Use-centric mining of customer reviews

Thomas Lee
Department of Operations and Information Management
University of Pennsylvania, Philadelphia, PA 19104
thomas.lee@wharton.upenn.edu

Abstract. Prior research involving customer reviews focuses on individual consumers and/or specific products. By contrast, use-centric mining aggregates over all reviews for all products in a category. Specifically, we induce a category-specific ontology from reviews and use that ontology to automatically extract product features and uses. We then use frequent-item sets to match product uses with product features. For example, what features do reviews suggest are important when purchasing a digital camera for travel? Use-centric mining may lead to recommendations based upon the correspondence between features and applications. In this paper, we focus on feature extraction and use-centric mining. We report preliminary results from a database of more than 300 digital camera customer reviews.

1. INTRODUCTION

A side effect of manufacturing trends like mass customization is the dizzying selection of makes and models facing the modern consumer in virtually all product categories. For example, a typical big box retailer might stock 30-40 products in the digital camera category. Online, Bestbuy.com lists more than 100 models in that category. Amazon.com lists more than 600 digital cameras and camera packages.

To help a customer shop within a product category, a knowledgeable clerk or informed friend might first identify how the customer hopes to use the item. Based upon the customer's intended uses, the clerk or friend can identify category-specific attributes and values. The resulting menu of recommendations associates products with the feature bundles most appropriate for the customer's intended uses. For example, customers shopping for a digital camera to use *while on vacation* might limit their consideration set to cameras whose *size* is *compact*. Attributes and values like *size* and *compact* are necessarily category specific. A compact dishwasher is much larger than a compact digital camera. There are three critical steps in this heuristic: obtain customer uses, associate uses with product features, and align bundles of product features with a set of product alternatives. While human experts rely upon experience to help customers match uses to product features, this research envisions a category-specific automated decision support tool to assist shoppers. Specifically, this research focuses on step two: learning the relationship between uses and product features from the knowledge contained within the online customer reviews of a single product category. We call this process use-centric mining of customer reviews.

Prior research involving customer reviews echoes one of two themes: product focus, or customer focus. Product-focused researchers cluster reviews by manufacturer and model. The reviews for a specific make and model help manufactures discern positive and negative aspects of that particular design both for incremental improvements and new product development [8, 13]. Moreover, prospective buyers use product-specific reviews to evaluate a particular item [8]. Customer-focused research on product reviews, by contrast, models individual consumers. By aggregating the reviews (or purchase history) of a single reviewer, researchers can create a user profile to predict or recommend future purchases [3].

We propose use-centric mining as a complementary approach to leveraging reviews. For a specific product category, we propose to learn those product features that align with particular uses and applications. Unlike product-focused research, we aggregate the reviews from all makes and models in a single product category. Unlike customer-focused research, we consider only those reviews in the designated product category. For use-centric mining, we first semi-automatically induce a category-specific ontology (a structured, controlled vocabulary) of product features and product usages. Using the ontology, each review is automatically mapped into a vector representation. Finally, we mine the review vectors for those frequent item sets that associate particular usages with particular features.

Because we employ separate strategies for inducing ontologies of features and of usages, we defer detailed discussions to separate papers. In this paper, we describe our ontology-based extraction-as-classification approach to transform each review into a 0,1 Boolean vector that we mine for association rules that ally uses and features. After discussing related work, we briefly summarize ontology induction and then focus on the vector representation, extraction, and use-centric mining. We report on an experiment using over 300 Epinions customer reviews for digital cameras [1].

2. RELATED WORK

This paper presents preliminary work on leveraging an ontology to extract features and uses, thereby transforming customer reviews into vectors suitable for data mining. Accordingly, we build upon three bodies of related research: Ontology-based information extraction (IE), extraction as classification, and text-mining on customer reviews.

In ontology-based IE, every extraction attribute is associated with a concept frame that captures related concepts, constants, and keywords that augment extraction patterns [5]. Past work on ontology-based IE assumed pre-existing ontologies; we follow research on automated ontology generation [10, 12] to induce a keyword lexicon entirely from the customer reviews.

Instead of explicitly extracting feature values, we use the ontology to build patterns that classify whether (the extracted string from) a review refers to a particular feature. Kushmerick similarly extracts/classifies fields from business card entries and change-of-email-address notices [9]. The approach is distinct from efforts that combine extraction and classification by reducing extraction to the problem of classifying whether tokens are extraction delimiters [4, 6].

One thread of research applying IE to customer reviews focuses on the structured portions of review text. Chueng et al. collected the product rankings from all reviews written by an individual customer to construct that person's profile [3]. As noted earlier, we attempt to model the product category, not individual consumers.

A second body of research on customer reviews applies natural language processing (NLP) techniques to extract sentiments and features from unstructured comments [8, 13]. Research on sensitivity analysis attempts to identify the tenor of a comment (e.g. whether a customer is pleased, angry, etc.). We believe that sentiment is orthogonal to our task. We are interested in identifying which features are used to evaluate a product's performance, regardless of whether the product performs well or poorly with respect to those measures.

Related work on feature extraction from customer reviews combines NLP with frequent item set mining. The theory is that frequently occurring noun phrases (NPs) across a corpus of reviews are likely to constitute product features [8]. We believe that this approach is complementary to our ontology-based approach. Hu and Liu process reviews for each product separately [8]; an ontology could reconcile variations in feature terminology across manufacturers, extending the NP count to all reviews in a product category. Furthermore, our ontology induction process (touched on below) incorporates perceptual features (e.g. compactness, ease-of-use) and other characteristics for feature extraction that are not NPs and hence would be overlooked by strict NP parsing.

3. USE-CENTRIC MINING OF CUSTOMER REVIEWS

The goal of use-centric mining is to develop a model of the customer's perspective on a specific product category. We do this in a multi-step process: (1) develop a category-specific ontology of product features and uses. (2) extract/classify features and uses to transform each review into a Boolean vector representation. (3) mine the vectors and filter the resulting association rules.

3.1. Develop the category-specific ontology. Although ontology induction is not the focus of this paper, it is worth noting both for the experiments and the subsequent discussion. We actually use two ontologies: one for features and one for uses. For each, we develop a three step induction process: (1) automatically generate a list of terms. (2) cluster and label the terms by semantic concept. (3) automatically generate extraction parameters for each concept class.

For the features ontology, terms are automatically assembled in an unsupervised manner from two sources. First, we retrieve the list of technical specifications for every product in the category from its respective "product detail" page and extract each product attribute. In addition, every Epinions.com review includes a short list of pros and cons (see Fig 1a). We parse every pro and con phrase, ignoring the distinction between positive and negative sentiment (see comments below). Terms are normalized using shallow linguistic processing (e.g. stop-words, capitalization), ordered by frequency count, and then clustered and named by hand. For each cluster, after removing redundant tokens (plural word forms etc.) every constituent phrase is compiled into a string-literal regular expression (RE) to be used in extraction/classification. Thus, an occurrence of the string literal *handy* or of *pocket-size* within a digital camera customer review is interpreted as an occurrence of the concept *compactness* within that review (see Fig 1b). The usage ontology is currently constructed entirely by hand. Student researchers read every review and the corresponding clusters were created manually. Our future plan is to apply supervised learning extraction/classification to the task.

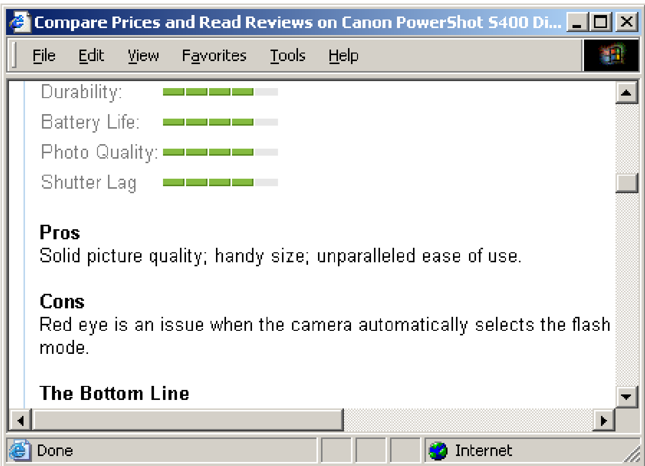


Figure 1a. Pros and Cons in an Epinions Review

```
<!-- Ontology entries for product
and perceptual features>
<Concept>
  <CName>Compactness</CName>
  <TermList>
   <Term>handy size</Term>
  …
  </TermList>
  <ClassifyExtract>
   <RegExp>'handy'</RegExp>
   <!--handy size—redundant term>
   <RegExp>'pocket-size'</RegExp>
  …
  </ClassifyExtract>
```

Figure 1b. Constructing an ontology entry

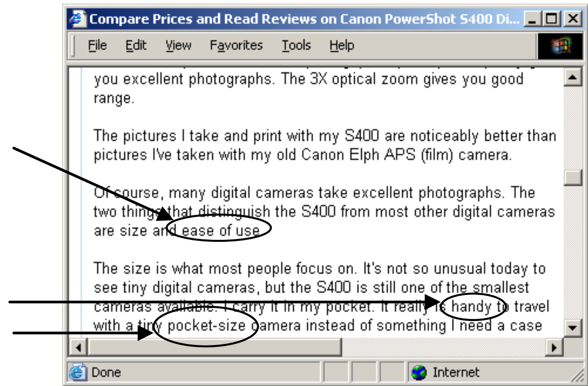
3.2. Extract/classify features and uses. We model individual product reviews as a vector of Boolean variables. Every vector has exactly $n+m$ attributes corresponding to the n product features and m product usages defined in the respective ontologies. A review is translated into a vector by calling *BuildVector* on the text of the review and an ontology (see Fig 2). Because we have two ontologies, we call *BuildVector* twice for each review and join the resulting usage- and feature-vectors. Intuitively, *BuildVector* processes a review by attempting to extract every feature and use mentioned in the review text. Then, because of possible semantic differences between different manufacturers and/or reviewers, we need to normalize the extracted terms to the ontology terms. For example, different reviewers may use *handy-sized* or *pocket-size* but both refer to the ontology term *compactness*.

```
BuildVector(review, ontology)
   Declare vector
   For each concept in the ontology
      vector[index(concept)] = 0
      For each extract/classify re representing the concept
        if re.match; vector[index(concept)] = 1; break
    Return (vector)
```

Figure 2. BuildVector

Because our ultimate objective is to map features and uses onto a review vector, we combine extraction and classification in a single step. *BuildVector* calls extract/classify for each concept cluster in the ontology (i.e. each vector index). As noted earlier, every ontology concept is associated with a set of string-literal regular expressions (see Fig 1b). Thus, we treat each RE as a classification rule for its corresponding concept. Given a customer review, for each ontology concept, evaluate every corresponding RE on the review text. If any RE evaluates to true (i.e. if there is a match), we set the corresponding review-vector index to true. Figure 3 depicts two RE matches for *compactness* and one RE match for *ease of use*. We envision more sophisticated and/or efficient extraction/classification routines in the future. In this way, even simple RE search reduces the problem of feature and usage extraction to one of classification.

```
<!-- Ontology entries><Concept>
<CName>Ease of use</CName> …
<ClassifyExtract>
  <RegExp>'ease[ -]of[ -]use'
  </RegExp>
  <RegExp>'easy[ -]to[ -]use'
  </RegExp>
…</Concept>
…<CName>Compactness</CName>
…<ClassifyExtract>
  <RegExp>'handy'</RegExp>
  <RegExp>'pocket-size'</RegExp>
…</Concept>
<Concept> …
```



Review Vector	Ease of use	Compactness	Zoom	Lens Cover	…
ProdID-ReviewerID	1	1	0	0	

Figure 3. Extract/classify to transform a review into a vector

3.3. Mine and filter rules. We explicitly selected a Boolean vector representation for modeling customer reviews to facilitate conventional frequent item-set and association rule mining [7]. A hazard of traditional association rule mining is the need to filter thousands of irrelevant rules. Apart from setting thresholds on support and confidence, our objective of use-centric mining imposes an inherent filter; we wish to align uses and features. Colloquially speaking, suppose a customer seeks a digital camera for her upcoming trip to London. For this customer, we mine the set of all digital camera reviews for those camera features that follow from the corresponding use (e.g. travel). Association rules that align feature to feature or have features in the rule body and usages in the rule head are of no use to our customer. Our customer knows her prospective use and would like to find features. Thus, we automatically filter to consider only rules where the rule body is limited to one or more usages.

4. EXPERIMENTAL EVALUATION

We introduce the concept of use-centric mining of customer reviews. Of the three steps in use-centric mining, we focus on extraction-as-classification to map reviews into vectors and association rule mining over the vectors. Evaluation proceeds in two stages. First, we assess the correctness of ontology-based extraction/classification by assessing automatically extracted data against hand-labeled data. Second, because our ultimate objective is to generate association rules over the review vectors, we generate rules from the extracted vectors, filter those rules as indicated in §3.3, and report standard statistical measures (support and confidence). As an additional, informal test, we ask whether the rules conformed to our subjective intuitions. In other words, would a lay reader with some limited knowledge subjectively concur with the suggested use-feature relationships.

We begin with a test set of over 300 digital camera reviews from Epinions.com[1]. Collectively, the reviews represent approximately 300 reviewers (some reviewers review multiple products) and 13 different models. The cameras are produced by 9 different manufacturers, range in resolution from 2MP (mega pixels) to more than 6MP, and are priced from below $100 to over $1000.

Usage and feature terms are collected into ontologies over the sample data as indicated in §3.1. Clustering results in 40 physical and perceptual product features and 28 use classes comprising a total of 68 ontological concept classes, each with its associated list of string literal extract/classify RE. We compare the results of *BuildVector* to its hand-coded counterpart and its hand coded

To evaluate our extraction/classification routine, researchers manually coded each review. The manual coding serves as a reference for comparison. Thus, for every concept in the ontology, we calculate the precision: *# reviews <u>correctly</u> classified as <u>discussing</u> the concept ÷ # reviews that were classified (correctly and incorrectly) as discussing the concept* and recall: *# reviews <u>correctly</u> classified* *as discussing the concept ÷ total # reviews that actually do discuss the concept*. These metrics, along with the F-measure, the weighted, harmonic mean of precision and recall [11], are standard measures used in information extraction. In addition, we report sensitivity, specificity and accuracy, traditional measures used in text classification. Sensitivity and recall are defined identically, specificity is the *# reviews that are correctly classified as not discussing the concept ÷ total # reviews that do not discuss the concept*, and accuracy is a weighted sum of sensitivity and specificity [7]. An excerpt is included as Table 1.

Ontology concept	F (α=0.5)	Precision	Recall/Sensitivity	Specificity	Accuracy
Ergonomic	0.6049	0.5506	0.6712	0.8305	0.7929
Compactness	0.6043	0.5045	0.8760	0.3833	0.5890
Download/transfer	0.6220	0.4769	0.8942	0.5024	0.6343
Batteries/power	0.8449	0.8233	0.8676	0.6381	0.7896

Table 1. Excerpted performance measures for extract/classify

Even perfect performance in extraction and classification is of little use if the resulting model of the product category space is not descriptive. We build a model of frequent item sets and association rules from our automatically extracted review vectors using apriori as implemented by Christian Borgelt and available in the Gnome Data Mine Tools package [2]. We first filtered the resulting rule set by support greater than 10% and a confidence greater than 80%, generating 27,695 rules. Further constraining rule bodies to only product usages pruned the result to 91 rules (see §3.3). We subjectively separated the rules into those we would intuitively expect and those that might be less obvious in a colloquial sense (see Table 2). For example, one would expect that a beginning photographer or someone purchasing their first camera would be interested in ease-of-use. Likewise, customers purchasing cameras for travel would naturally consider compactness; that landscape and nature photographers would express similar interests is less expected modulo the independence of travelers and either landscape or nature photography.

Expected rules			Less obvious rules		
Rule	Supp%	Conf %	Rule	Supp%	Conf%
beginner ÷ ease of use	27.0	86.0	Web/email ÷ video	11.9	82.6
travel ÷ compact	22.6	87.8	landscape ÷ compact	18.9	90.9
print/enlarge ÷ download/transfer	17.0	90.0	nature ÷ download/transfer	11.9	86.4
Web/email ÷ download/transfer	13.2	91.3	nature ÷ compact	12.6	90.9

Table 2. Excerpted association rules from mining review vectors

5. DISCUSSION

While the concept of aggregating reviews may have appeal, there are some limitations to the data set. Reviews are a self-selected subset of positive data (i.e. customers who both purchase and choose to review). Moreover, reviewers may exhibit particular biases; there is no reason to believe that reviewers always tell the truth or discuss features and uses in any systematic manner. However, the community of users provides a distinct perspective to prospective purchasers. We believe that aggregation over a sufficiently high review count may compensate for individual biases. Moreover, for our purposes, sentiment (e.g. whether the reviewer is pleased or not) and even reviews that compare multiple products should not confound. For example, if a review indicates that one camera is better for travel than a competing product because of size, the important relationship is travel and size.

Though our ontology-based approach requires a separate ontology for each product category, we believe that the body of customer reviews, pre-clustered by category, provides an ideal knowledge base for experiments in automated ontology induction. Tightly constraining the application to product recommendations may further simplify automated induction. However, ontology-based extraction is vulnerable to NLP problems like negation and POS contexts. For example, we treat every occurrence of 'trip' as a reference to the 'vacation' concept class whether it is used as a Verb rather than a Noun or if the reviewer wrote "I never go on trips." More sophisticated Key-Word-In-Context (KWIC) classification may help as might complementary work combining sentiment terms and POS tags[8]. Ultimately, our final objective, rule-mining, does not depend upon the ontology. Thus, other unsupervised or even supervised IE techniques are an option.

Although the flat state of our current ontology lends itself to a Boolean vector representation, many of the concept classes are naturally hierarchical (e.g. travel v. travel to London or travel v. travel for work v. travel for pleasure etc.) More powerful categorical or multi-level mining strategies are needed to address finer-grained distinctions.

Finally, we should consider additional rule filtering techniques. The chi-squared statistic can account for the association rule assumption of variable independence (e.g. nature, landscape, and travel). Beyond support and confidence, however, we have no intuitions about objective, statistical measures of "interestingness". While the simple filter constraining rule-bodies limited the rules for digital cameras, the filter may be less effective in other domains. Moreover, though we are interested in relating uses to features, association rules between features may also help define the product category space.

6. CONCLUSIONS

In this paper, we present an exploratory study introducing the concept of use-centric mining. Of the three steps in use-centric mining, we focus on extraction-as-classification to transform reviews into vectors and then mine for association rules over the vectors. We offer three principal contributions: (1) the concept of aggregating over all reviews to model a product category space rather than individual consumers or distinct products (2) inducing an ontology to build classifiers for extraction (3) association rule-mining over a vector representation created through multi-dimensional classification. We are currently coding additional camera reviews to expand the test set and collecting reviews in several other product categories for additional trials of our aggregation intuition. At this stage, we do not know whether our techniques will work in other domains. Additional product categories may also provide some intuitions on what types of features and uses might be most suited to our analytical approach. For these trials, we are incorporating POS tags and KWIC into our extract/classify patterns. Finally, we are exploring more sophisticated, multi-level mining models for finer-grained usage and feature analysis. Ultimately, our goal is to develop automated tools to support consumer purchasing.

7. REFERENCES

[1] "Epinions.com Customer Reviews," accessed: 5 July 2004, 2004, http:// www.epinions.com/ Digital_Cameras.

[2] C. Borgelt and R. Kruse, "Induction of Association Rules: Apriori Implementation," presented at 15th Conf on Computational Statistics (Compstat), Berlin, Germany, 2002.

[3] K.-W. Cheung, J. T. Kwok, M. H. Law, and K.-C. Tsui, "Mining customer product ratings for personalized marketing," *Decision Support Systems*, vol. 35, pp. 231-243, 2003.

[4] W. Cohen and W. Fan, "Learning Page-Independent Heuristics for Extracting Data from Web-pages," presented at WWW8, Toronto, Canada, 1999.

[5] D. W. Embley, D. M. Campbell, Y. S. Jiang, S. W. Liddle, Y.-K. Ng, D. Quass, and R. D. Smith, "Conceptual-model-based data extraction," *Data & Knowledge Engineering*, vol. 33, 1999.

[6] A. Finn and N. Kushmerick, "Multi-level Boundary Classification for Information Extraction," presented at ECML/PKDD 2004, Pisa, Italy, 2004.

[7] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*. San Francisco: Morgan Kaufmann, 2001.

[8] M. Hu and B. Liu, "Mining and Summarizing Customer Reviews," KDD04, Seattle, WA, 2004.

[9] N. Kushmerick, E. Johnston, and S. McGuinness, "Information Extraction by Text Classification," presented at IJCAI-01 Workshop on Adaptive Text Extraction and Mining, Seattle, WA, 2001.

[10] B. Liu, C. W. Chin, and H. T. Ng, "Mining Topic-Specific Concepts and Definitions on the Web," WWW 2003, Budapest, Hungary, 2003.

[11] J. Makhoul, F. Kubala, R. Schwartz, and W. R, "Performance measures for information extraction," presented at DARPA Broadcast News Workshop, Herndon, Virginia, 1999.

[12] G. Modica, A. Gal, and H. Jamil, "The Use of Machine-Generated Ontologies in Dynamic Information Seeking," presented at CoopIS 2001, Trento, Italy, 2001.

[13] T. Nasukawa and J. Yi, "Sentiment Analysis: Capturing Favorability Using Natural Language Processing," presented at K-CAP`03, Sanibel Island, Florida, 2003.