A Framework to Discover Significant Product Aspects from E-commerce Product Reviews

Saratchandra Indrakanti
eBay Inc.
San Jose, California
sindrakanti@ebay.com

Gyanit Singh
eBay Inc.
San Jose, California
gysingh@ebay.com

ABSTRACT

Product reviews increasingly influence buying decisions on e-commerce websites. Reviewers share their experiences of using a product and provide unique insights that are often valued by other buyers and not available in seller provided descriptions. Product-specific opinions expressed by reviewers and buyer perspectives provided by them can be employed to power novel buyer-centric shopping experiences, as opposed to existing e-commerce experiences tailored to product catalogs. Product aspects that have been opined upon and collectively discussed by the reviewers in product reviews can be identified and aggregated to capture such insights. However, owing to the vast diversity of products listed on modern e-commerce platforms, usage of colloquial language in reviews and vocabulary mismatch between seller(manufacturer) and buyer terminology; identifying such significant product aspects becomes a challenging problem at scale. In this paper, we present a framework for product aspect extraction and ranking developed to identify product aspects from reviews and quantify their importance based on collective reviewer opinions. We further examine the value of incorporating domain-specific knowledge into our model, and show that domain-specific knowledge significantly improves performance of the model.

CCS CONCEPTS

• **Information systems → Information extraction**; *Sentiment analysis*; *Summarization*; • **Computing methodologies** → Language resources;

KEYWORDS

Aspect Mining, Opinion Mining, Product Reviews, , E-commerce, Sentiment Analysis

ACM Reference Format:
Saratchandra Indrakanti and Gyanit Singh. 2018. A Framework to Discover Significant Product Aspects from E-commerce Product Reviews. In *Proceedings of ACM SIGIR Workshop on eCommerce (SIGIR 2018 eCom)*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

In: J. Degenhardt, G. Di Fabbrizio, S. Kallumadi, M. Kumar, Y.-C. Lin, A. Trotman, H. Zhao (eds.): Proceedings of the SIGIR 2018 eCom workshop, 12 July, 2018, Ann Arbor, Michigan, USA, published at http://ceur-ws.org

1 INTRODUCTION

The browse and navigational experiences on most e-commerce websites today are tailored to product catalogs and manufacturer-provided product attributes. However, buyer-centric navigational experiences constructed based on buyer-provided product insights can potentially enhance the online shopping experience and lead to better user engagement [28]. For instance, the product catalog may be associated with objective values such as *15 inch* or *retina* corresponding to the attribute *display* for a laptop, in contrast to qualitative terminology such as *crisp display* or *large display* that can assist buyers in their shopping journey. Meanwhile, product reviews written by reviewers sharing their experiences with a product, have evolved into community powered resources that provide qualitative insights into products from a buyer perspective. By capturing valuable buyer perspectives of products that are not available elsewhere in seller or manufacturer provided descriptions or metadata, reviews have been playing an increasingly important role in shaping buying decisions on most popular e-commerce websites. Discovering such buyer-provided product insights and opinions from product reviews can help power novel and engaging online shopping experiences, such as the ones shown in Figure 1.

While manufacturer-provided attributes associated with product catalogs generally comprise of structured metadata, product reviews comprise of unstructured text. Product-specific insights in natural language text can be identified by extracting opinionated words and opinion targets, also referred to as *aspects*. However, this task introduces several challenges. Product reviews are written in a generally colloquial style as opposed to technical vocabulary, which introduces ambiguity in identifying product-specific aspects. For instance, in the sentence *"This camera is easy to use"*, *use* is a colloquial term with *easy to use* representing a product-specific opinion, while *use* in *"I use a large screen"* does not. Further, the vast diversity of products listed on modern e-commerce platforms ranging from electronics to media to art have differences in the nature of aspects discussed in reviews. The name of a person such as the author in a book review can be a useful product-specific aspect, whereas this may not be the case with respect to the name of a person such as a friend who suggested the product in a review about a gadget. Aspect extraction techniques must be robust enough to account for the domain-specific nuances.

Certain words that have been extensively used in reviews for a product, may not be very informative aspects for the product. For instance, *camera* is extensively used in *digital cameras* category. This is not a very informative word as an aspect for a product in this category, although it could be a meaningful aspect for a smartphone. Further, words that capture interpersonal relationships such as *friend*, *brother* etc. are not relevant as product aspects. Synonyms such as *Images* and *pictures* are used interchangeably in reviews for cameras, and a single aspect that represents them must be selected. To that end, a scalable framework developed to identify and rank opinionated product-specific aspects from product reviews, that can be leveraged to power buyer-centric shopping experiences for modern e-commerce websites is presented in this paper.

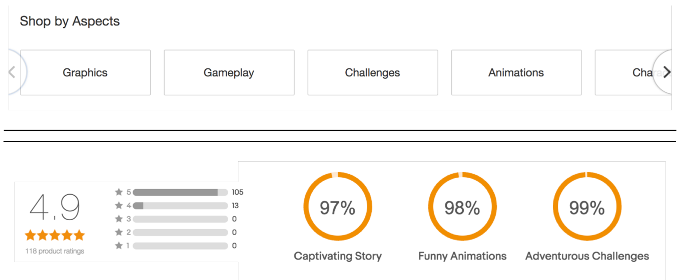

Figure 1: Examples of novel buyer-centric e-commerce experiences that can be constructed based on vocabulary and opinions form product reviews. The interface at the top lets users navigate the website using aspects from reviews. The bottom interface conveys aggregate opinions on a product.

In the context of this work, aspects are attributes or features of a product discussed in reviews, upon which the reviewer expresses an opinion. They are also referred to as opinion targets in the field of opinion mining. For instance, *battery life*, *screen size*, *camera resolution*, *operating system* could be aspects discussed in smartphone reviews. Consider the following excerpt from a review for a video game: *"Compelling gameplay and story and beautiful graphics"*. The reviewer expresses the opinion *compelling* on the aspect *gameplay*. Different reviewers discuss a variety of aspects in their reviews for a product, and express varying opinions. In order to identify *important* aspects for a product from its reviews, the collective opinions expressed by reviewers on the aspects must be aggregated to quantify their importance. This entails firstly, identifying opinion and target word candidates (aspect extraction), and next, ranking them to quantify reviewer emphasis (aspect ranking). To achieve this, we propose a graph based framework to discover and rank aspects from product reviews. In the proposed framework, first, aspect candidates are identified from each sentence by exploiting certain word dependencies from its sentence dependency structure. Then, an aspect graph is constructed for a given product based on the identified dependencies. Graph centrality measures are employed in conjunction with domain-specific knowledge derived in the form of structured metadata from product catalogs to quantify the importance of aspects and rank them.

There are several motivating factors contributing to our choice of adopting an unsupervised approach to aspect extraction and intuition behind incorporating domain-specific knowledge from product catalogs as introduced above. Firstly, the framework must be robust and able to scale across a large number of diverse product categories present on e-commerce websites. Owing to the constantly evolving nature of e-commerce content, inventory and categories, unsupervised approaches are a natural fit since they do not rely on training data. Although supervised models generally have performed better in terms of precision and recall, procuring properly annotated data representative of a large and diverse set of categories of products for training supervised models such as conditional random fields [11] could be a daunting task. Next, we apply domain-specific knowledge derived from manufacturer-provided aspects (MPAs) in product catalogs as a post-processing step to mitigate the occurrence of false positives among the extracted aspects. While semi-supervised models that rely on seed words have been proposed for aspect extraction [22], compiling a set of seed words for every category of products may not be feasible. Although it is intuitive to attempt applying domain specific MPAs as seed words, this may not be effective owing to the mismatch between manufacturer and buyer vocabulary. Further, such domain-specific knowledge may not be available or may be very sparse in many product categories, as depicted in Figure 2.

While aspect extraction and ranking methods have been well researched individually, very few works have explored this as a unified problem in conjunction with incorporating domain-specific knowledge to improve accuracy of the methods. Enhancing existing literature, we present a unified graph-based framework that facilitates both identification of significant aspects from product reviews and ranking them, while leveraging domain knowledge and being applicable to large e-commerce websites that have a diverse catalog of products. The authors of [21] proposed an aspect extraction method that selects aspects by applying dependency rules to sentence dependency trees. We use a similar approach to aspect extraction, however, in addition to that we developed a graph centrality based method to rank aspects. While the intuition behind our centrality-based ranking is comparable to [29], where the authors proposed a graph-based algorithm inspired from pagerank to rank aspects, we extend it to incorporate domain knowledge to the task.

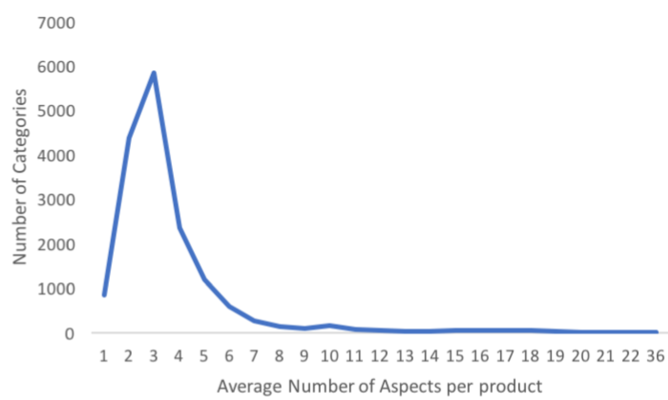

Figure 2: Plot showing the distribution of average number of manufacturer-provided aspects(MPAs) per product in a category,(mean = 3.5, Std. dev = 1.7). Many categories on e-commerce websites have few MPAs per product.

We evaluated the proposed framework on a human-judged evaluation dataset generated by sampling product reviews from a diverse set of e-commerce product categories. Experiments on the evaluation dataset indicate that our framework can scale across the thousands of diverse product categories present on e-commerce websites. Further, we show that our approach of incorporating domain knowledge improves positively contributes to precision of the model. We observe a 9.8% overall improvement in mean average precision when domain knowledge is applied.

The main contributions of this paper can be summarized as follows:

- We propose a scalable graph-based framework for aspect extraction and ranking from e-commerce product reviews.
- We examine the benefits of applying domain-specific knowledge to the problem of aspect extraction and ranking.

2 RELATED WORK

Aspect extraction models have been proposed by several researchers over the past decade. Authors of [9] first introduced an unsupervised model based on frequent itemset mining. An improved method was proposed in [19], where PMI statistics were incorporated from an online text corpus to improve precision. A rule-based model that leverages observed linguistic patterns from sentence dependency trees to extract aspects was presented in [21]. A combination of rules and sentence dependency structure was used to generate a graphical structure based on sentiment and aspect pairs, and product aspects were identified using an algorithm based on page-rank was used to rank aspects in [29]. Generally, in the area of text mining, graph based models that employ centrality measures to identify significant phrases in text have proven to be effective [13]. Discourse-level opinion graphs have been proposed by [24] to interpret opinions and discourse relations. The authors presented an unsupervised model to automatically select a set of rules that utilize the sentence dependency structure in [12]. Syntactical structure of sentences along with various statistical measures have been used in works such as [23], [30], [10].

Different variations of topic models have been explored by researchers in extracting aspects from reviews. Multi-grain topic models that extends topic modeling techniques such as LDA and PLSI to extract local and global aspects for products within a category were proposed in [25]. LDA was applied to identify latent topics in [3] and representative words for topics were selected as aspects. Probabilistic graphical models that extend LDA and PLSI were proposed in [14] to extract aspects and determine a rating based on sentiment expressed with respect to each aspect.

Semi-supervised models have been introduced to guide certain unsupervised models towards more precise aspects by using a domain specific set of seed words. A double propagation approach to extract opinion words and targets iteratively by bootstrapping with a seed opinion lexicon of small size was proposed in [22]. The semi-supervised model proposed in [15] uses seed words provided for a few aspect categories and extracts and clusters aspect terms into categories. Seed words extracted from product descriptions were used to group reviews and a labeled extension of LDA was used to extract aspects in a semi supervised way in [26].

Several supervised models have been proposed to extract aspects from reviews. A model based on conditional random fields was trained with features that were built using parts of speech tags of tokens and dependency path distances in [11]. A model based on Convolutional neural networks that has word embeddings provided as input features and extended to employ linguistic patterns is presented in [20].

3 METHODOLOGY

The methodology can be structured into three major phases: 1) Aspect extraction 2) Aspect graph construction and 3) Aspect ranking and post-processing. An overview of the proposed methodology is provided in Figure 3. For a given product, we first select potential aspect candidates by applying Dependency tree pruning algorithm (DPT) on review sentences, as part of aspect extraction. Next, construct an aspect graph by aggregating the relations returned by DPT, and compute centralities for nodes of the graph. We then apply domain-specific knowledge along with other post-processing steps aimed at reducing the occurrence of false positives and compute a ranking of the aspects. Post-processing steps include synonym-based clustering of aspects, compiling a list of non-exclusive aspects that can be demoted, and applying domain-specific MPAs to boost relevant aspects. These methods are formally described in detail in this section.

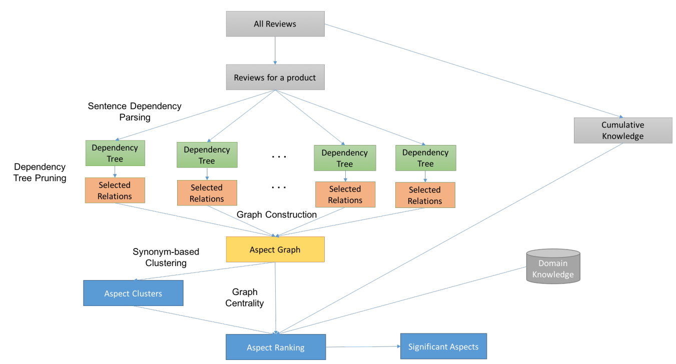

Figure 3: Overview of the proposed aspect extraction and ranking framework. The work flow from review aggregation to discovery of important aspects is depicted.

3.1 Aspect Definition

An aspect can be defined as an attribute or feature of a product, that is discussed in the product review content and upon which the reviewers express an opinion. Product aspects are most frequently nouns [16], however certain verb forms can also occur as aspects. It must be noted that not all nouns are aspect candidates. Consider the following excerpt from a review: *"Sound quality is amazing, and battery lasts long enough as well"*, where *sound quality* and *battery* are aspect candidates. However, *"I worked in Electronics for 35 years."* has no aspect candidates, although *electronics* and *years* are nouns. The challenge in discovering aspect candidates mainly involves identifying those words that 1) the reviewer has expressed an opinion on, and 2) are attributes of the product and describe its features.

3.2 Aspect Extraction

For a given product, aspect candidates are extracted from each sentence of every review, based on its sentence dependency tree.

Sentence dependency trees capture the grammatical relations between words that comprise a sentence. The dependencies are binary asymmetric relations between a word identified as head (generally a verb) and its dependents [6]. The nature of the relationship is denoted by a dependency label associated with the edge connecting the two words in the relationship. For instance, Figure 4 depicts the dependency tree for the following sentence: *"The framerate was high, battery life was long, the visual effects looked as polished as today's consoles"*. A detailed description of dependency trees can be found in [6]. The open source library Spacy [8] is employed to generate dependency trees owing to its combination of speed and accuracy [5]. Punctuation is retained and no pre-processing is performed prior to producing dependency trees, since lemmatization or other pre-processing steps may affect the accuracy of dependency tree generation. Next, we describe the sentence tree pruning algorithm which returns dependency relations associated with aspect candidates.

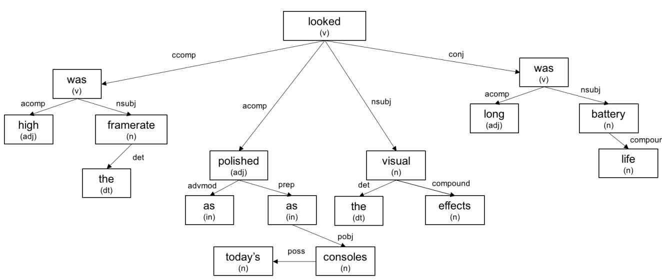

Figure 4: Sentence dependency tree generated for an example review sentence *"The framerate was high, battery life was long, the visual effects looked as polished as today's consoles"*. Nodes are individual words in a sentence, while edges represent dependencies between words.

3.2.1 Dependency Tree Pruning. Dependency tree pruning algorithm (DTP) prunes the dependency tree generated for each sentence to retain relations that are associated with aspect candidates. Specifically, for every aspect candidate we retain the dependencies that capture the relations between the aspect, opinion, and the head word of the aspect. Aspect candidates are identified subject to a set of dependency rules that are applied to the dependency tree. The set of rules R, some of which have been defined by the authors of [21], that determine if a word is an aspect candidate, are described below:

(1) A noun n is an aspect if there exists a parent-child relation between n and another word a, and a is either an adjective or an adverb. Here the set of words a that satisfy this condition constitute the opinion.
(2) Adjective-adverb sibling rule: A noun n is selected as an aspect, and a as the opinion if there exists a word a that shares the same parent (head term), and a is either an adjective or adverb.
(3) If a word v has a direct object relation with a noun n, and v is a verb, then n is selected as an aspect, with v being the opinion.
(4) If A verb v has a noun n as a parent (head term), and v has a parent-child relation with another word a which is an adjective, then n is selected as an aspect, with a being the opinion.
(5) If a noun n_1 in a conjunct relation or a prepositional relation with another noun n_2, and n_2 is parent-child relation with an adjective a, n_1 is selected as an aspect with a being
(6) If a word v is in an open clausal relationship with another word a, with v being a verb and a being an adjective or adverb, then select v as an aspect and a as opinion.

Given a sentence s, its dependency tree $D_s : \{dep\langle w_1, w_2\rangle\}$, where $w_1, w_2 \in s$ are any two words in the sentence s with a dependency relation dep, is generated. Aspect candidates α satisfying atleast one of the rules in R are selected by DTP along with the associated opinion and head word dependencies. The satisfying relations $dep\langle head_\alpha, \alpha\rangle, dep\langle O_\alpha, \alpha\rangle$ are returned, where $head_\alpha$ is the head term of α and O_α is the corresponding opinion. Figure 5 shows the result produced by the algorithm for the example sentence.

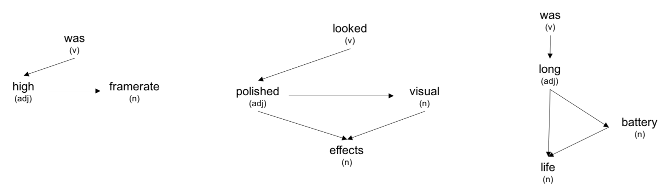

Figure 5: Dependency relations returned after applying the dependency tree pruning algorithm on the example sentence.

3.3 Aspect Graph

In order to capture the collective opinions expressed on aspects by all the reviewers for a product, an aspect graph is constructed from the relations returned by DTP. DTP is applied on each sentence in the review corpus for a product, to return a set of relations associated with aspect candidates for each sentence. As a pre-processing step, the review corpus is run through a lemmatizer to identify the canonical form of each word. The words that share the same canonical form are replaced with a representative word selected based on frequency of occurrence in the corpus. Each of the relations is added to aspect graph $G_p = (V, E)$, a directional graph constructed for the given product P. A node $\eta \in V$ in the graph is a tuple (w, pos_w, t_w) representing a word w, its parts-of-speech tag pos_w and its type t_w where $t \in \{head, opinion, aspect\}$, while the dependency relations $\eta_1 \longrightarrow \eta_2$ returned by DTP denote edges $e_{12} \in E$. Weight ω_{12} of the edge e_{12} is the frequency of such relations in the corpus. While, weight could be extended to include other properties such as aggregate sentiment associated with an aspect candidate, we limit it to frequency in this discussion. Figure 6 shows an example aspect graph constructed based on the sentence discussed previously.

3.4 Aspect Ranking

We rank candidate aspects based on their measured importance in the aspect graph. To that end, we utilize graph centrality measures

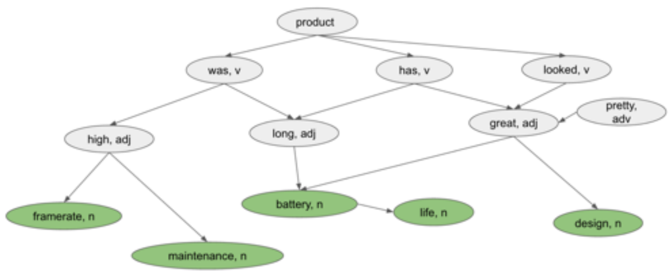

Figure 6: Illustration of an aspect graph constructed from relations returned by DTP.

for the aspect graph to quantify the importance of aspects as expressed collectively by reviewers in the product reviews. Various graph centrality measures exist that capture different properties of a graph, and offer varying perspectives to measuring the importance of a node in a graph [2]. The aspect ranking problem is formulated as follows: Select top k nodes $\eta(w, pos_w, t_w) \subset V$ from aspect graph $G_p = (V, E)$ for product p, when ranked by ranking measure ρ_G and $t_\eta = aspect$.

3.4.1 Ranking Measures. The centrality measures used for ranking aspects, ρ_G, are discussed in this section. Graph centrality measures quantify the importance of nodes in a graph. Since importance is subjective, there exist various centrality measures that capture different properties of the structure of a graph and the influence of specific nodes. Centrality measures have been extensively used in varying applications such as extracting keywords from text, with encouraging results; an overview of this area is available in [1].

In this work, we explore the application of in-strength centrality and page-rank [17] to rank the nodes in the aspect graph. In-strength centrality for a node is defined as the sum of weights of all incoming edges to it. It translates to the number of times a given aspect candidate has occurred in contexts where an opinion has been expressed about it. Although it is a simple measure, in-strength centrality has been found to be an effective measure in studies such as [27], where the authors applied various centralities to noun phrase networks in extracting keywords from abstracts.

While ranking nodes based on graph centrality can assist in discovering important aspects, there could be several factors that affect the quality of the selected aspects leading to false positives. In the following sections, we introduce specific post-processing techniques: incorporating domain-specific knowledge, synonym-based clustering of aspects and aspect exclusivity to reduce false positives and improve the accuracy of the framework.

3.4.2 Domain-specific knowledge. Domain-specific knowledge for a product can be obtained from structured metadata available in product catalogs. Many products are associated with certain aspect names and values provided by the manufacturer, referred to as MPAs, as part of their technical specifications. For example, Microsoft Xbox One S may be associated with MPAs such as *Device Input Support, Console color, Internet Connectivity, Hard Drive Capacity*. We aggregate MPAs within a category of products to create a domain-specific MPA dictionary. Extracted aspects are re-ranked based on a match with an entry in the MPA dictionary or one of its synonyms. Matching aspects are promoted ahead of the ones that have no matches in the centrality-based ranking.

3.4.3 Synonym-based Clustering. We use 300-dimensional word embeddings for one million vocabulary entries trained on the Common Crawl corpus using the GloVe algorithm [18] to compute word similarities. Synonym clusters of similar aspects such as *picture* and *image* are formed by grouping pairs which have a cosine similarity of word vectors greater than a threshold $\eta \in [0, 1]$. An agglomerative hierarchical clustering approach is adopted to form the clusters, while η is empirically learnt from a small dataset of synonyms for this task. A representative word is selected for each cluster, based on the ranking measure for the node representing it in the aspect graph.

3.4.4 Aspect Exclusivity. A set of aspects that are not exclusive to a few categories, but are widely used in reviews across all categories of products is generated. Non-exclusive aspects generally may not offer value in describing attributes very specific to a product and are demoted. For example, although the aspect *features* occurs very frequently in reviews of a variety of products, it is a very broad in scope and offers little knowledge about a product, in contrast to an aspect like *suction power*, which is very specific to vacuum cleaners. We generate category-wise review corpora, by aggregating all reviews of products belonging to the same category. We generate a set of documents $D : \{d_1, d_2, ..d_n\}$, corresponding to categories $C : \{c_1, c_2, ..c_n\}$, where the document d_i contains reviews of all products in c_i. An aspect α is considered to be non-exclusive if $|d_i : s \in d_i| > m$, i.e. it occurs in more than m categories.

4 EVALUATION

4.1 Evaluation Dataset

Evaluation of the proposed methods was performed on a human-judged evaluation dataset generated by sampling product reviews from a diverse set of e-commerce product categories including those in areas such as media, electronics, books, health & beauty, home & garden etc. While, there exist several aspect evaluation datasets published previously including [4], [29], [22], we opted to compile a fresh dataset for two main reasons. Firstly, many of the existing datasets are focused on very specific categories such as electronics or restaurants. Modern e-commerce websites have a much broader variety of product categories, and the evaluation dataset must be representative of them. Next, many of the existing datasets have been built to evaluate the effectiveness of aspect extraction from individual reviews. However, the focus of this work is in identifying important aspects collectively discussed by all the reviewers of a product.

The evaluation dataset consisted of approximately 60,000 reviews for 427 products in 126 categories, and their respective human evaluated aspects. The products were sampled in a fashion that is representative of the e-commerce categories that receive product reviews. The selected products had an average of 140 reviews per product. Manually reading all the reviews in the evaluation dataset to identify every aspect occurring in the reviews can be a very demanding task for human judges. Further, such a process is prone to errors owing to fatigue associated with reading a large number

Blender	Video Game	Vacuum Cleaner	Face Powder	Pet Medicine	Coffee Maker	Movie
smoothies	game play	suction	foundation	infestation	brewer	effects
speed	graphics	attachments	brush	retriever	heating	characters
recipes	challeneges	maneuverability	acne	cats	cup size	graphics
performance	story	dyson	touch	flea treatments	flavour	expressions
waste	gamer	edges	ingredients	application	steel	tradition
warranty	animations	floor	summer	hair	k cup	animation

Table 1: Top aspects identified for some of the products in the dataset. False positives are shown in red.

of reviews, inconsistency in individual interpretation in identifying aspects [7]. To simplify the evaluation task for the human judges, we generated a list of potential aspects for each product, determined based on their part of speech tag and frequency thresholds, and requested the evaluators to indicate if they thought the candidate was relevant. The evaluation dataset had an average of 31 aspect candidates provided per product and the evaluators were asked to provide a binary decision on each candidate. The aspect candidates for each product received 3 votes and a majority was considered as the final evaluation.

4.2 Results

Experiments were performed on the evaluation dataset to examine: 1) The performance of the centrality measures in comparison to a tf-idf baseline. 2) The contribution of domain-specific knowledge to the current task, and 3) Differences in performance of the methods in contrasting product categories.

The results produced after evaluating in-strength centrality and page-rank methods using the evaluation dataset are shown in Tables 2 and 3. We use tf-idf as a baseline to compare the performance of the centrality-based methods. To compute tf-idf, all the reviews for a given product are considered as a document, while the corpus consists of documents representing each product. In order to maintain consistency, the same pre-processing and post-processing methods used in aspect graph construction are applied to computing tf-idf. Table 2 shows the results for all 427 products in the evaluation set with and without domain knowledge incorporated. Table 3 compares the performance of the models on electronics and media categories separately. Electronics categories consist of product related to smartphones, computers, printers etc., while media categories comprise of books, DVDs etc. There were 76 products belonging to media related products in the evaluation dataset, while there were 103 electronics products.

While, the graph centrality-based measured perform better than the baseline tf-idf, we have observed that although it is a simpler measure, in-strength centrality generally performs very similar to page-rank. Owing to the limited depth of the aspect graphs and smaller number of candidate aspects per graph, there are limitations to leveraging the properties of page-rank, and simpler measures can be effective in this case. Further, we also observed that these methods perform better in more structured and well-defined categories such as electronics than categories such as media, as can be seen from Table 3.

We also investigated the influence of domain knowledge (using MPAs) on aspect ranking. First, MPAs associated with products in the evaluation dataset were compared to the aspects identified by the evaluators. Only 23.8% of MPAs match with the aspects identified by evaluators, emphasizing the mismatch between reviewer and manufacturer vocabulary. Further, Table 2 compares the precision@k obtained for in-strength and pagerank measures with and without using MPAs. We can see that MPAs have a positive influence on performance and aspect discovery can benefit from utilizing them.

5 SUMMARY

Product reviews are repositories of valuable buyer-provided product insights that other prospective buyers on e-commerce websites trust. Product insights discovered from reviews can power novel and engaging buyer-centric browsing and shopping experiences on e-commerce websites, in contrast to the existing experiences tailored to product catalogs. In this paper, we present methods to extract such insights from product reviews and quantify their importance based on collective opinions expressed by reviewers.

To capture top product insights from reviews, we present a framework to identify product aspects based on sentence dependency structure, and rank them by applying graph centralities. We also incorporate domain knowledge into out framework and study the contribution of domain knowledge to this task. The method we proposed is unsupervised and can scale across a diverse set of categories. We evaluate the proposed methods on an evaluation dataset that is representative of the product categories on major e-commerce platforms. The results show that the proposed framework can be applicable across a diverse set of product categories and that domain knowledge can positively contribute to this task.

REFERENCES

[1] Slobodan Beliga, Ana Meštrović, and Sanda Martinčić-Ipšić. 2015. An overview of graph-based keyword extraction methods and approaches. *Journal of information and organizational sciences* 39, 1 (2015), 1–20.

[2] Stephen P Borgatti and Martin G Everett. 2006. A graph-theoretic perspective on centrality. *Social networks* 28, 4 (2006), 466–484.

[3] Samuel Brody and Noemie Elhadad. 2010. An unsupervised aspect-sentiment model for online reviews. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 804–812.

[4] Zhiyuan Chen and Bing Liu. 2014. Topic modeling using topics from many domains, lifelong learning and big data. In *International Conference on Machine Learning*. 703–711.

[5] Jinho D Choi, Joel R Tetreault, and Amanda Stent. 2015. It Depends: Dependency Parser Comparison Using A Web-based Evaluation Tool.. In *ACL (1)*. 387–396.

[6] Marie-Catherine De Marneffe and Christopher D Manning. 2008. *Stanford typed dependencies manual*. Technical Report. Technical report, Stanford University.

[7] Pinar Donmez, Jaime G Carbonell, and Jeff Schneider. 2009. Efficiently learning the accuracy of labeling sources for selective sampling. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 259–268.

[8] Matthew Honnibal and Mark Johnson. 2015. An Improved Non-monotonic Transition System for Dependency Parsing. In *Proceedings of the 2015 Conference on*

	Including MPAs			Excluding MPAs		
	In-Strength	**PageRank**	**Tf-Idf**	**In-Strength**	**PageRank**	**Tf-Idf**
P@1	.761	.758	.691	.712	.728	.663
P@3	.731	.726	.674	.681	.702	0.642
P@5	.702	.692	.658	.652	.668	.631
P@7	.676	.681	.632	.639	.644	.612
MAP	.647	.655	.597	.613	.619	.559
Recall	.621	.603	.541	.523	.511	.491
F1	.634	.628	.568	.564	.560	.523

Table 2: Precision@k (P@k) , MAP, recall and f1 values with and without using MPAs obtained using In-strength, pagerank and tf-idf(baseline) on all products in the evaluation dataset. MPAs (domain knowledge) have a positive contribution to the performance of the model.

	Electronics			Media		
	In-Strength	**PageRank**	**Tf-Idf**	**In-Strength**	**PageRank**	**Tf-Idf**
P@1	.774	.783	.713	.636	.643	.561
P@3	.749	.758	.688	.622	.624	.548
P@5	.718	.727	.669	.613	.619	.532
P@7	.691	.711	.646	.598	.603	.511
MAP	.673	.684	.621	.574	.582	.493
Recall	.649	.635	.572	.562	.574	.513
F1	.661	.658	.595	.568	.578	.502

Table 3: Comparison of Precision@k (P@k), MAP, recall and f1 values obtained for Electronics categories and Media categories. Improvement from the baseline in media indicates that the model can scale across various categories.

Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Lisbon, Portugal, 1373–1378. https://aclweb.org/anthology/D/D15/D15-1162

[9] Minqing Hu and Bing Liu. 2004. Mining opinion features in customer reviews. In *AAAI*, Vol. 4. 755–760.

[10] Saratchandra Indrakanti, Gyanit Singh, and Justin House. 2018. Blurb Mining: Discovering Interesting Excerpts from E-commerce Product Reviews. In *Companion of the The Web Conference 2018 on The Web Conference 2018*. International World Wide Web Conferences Steering Committee, 1669–1675.

[11] Niklas Jakob and Iryna Gurevych. 2010. Extracting opinion targets in a single-and cross-domain setting with conditional random fields. In *Proceedings of the 2010 conference on empirical methods in natural language processing*. Association for Computational Linguistics, 1035–1045.

[12] Qian Liu, Zhiqiang Gao, Bing Liu, and Yuanlin Zhang. 2015. Automated Rule Selection for Aspect Extraction in Opinion Mining.. In *IJCAI*. 1291–1297.

[13] Rada Mihalcea and Paul Tarau. 2004. TextRank: Bringing order into texts. Association for Computational Linguistics.

[14] Samaneh Moghaddam and Martin Ester. 2011. ILDA: interdependent LDA model for learning latent aspects and their ratings from online product reviews. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. ACM, 665–674.

[15] Arjun Mukherjee and Bing Liu. 2012. Aspect extraction through semi-supervised modeling. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*. Association for Computational Linguistics, 339–348.

[16] Hiroshi Nakagawa and Tatsunori Mori. 2002. A simple but powerful automatic term extraction method. In *COLING-02 on COMPUTERM 2002: second international workshop on computational terminology-Volume 14*. Association for Computational Linguistics, 1–7.

[17] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. *The PageRank citation ranking: Bringing order to the web.* Technical Report. Stanford InfoLab.

[18] Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global Vectors for Word Representation.. In *EMNLP*, Vol. 14. 1532–1543.

[19] Ana-Maria Popescu, Bao Nguyen, and Oren Etzioni. 2005. OPINE: Extracting product features and opinions from reviews. In *Proceedings of HLT/EMNLP on interactive demonstrations*. Association for Computational Linguistics, 32–33.

[20] Soujanya Poria, Erik Cambria, and Alexander Gelbukh. 2016. Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems* 108 (2016), 42–49.

[21] Soujanya Poria, Erik Cambria, Lun-Wei Ku, Chen Gui, and Alexander Gelbukh. 2014. A rule-based approach to aspect extraction from product reviews. In *Proceedings of the second workshop on natural language processing for social media (SocialNLP)*. 28–37.

[22] Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. 2011. Opinion word expansion and target extraction through double propagation. *Computational linguistics* 37, 1 (2011), 9–27.

[23] Christopher Scaffidi, Kevin Bierhoff, Eric Chang, Mikhael Felker, Herman Ng, and Chun Jin. 2007. Red Opal: product-feature scoring from reviews. In *Proceedings of the 8th ACM conference on Electronic commerce*. ACM, 182–191.

[24] Swapna Somasundaran, Galileo Namata, Lise Getoor, and Janyce Wiebe. 2009. Opinion graphs for polarity and discourse classification. In *Proceedings of the 2009 Workshop on Graph-based Methods for Natural Language Processing*. Association for Computational Linguistics, 66–74.

[25] Ivan Titov and Ryan McDonald. 2008. Modeling online reviews with multi-grain topic models. In *Proceedings of the 17th international conference on World Wide Web*. ACM, 111–120.

[26] Tao Wang, Yi Cai, Ho-fung Leung, Raymond YK Lau, Qing Li, and Huaqing Min. 2014. Product aspect extraction supervised with online domain knowledge. *Knowledge-Based Systems* 71 (2014), 86–100.

[27] Zhuli Xie. 2005. Centrality measures in text mining: prediction of noun phrases that appear in abstracts. In *Proceedings of the ACL student research workshop*. Association for Computational Linguistics, 103–108.

[28] Qian Xu and S Shyam Sundar. 2014. Lights, camera, music, interaction! Interactive persuasion in e-commerce. *Communication Research* 41, 2 (2014), 282–308.

[29] Zhijun Yan, Meiming Xing, Dongsong Zhang, and Baizhang Ma. 2015. EXPRS: An extended pagerank method for product feature extraction from online consumer reviews. *Information & Management* 52, 7 (2015), 850–858.

[30] Jianxing Yu, Zheng-Jun Zha, Meng Wang, and Tat-Seng Chua. 2011. Aspect ranking: identifying important product aspects from online consumer reviews. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics, 1496–1505.