Aspect Selection for Social Recommender Systems

Yoke Yie Chen[1], Xavier Ferrer[2,3], Nirmalie Wiratunga[1], and Enric Plaza[2]

[1]IDEAS Research Institute, Robert Gordon University, Aberdeen, Scotland
{y.y.chen,n.wiratunga}@rgu.ac.uk
[2]Artificial Intelligence Research Institute (IIIA-CSIC)
Spanish National Research Council (CSIC)
Campus UAB, Bellaterra, Catalonia (Spain)
[3]Universitat Autònoma de Barcelona
Bellaterra, Catalonia (Spain)
{xferrer,enric}@iiia.csic.es

Abstract. In this paper, we extend our previous work on social recommender systems to harness knowledge from product reviews. By mining product reviews, we can exploit sentiment-rich content to ascertain user opinion expressed over product aspects. Aspect aware sentiment analysis provides a more structured approach to product comparison compared to one that is not. However, aspects extracted using NLP-based techniques remain too large and lead to poor quality product comparison metrics. To overcome this problem, we explore the utility of feature selection heuristics based on frequency counts and Information Gain (IG) to rank and select the most useful aspects. Here an interesting contribution is the use of Top ranked products from Amazon to formulate a binary classification over products to form the basis for the supervised IG metric. Experimental results on three related product families (Compact Cameras, DSLR Cameras and Point & Shoot Cameras) extracted from Amazon.com demonstrate the effectiveness of incorporating feature selection techniques for aspect selection in recommendation task.

Keywords: social recommenders, online reviews, feature selection

1 Introduction

Recommender systems provide a ranked list of products to assist user purchase needs. With content-based systems products similar to those that have been liked by the user are ranked over others [6]. Central to this is the ability to establish similarity between the target 'liked' product and the rest. How best to represent products to achieve effective product comparison is an area of interest to Case-Based Reasoning (CBR) in the context of recommender systems [10]. Increasingly effort is being focused on incorporating knowledge from product reviews into product representation. In particular, the rich information embedded in product reviews permit recommender systems to learn implicit preferences of users by considering product aspects (also called features) mentioned in product reviews [1].

Our previous work proposed a social recommender system used two social media knowledge sources: online product reviews and purchase preferences. As a result, recommendation was improved by the combination of aspect based sentiment analysis with preference knowledge [2]. More importantly, we showed that recommendations generated based on aspect-based sentiment analysis to be far superior to one that is agnostic of aspects. However, most NLP-based aspect extraction techniques rely on POS tagging and syntactic parsing which are known to be less robust when applied to informal text. As a result, it is not unusual to have a large numbers of spurious content to be extracted incorrectly as aspects. Methods to infer aspect importance and thereafter rank them for selection are needed to achieve a manageable aspect subset size.

Feature selection is known to enhance accuracy in supervised learning tasks such as text classification by identifying redundant and irrelevant features [13]. In this paper we address the problem of selecting important aspects using feature selection heuristics. Specifically, we explore two feature selection approaches to evaluate aspect usefulness: Information Gain(IG) and aspect frequency. In our solution, we capitalise on Top ranked products from Amazon to formulate a binary classification over products to form the basis for the supervised IG metric. In addition, we investigate the transferability of selected aspects from a particular product family (e.g. Compact Cameras) to other related product families (e.g. DSLR Cameras and Point & Shoot Cameras).

The rest of the paper is orrganised as follows: In Section 2 we present related research. Next in Section 3 we describe the process of aspect extraction and feature selection heuristics. Finally, evaluation results are presented in Section 4 followed by conclusions in Section 5.

2 Related Work

Recent work in social recommender systems utilise sentiment analysis as key features for product representation. An interesting ideas here is to compare products not simply on the basis of sentiment polarity (i.e. positive or negative sentiment scores) but on the basis of similar sentiment over product aspects. This then requires that aspects are extracted from product reviews before they can be associated with polarity scores [4, 5]. The fundamental to this comparison is the relevance of the product aspects extracted from online reviews.

Frequency of aspects is commonly used as a heuristic to rank and select the best aspects from product reviews. This frequency score can further be combined with sentiment scores to bias these rankings when the task involves opinionated content [14]. Equally frequency can also be combined with similarity knowledge whereby aspects that contribute most to product similarity computations are considered more relevant than those that do not [9].

Unlike frequency-based heuristics supervised selection heuristics have been successfully employed to reduce dimensionality and achieve significant gains in accuracy for text classification [12]. In this paper we explore how the supervised Information Gain (IG) heuristic can be adopted in the context of social recommenders to reduce the dimensionality of product aspects. Whilst Vargas-Govea et al [11] have also used a supervised selection method in the context of semantic based restaurant recommender systems they did so to identify influential contextual features using user rating values as the class label. Unlike with typical classification tasks where class labels are explicitly defined, in our work the notion of class and its boundaries need to be considered carefully to enable the application of IG for aspect selection.

3 Review based Product Recommendation

Central to a social recommender system is the source of opinionated content in the form of product reviews. As depicted in Figure 1 this source can be harnessed to generate a product ranking using the following three steps:

1. Extract product aspects from reviews and quantify the strength of sentiment over these aspects within the range of [-1,1];
2. select the best aspects according to a selection heuristics; and
3. generate recommendations using evidence from sentiment based strategies.

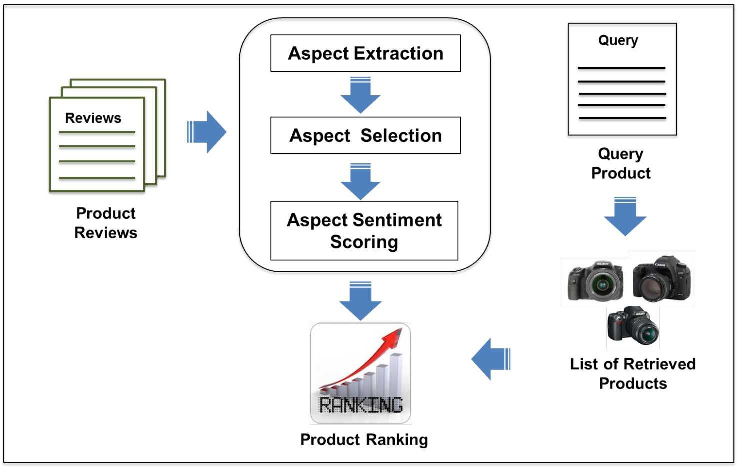


Fig. 1: Overview of social product recommender process.

3.1 Aspects Extraction from Product Reviews

Grammatical extraction rules [7] are used to identify a set of candidate aspect phrases from sentences. These rules operate on dependency relations in parsed sentences[1]. Figure 2 lists the rules that we have employed in this work. The rule conclusions contain the constructs that form the extracted aspect following rule activation. Here N is a noun, nn is a compound-noun, A an adjective, V a verb, h a head term, m a modifier. Candidate phrases include $\langle h, m\rangle$, $\langle N, A\rangle$, $\langle N, V\rangle$, $\langle h+N, m\rangle$ and $\langle N+h, m\rangle$. For each candidate, non noun (N) words are eliminated and the remainder forms the set of aspects.

$$
\begin{aligned}
&DP = \text{set of dependency pattern rules}\\
&\{\\
&\quad dp_1 : amod(N, A) \rightarrow \langle N, A\rangle,\\
&\quad dp_2 : acomp(V, A) + nsubj(V, N) \rightarrow \langle N, A\rangle,\\
&\quad dp_3 : cop(A, V) + nsubj(A, N) \rightarrow \langle N, A\rangle,\\
&\quad dp_4 : dobj(V, N) + nsubj(V, N') \rightarrow \langle N, V\rangle,\\
&\quad dp_5 : \langle h, m\rangle + nn(h, N) \rightarrow \langle N + h, m\rangle,\\
&\quad dp_6 : \langle h, m\rangle + nn(N, h) \rightarrow \langle h + N, m\rangle\\
&\}
\end{aligned}
$$

Fig. 2: Extraction rules.

3.2 Aspect Selection

Aspects extracted are not equally important and therefore are subjected to selection-based dimensionality reduction. Given a set of products, $\mathcal{P}$, a set of aspects, $\mathcal{A}$, that appear in online reviews, $\mathcal{R}$. A product p is represented as $\vec{x} = \{x_1, ...x_{|\mathcal{A}|}\}$ where x is binary valued and correspond to the presence or absence of an aspect $a \in \mathcal{A}$. The aim of feature selection is to reduce $|\mathcal{A}|$ to a smaller aspect subset size n by selecting aspects ranked according to the score assigned by the feature selection technique. The selected aspects then form a new aspect vector $\vec{x'}$ and a corresponding reduced aspects set $\mathcal{A}$' for product p, where $\mathcal{A}' \subset \mathcal{A}$ and $|\mathcal{A}'| \leq |\mathcal{A}|$. The algorithm used to rank aspects for selection is shown in Algorithm 1. Here $S = \{p_1, ..., p_w\}$ denotes the sample of products for training purpose.

Aspect Selection by Frequency Frequency of extracted aspects is calculated according to the number of times an aspect occurs over the set of reviews. Accordingly aspects are ranked based on their FrequencyRank scores computed as follows:

$$\textsc{FrequencyRank}(a_i) = \frac{f(a_i)}{\sum_{j=1}^{\mathcal{A}} f(a_j)} \tag{1}$$

[1] Sentences are parsed using the Stanford Dependency parser [3]

Algorithm 1 : Aspect Selection

n = aspect subset size
for each $a \in \mathcal{A}$ **do**
  Calculate aspect score using S
end for
Sort aspects based on frequency or IG scores
$\mathcal{A}' = \{a_1, ..., a_n\}$
return $\mathcal{A}'$

where aspect score $f(a_i)$ returns the relative frequency of an aspect a_i appearing in reviews $\mathcal{R}$. Here frequent occurrence of aspects in online reviews is perceived as important. However, frequency based approaches have a tendency to select general aspects such as "camera", "use" and "quality" which fail to provide sufficient context for product comparisons (see Table 3). Instead of relying simply on frequency what is required here is a heuristic that can discriminate between aspects that can discern between recommendation strategies. One such strategy discussed next considers whether recommended products appear in the top selling list or not.

Aspect Selection with Information Gain Features that are able to discriminate between classes are considered important in text classification. In the absence of predefined class labels we use a product ranking benchmark to derive class labels, whereby a rank position is used as a class boundary to separate top ranked from the rest of the products. Here we use a binary class such that c is either 0 meaning that the product is in the Top ranked set; or is 1 meaning it is not among the Top ranked products. In this way each of the products in the product sample, S, can be assigned a binary label. Accordingly we rewrite the product notation as a pair $(\vec{x}, c)$ where c is binary class label for p. Essentially, increasing the rank position that derives the class boundary for products will lead to a skewed class distribution; that is it will result in decreasing the number of products belonging to $c = 0$ whilst increasing the size of class c=1. Given this supervised context, the discriminative power of an aspect a given the classes is computed as follows:

$$\text{IG}(X, C) = \sum_{x \in 0,1} \sum_{c \in 0,1} P(X = x, C = c).log_2 \frac{P(X=x, C=c)}{P(X=x).P(C=c)} \quad (2)$$

3.3 Aspect Sentiment Scoring

Products that are favorably mentioned in reviews should ideally be ranked higher for recommendation. Given a target query product and its set of similar products we can rank these based on their sentiment (positive and negative) scores. Essentially such a product score is an aggregation of sentiment scores over the selected subset of aspects.

$$score(p_i) = \frac{\sum_{j=1}^{|\mathcal{A}'|} SentiScore(p_i, a_j)}{|\mathcal{A}'|} \tag{3}$$

Where the sentiment of the product p_i is associated with individual aspects a_j and $|\mathcal{A}'|$ is the aspect set for product p_i. Here, *SentiScore* of an aspect is derived from product reviews using SmartSA [8] and is computed as:

$$SentiScore(p_i, a_j) = \frac{\sum_{m=1}^{|\mathcal{R}_j^i|} SentiScore(r_m)}{|\mathcal{R}_j^i|} \tag{4}$$

where $\mathcal{R}_j^i$ is a set of reviews for product p_i related to aspect a_j and $r_m \in \mathcal{R}_j^i$.

4 Evaluation

Primary aim of our evaluation is to study the impact of aspect selection on recommendation quality. To do this, we evaluate how well the recommendation system works in practice on Amazon.com data. We conveniently use Amazon's product *Star-Ratings* as the benchmark ranking to derive a comparison metric based on rank improvement. A secondary aim is to explore the transferability of the aspects learned from a particular product family (e.g. Compact cameras) to other related product families (e.g. DSLR cameras and Point & Shoot cameras).

4.1 Amazon Datasets

We crawled 1179 Amazon products during September 2015 from three different Amazon Digital Cameras categories: Point & Shoot (PAS), Digital SLR (DSLR) and Compact System Cameras (COMPACT). The products extracted contain more than 100,000 different user generated reviews. Since we are not focusing on the cold-start problem, we use 1st January 2010 and less than 15 reviews as the pruning factor for the three product families. Finally, any synonymous products are united leaving us data for 98 COMPACT, 102 DSLR and 93 PAS products (see Table 1).

The aspect extraction algorithm extracted 300-450 unique aspects for DSLR, PAS and COMPACT. On average, each product is defined by 220 different aspects, with standard deviations of 115, 110 and 86 aspects for DSLR, COMPACT and PAS cameras respectively. Importantly, more than 50% of the products shared at least 100 different aspects with other products of the same category, while almost 30% shared more than 150 aspects (more than 200 for COMPACT) on average. The fact that there are many shared aspects between products of the same category is reassuring for product comparison.

Category	**DSLR**	**Compact**	**PAS**
No. of Products	102	98	93
No. of Reviews	7451	6349	11,202
Aspects Mean (Std. Dev.)	226.78 (115.40)	267.25 (110.28)	186.39 (86.72)
No. of Different aspects	438	424	308

Table 1: Statistic of Amazon DSLR, Compact and PAS Camera Datasets

4.2 Evaluation Metrics

In the absence of a manual qualitative estimate of recommendation or access to user specific purchase trails, we derived approximations from the Amazon data we had crawled. For this purpose, using a *leave-one-out* methodology, the average gain in rank position of recommended products over the left-out query product is computed relative to a benchmark product ranking for each of the three categories DSLR, COMPACT and PAS.

$$RankImprovement\%(RI) = \frac{\sum_{i=1}^{n=3} benchmark(P_q) - benchmark(P_i)}{n * |\mathcal{P} - 1|} \quad (5)$$

where n is the number of the top ranked products and *benchmark* returns the position on the benchmark list. The greater the gain over the query product the better. For instance, suppose the query product is ranked 40th on the benchmark list of 81 unique products $\mathcal{P}$, and the recommended product is ranked 20th on this list, then the recommended product will have a relative benchmark RI of 25%.

We generated three benchmark lists according to Amazon's *Star-Rating*s of the three camera families we crawled. In cases where two or more products had the same star-rating, the products were ordered by the number of comments.

4.3 Ranking Strategies

The retrieval set of a query product consists of products that share a similar number of k aspects such that higher values of k denote lower number of products retrieved. This retrieval set is ranked using the sentiment-based recommendation strategies presented in Section 3.3. Central to this strategy is the selection of aspects using the following feature selection methods:

- BASE: recommend using aspect sentiment analysis with all aspects (see Equation 3);
- FREQUENCYRANK (FR): same as BASE but only considering a subset of aspects selected by FR (see Equation 1);
- INFORMATIONGAIN (IG): same as BASE but only considering a subset of aspects selected by IG (see Equation 2).

The experiments were performed using 5 fold cross validation. To assess the transferability of the important aspects learned from different camera family, we apply the selected aspects learned from a particular family to other two related product families in ranking the products.

4.4 Recommendation Performance using IG

The objective of using feature selection technique is to exploit important aspects generated by these techniques to rank products. We assess IG's effect on recommendation performance by manipulating class sizes and aspect subset sizes. Figure 3 shows for increasing class size the performance of each product family in terms of average RI on benchmark *Star-Rating*. Here, the average RI is computed using different k shared aspects where k is from 0 to 240. The result shows that a small class size is seen to lead to better performance. Note from Section 3.2 that class size relates to the top products rank position being used to create class boundary separation. For instance, we observed in Figure 3 that the performance of COMPACT improved from 5% to 10% but starts to fall after 10%. Similar observations can be made on DSLR and PAS where their performance starts to drop after 15%.

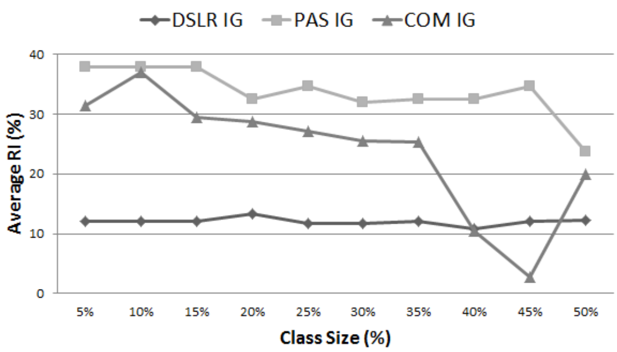


Fig. 3: Average RI for all Products at Different Class Size

Figure 4 presents average RI for all product families when selecting aspects at different aspect subset size using IG at class size 10%. In general, the average RI of all product families is at its best when 90 aspects were selected and remains constant for $n > 90$. It is interesting to note that when $n < 90$, products are compared using a smaller number of aspects. For example, only 40% of PAS and COMPACT contain more than 25 aspects in the aspect subset size of 50. This explains the fluctuations in average RI for both families when considering low values of n. Based on the observations in both experiments, from this point onwards we use fixed aspect subset size $n = 90$ and a class size of 10% for the rest of the experiments.

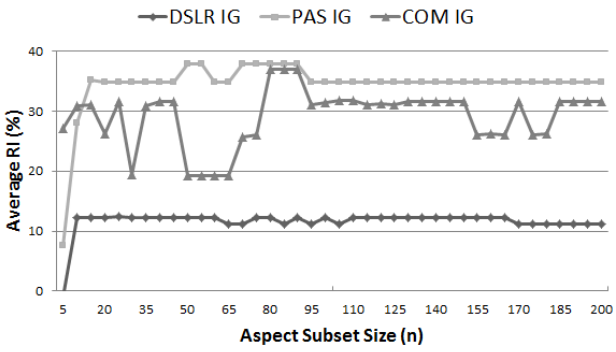


Fig. 4: Average RI for all Products at Different Aspect Subset Size

4.5 Comparison of Feature Selection Techniques

The graphs in Figures 5, 6 and 7 illustrates the results of our comparison using RI at increasing k number of shared aspects. An overall view of these graphs shows that IG performs best for all three product families, being the results obtained in DSLR only slightly better than FR. However, we observed the RI of IG is 15% more than BASE on average, obtaining an absolute RI of more than 40% for PAS category. This means for every query product over a set of 90 products, we are able to recommend a better product ranked 40 positions higher on average. It is also worth pointing out that the performance of FR improves the recommendations of all three categories at 5% on average compared to BASE. These results show that selecting a subset of aspects which are important provides a significant improvement on recommendation.

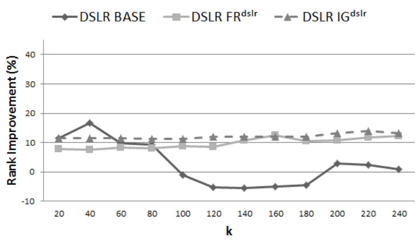


Fig. 5: RI with Aspect Selection on DSLR.

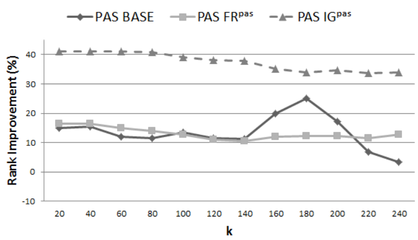


Fig. 6: RI with Aspect Selection on PAS.

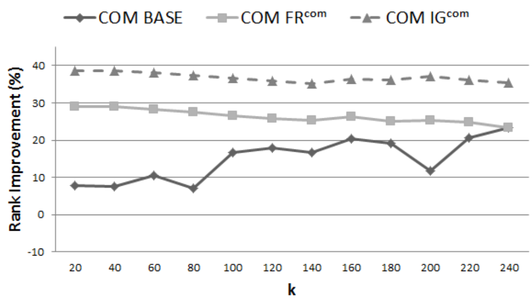


Fig. 7: RI with Aspect Selection on COMPACT.

4.6 Similarity of Product Families

In Table 2, we studied the similarity of product aspects between the three related product families by computing the Jaccard similarity coefficient between the sets of aspects of each category. Furthermore, we computed the frequency of each aspect over the three camera families and created a ranking of most frequent aspects for each family (see Table 3) and applied Spearman rank correlation coefficient to compare those ranked lists of aspects. As can be observed, DSLR and COMPACT share a similar set of aspects with a 0.72 Jaccard coefficient (even higher when considering top 20 products), whilst the set of aspects used in PAS is slightly different (with a Jaccard coefficient of 0.61 between PAS and DSLR, and 0.58 between PAS and COMPACT). Furthermore, the Spearman rank correlation coefficient value shows that the aspects shared between categories have similar frequencies values. For instance aspect *lens* in category DSLR is the second most used aspect whilst it occupies the third position in COMPACT (both categories have a 0.87 Spearman rank correlation).

	All Aspects		
	DSLR+Comp	DSLR+PAS	Comp+PAS
Jacc.	0.72	0.58	0.61
Spear.	0.87	0.75	0.76
	Top 20 Aspects		
Jacc.	0.81	0.60	0.66
Spear.	0.80	0.62	0.64

Table 2: Aspect Similarity for Different Camera Families

DSLR	**Compact**	**PAS**
camera	camera	camera
use	lens	picture
lens	use	use
picture	focus	photo
video	picture	video
focus	quality	quality
time	image	zoom
shoot	time	battery
image	photo	time
quality	shoot	shot

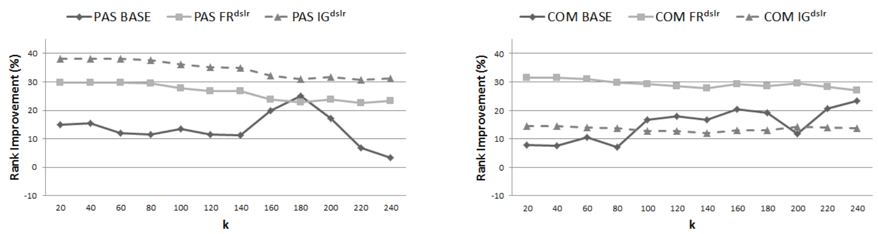
Table 3: Top 10 Most Frequent Aspects by Product Families

4.7 Transferability of Aspects

In Table 2, we observed the product aspects from three related product families have some degree of similarity. Here, we assess the transferability of the aspects observing if important aspects learned from one product family are able to improve recommendation performance on other product families. Figures 8-10 show the RI for three product families using FR and IG in aspect selection. Here PAS $\mathrm{FR}^{\mathrm{DSLR}}$ indicates that results presented correspond to FR strategy for PAS using DSLR selected aspects. Similarly, PAS $\mathrm{IG}^{\mathrm{DSLR}}$ indicates PAS results using DSLR aspects selected by IG.

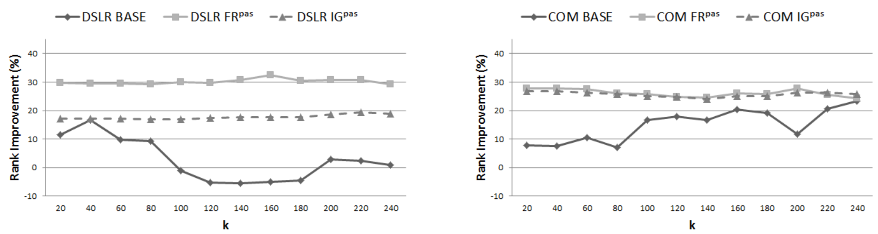
The benefit of aspects transferability can be observed when FR is used in aspect selection. For instance, Figure 9a and 10b shows FR provides significant improvements in recommendation for DSLR and PAS respectively. Furthermore, we observed that COMPACT $\mathrm{FR}^{\mathrm{DSLR}}$ (Figure 8b) obtains similar RI to COMPACT $\mathrm{FR}^{\mathrm{COMPACT}}$ (Figure 7), indicating the selected aspects of both families are similar. This result is expected given high aspects correlation between the frequent aspects of DSLR and COMPACT. The result obtained using IG is mixed. One explanation for its poor performance is that the product families does not share similar subset of aspects, resulting a dropped in RI of IG in Figure 8b and 10a. This indicates that aspects selected by IG are domain dependent as such provide little benefits to other product families.

The high transferability of the aspects using FR suggest that general aspects are suitable to be used in recommending cameras products. However, we observed that not all product families benefit from transfer of aspects. For instance, COMPACT does not benefit from aspects learned from other products. Essentially, best results are achieved when domain dependent aspects are learned using IG in COMPACT and PAS.



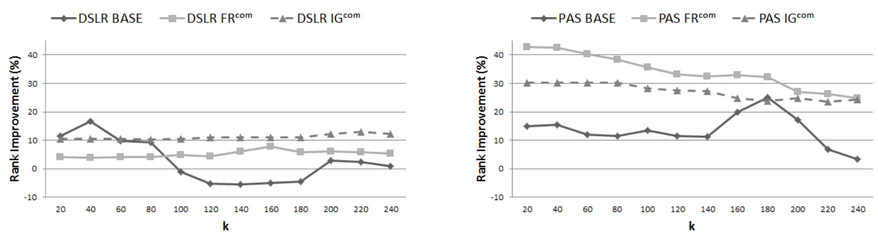
(a) PAS (b) COMPACT

Fig. 8: RI using Transferred Aspects from DSLR



(a) DSLR (b) COMPACT

Fig. 9: RI using Transferred Aspects from PAS



(a) DSLR (b) PAS

Fig. 10: RI using Transferred Aspects from COMPACT

5 Conclusion

In this paper we extended our previous work on social recommender systems to harness knowledge from product reviews, and explored the utility of frequency based approach and supervised Information Gain to rank and select the most useful aspects for recommendation. The benefits are demonstrated in a realistic recommendation setting using benchmarks generated from *Star-Rating*. We confirmed that aspect selection using feature selection techniques help improved recommendations of the three datasets; the best results are obtained using Information Gain when considering only a small subset of aspects. On the other hand, we presented how frequency based aspect selection technique are transferable between product families and that leads to better recommendation performance. However, better result are achieved when using domain dependent aspects. Our results show that Information Gain is promising in identifying important aspects and improve recommendations, but further work is needed to explore other feature selection techniques such as mutual information and the Chi-squared statistic.

Acknowledgments. This research has been partially supported by AGAUR Scholarship (2013FI-B 00034) and NASAID (CSIC Intramural 201550E022).

References

1. Li Chen, Guanliang Chen, and Feng Wang. Recommender systems based on user reviews: the state of the art. *User Modeling and User-Adapted Interaction*, pages 1–56, 2015.
2. Yoke Yie Chen, Xavier Ferrer, Nirmalie Wiratunga, and Enric Plaza. Sentiment and preference guided social recommendation. In *Case-Based Reasoning Research and Development*, pages 79–94. Springer, 2014.
3. M. De Marneffe, B. MacCartney, C. Manning, et al. Generating typed dependency parses from phrase structure parses. In *Proc. of Language Resources and Evaluation Conference*, pages 449–454, 2006.
4. R. Dong, M. Schaal, M. OMahony, K. McCarthy, and B. Smyth. Opinionated product recommendation. In *Inter. Conf. on Case-Based Reasoning*, 2013.
5. Jeff Huang, Oren Etzioni, Luke Zettlemoyer, Kevin Clark, and Christian Lee. Revminer: An extractive interface for navigating reviews on a smartphone. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*, pages 3–12. ACM, 2012.
6. Pasquale Lops, Marco De Gemmis, and Giovanni Semeraro. Content-based recommender systems: State of the art and trends. In *Recommender systems handbook*, pages 73–105. Springer, 2011.
7. S. Moghaddam and M. Ester. On the design of lda models for aspect-based opinion mining. In *Proc. Inter. Conf. on Information and Knowledge Management*, CIKM '12, 2012.
8. A. Muhammad, N. Wiratunga, R. Lothian, and R. Glassey. Contextual sentiment analysis in social media using high-coverage lexicon. In *Research and Development in Intelligent Systems*, pages 79–93, 2013.

9. Royi Ronen, Noam Koenigstein, Elad Ziklik, and Nir Nice. Selecting content-based features for collaborative filtering recommenders. In *Proceedings of the 7th ACM conference on Recommender systems*, pages 407–410. ACM, 2013.
10. Barry Smyth. Case-based recommendation. In *The adaptive web*, pages 342–376. Springer, 2007.
11. Blanca Vargas-Govea, Gabriel González-Serna, and Rafael Ponce-Medellın. Effects of relevant contextual features in the performance of a restaurant recommender system. *ACM RecSys*, 11, 2011.
12. Nirmalie Wiratunga, Ivan Koychev, and Stewart Massie. Feature selection and generalisation for retrieval of textual cases. In *Advances in Case-Based Reasoning*, pages 806–820. Springer, 2004.
13. Yiming Yang and Jan O Pedersen. A comparative study on feature selection in text categorization. In *ICML*, volume 97, pages 412–420, 1997.
14. Zheng-Jun Zha, Jianxing Yu, Jinhui Tang, Meng Wang, and Tat-Seng Chua. Product aspect ranking and its applications. *Knowledge and Data Engineering, IEEE Transactions on*, 26(5):1211–1224, 2014.