (1971), 378–382.

[13] Shen Gao, Xiuying Chen, Piji Li, Zhaochun Ren, Lidong Bing, Dongyan Zhao, and Rui Yan. 2019. Abstractive text summarization by incorporating reader comments. In *Proc. of the AAAI Conference*, Vol. 33. 6399–6406.

[14] Shima Gerani, Yashar Mehdad, Giuseppe Carenini, Raymond T. Ng, and Bita Nejat. 2014. Abstractive summarization of product reviews using discourse structure. In *Proc. of EMNLP*. 1602–1613.

[15] Ido Guy, Avihai Mejer, Alexander Nus, and Fiana Raiber. 2017. Extracting and ranking travel tips from user-generated reviews. In *Proc. of WWW*. 987–996.

[16] Ruining He and Julian McAuley. 2016. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *Proc. of WWW*. 507–517.

[17] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[18] Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proc. of KDD*. 168–177.

[19] Nan Hu, Indranil Bose, Noi Sian Koh, and Ling Liu. 2012. Manipulation of online reviews: An analysis of ratings, readability, and sentiments. *Decision support systems* 52, 3 (2012), 674–684.

[20] Soo-Min Kim, Patrick Pantel, Tim Chklovski, and Marco Pennacchiotti. 2006. Automatically assessing review helpfulness. In *Proc. of EMNLP*. 423–430.

[21] Theodoros Lappas, Mark Crovella, and Evimaria Terzi. 2012. Selecting a characteristic set of reviews. In *Proc. of KDD*. 832–840.

[22] Theodoros Lappas and Dimitrios Gunopulos. 2010. Efficient confident search in large review corpora. In *ECML PKDD*. Springer, 195–210.

[23] Piji Li, Zihao Wang, Lidong Bing, and Wai Lam. 2019. Persona-Aware Tips Generation. In *The World Wide Web Conference*. 1006–1016.

[24] Piji Li, Zihao Wang, Zhaochun Ren, Lidong Bing, and Wai Lam. 2017. Neural rating regression with abstractive tips generation for recommendation. In *Proc. of SIGIR*. 345–354.

[25] Stephen W Litvin, Ronald E Goldsmith, and Bing Pan. 2008. Electronic word-of-mouth in hospitality and tourism management. *Tourism management* 29, 3 (2008), 458–468.

[26] Linqing Liu, Yao Lu, Min Yang, Qiang Qu, Jia Zhu, and Hongyan Li. 2018. Generative adversarial network for abstractive text summarization. In *Thirty-second AAAI conference on artificial intelligence*.

[27] Mitchell Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Treebank. (1993).

[28] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint* abs/1301.37810 (2013).

[29] Thanh-Son Nguyen, Hady W Lauw, and Panayiotis Tsaparas. 2013. Using micro-reviews to select an efficient set of reviews. In *Proc. of CIKM*. 1067–1076.

[30] Slava Novgorodov, Ido Guy, Guy Elad, and Kira Radinsky. 2019. Generating product descriptions from user reviews. In *Proc. of WWW*. 1354–1364.

[31] Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proc. of EMNLP*, Vol. 14. 1532–1543.

[32] Irina Rish. 2001. An empirical study of the naive Bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, Vol. 3. 41–46.

[33] Ruben Sipos and Thorsten Joachims. 2013. Generating comparative summaries from reviews. In *Proc. of CIKM*. 1853–1856.

[34] Doug Snowball. 1980. Some effects of accounting expertise and information load: An empirical study. *Accounting, Organizations and Society* 5, 3 (1980), 323–338.

[35] Cheri Speier, Joseph S Valacich, and Iris Vessey. 1999. The influence of task interruption on individual decision making: An information overload perspective. *Decision Sciences* 30, 2 (1999), 337–360.

[36] Oren Tsur and Ari Rappoport. 2009. Revrank: A fully unsupervised algorithm for selecting the most helpful book reviews. In *Proc. of ICWSM*.

[37] Ingmar Weber, Antti Ukkonen, and Aris Gionis. 2012. Answers, not links: extracting tips from yahoo! answers to address how-to web queries. In *Proc. of WSDM*. 613–622.

[38] Alfan Farizki Wicaksono and Sung-Hyon Myaeng. 2012. Mining Advices from Weblogs. In *Proc. of CIKM*. 2347–2350.

[39] Alfan Farizki Wicaksono and Sung-Hyon Myaeng. 2013. Toward advice mining: Conditional random fields for extracting advice-revealing text units. In *Proc. of CIKM*. 2039–2048.

[40] Qing-Song Xu and Yi-Zeng Liang. 2001. Monte Carlo cross validation. *Chemometrics and Intelligent Laboratory Systems* 56, 1 (2001), 1–11.

[41] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical attention networks for document classification. In *Proc. of NAACL*. 1480–1489.

[42] Qiang Ye, Rob Law, and Bin Gu. 2009. The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management* 28, 1 (2009), 180–182.

[43] Di Zhu, Theodoros Lappas, and Juheng Zhang. 2018. Unsupervised tip-mining from customer reviews. *Decision Support Systems* 107 (2018), 116–124.