

Preference and Sentiment Guided Social Recommendations with Temporal Dynamics

Xavier Ferrer, Yoke Yie Chen, Nirmalie Wiratunga and Enric Plaza

Abstract Capturing users' preference that change over time is a great challenge in recommendation systems. What makes a product feature interesting now may become the accepted standard in the future. Social recommender systems that harness knowledge from user expertise and interactions to provide recommendation have great potential in capturing such trending information. In this paper, we model our recommender system using sentiment rich user generated product reviews and temporal information. Specifically we integrate these two resources to formalise a novel aspect-based sentiment ranking that captures temporal distribution of aspect sentiments and so the preferences of the users over time. We demonstrate the utility of our proposed model by conducting a comparative analysis on data extracted from Amazon.com and Cnet. We show that considering the temporal preferences of users leads to better recommendation and that user preferences change over time.

1 Introduction

Recommender systems traditionally provide users with a list of recommended items based on users preferences. The huge success of these systems in the retail sector demands innovative and improved recommendation algorithms. The dawn of the social web has created opportunities for new recommendation algorithms to utilise knowledge from such resources and so the emergence of social recommender system. These systems harness knowledge from user generated reviews to generate better recommendations by incorporating sentiment expressed in opinions to bias the recommendation list [5]. Similarly preference knowledge and temporal dynamics have also separately been applied to influence recommendations [18, 7].

Xavier Ferrer, Enric Plaza

Artificial Intelligence Research Institute (IIIA), Spanish Scientific Research Council (CSIC), Campus UAB, Bellaterra, Spain, e-mail: xferrer@iiia.csic.es, enric@iiia.csic.es

Yoke Yie Chen, Nirmalie Wiratunga

IDEAS Research Institute, Robert Gordon University, Aberdeen, AB25 1HG, e-mail: y.y.chen@rgu.ac.uk, n.wiratunga@rgu.ac.uk

Purchase choices are based on comparison of artefacts; which implicitly or explicitly involve comparison of characteristics or aspects of these artefacts. In particular a user’s purchase decision hints at the aspects that are likely to have influenced their decision and as such be deemed more important. Additionally it is also not unusual to expect that the criteria used for this comparison may also change with time. For example, in the domain of Cameras, the LCD display may have been an important aspect users were interested in the past but now this is given in almost every camera and so is likely to be an aspect of contention.

In recent work [1] we explored how preference knowledge can be captured and exploited within a social recommendation application. Our findings suggested that preference knowledge allows us to extract important aspects from reviews, in terms of those that are likely to have influenced the users’ purchase decision. However, would recency of reviews have an impact on aspect weights? How far back in time must we go before extracted weights improve the recommendations? Our main focus in this paper is to study temporal and preference context for social recommendations with a view to integrate these contexts with aspect-based sentiment analysis.

Our contribution is three-fold: firstly we demonstrate how sentiment distribution analysis can impact the quality of recommendations; and secondly show how a preference-based algorithm can be incorporated to derive rankings on the basis of preference relationships; and finally provide a formalism to combine sentiment and temporal information. Our results confirm that incorporating temporal information in aspect-based sentiment analysis is comparable to preference knowledge.

The rest of the paper is organized as follows: In Section 2 we present the background research related to this work. Next in Section 3 we describe aspect preference over time and how preference graphs can be generated by using a case study from Amazon.com. The process of aspect extraction and weight learning for sentiment analysis is presented in Section 4. Finally, our evaluation results are presented in Section 5 followed by conclusions in Section 6.

2 Related Works

Social recommenders recognise the important role of sentiment analysis of user reviews [5]. Instead of relying on user logs and sessions to model user preference [12], in this paper we infer aspect preferences from comparing the sentiment-rich content generated by users. However, extracting sentiment from natural language constructs is a challenge. Lexicons are often used to ascertain the polarity (positive or negative) and strength of sentiment expressed at word-level (e.g. SentiWordNet [6]). However sophisticated methods are needed to aggregate these scores at the sentence, paragraph and document level to account for negation and other forms of sentiment modifiers [15]. Increasingly aggregation is organised at the aspect level, since the distribution of a user’s sentiment is typically mixed and expressed over the aspects of the artefact (e.g. I love the *colour* but not too keen on *size*). Hu and Liu [8] propose an association mining driven approach to identify frequent nouns or noun phrases as aspects. Thereafter sentences are grouped by these aspects and sentiment scores assigned to each aspect group [13]. Whilst there are many other statistical

approaches to frequent noun extraction [16]; others argue that identifying semantic relationship in text provides significant improvements in aspect extraction [14]. Here we explore how semantic based extraction can be augmented by frequency counts.

Temporal dynamics is a crucial dimension for both content and collaborative approaches. Initial work on concept drift was applied to classification tasks with focus on optimising the learning time window [9]. More recently with recommender systems, temporal knowledge in the form of rules was used to predict purchase behaviour over time [2]. Similarly the temporal influence on changes in user ratings has been observed on the Netflix movies [10], and short and long-term preference changes on products [19]. The association of importance weights to aspects according to time is not new [4]. Here they generate aspect weights that are a function of time. Whilst our work also acknowledges the need for time-aware aspect weight learning, we exploit knowledge from both user review histories and preferences for this task.

3 Social Recommendation Model

An overview of our proposed process appears in Figure 1. The final outcome is a recommendation of products (hitherto referred to as artefacts) that are retrieved and ranked, with respect to a given query product. Central to this ranking are aspect weights, which are derived from two knowledge sources: sentiment rich user generated product reviews and preferences from purchased summary statistics. Generally preference knowledge is captured in a graph according to purchase behavior and reviews depending on recency will influence both weight extraction and ranking algorithms. Here, we are interested in exploring how aspects can be weighted. Accordingly alternative aspect weight functions, AW_i , will be explored by taking into account the time and preference contexts of aspects. These in turn will influence the sentiment scores assigned to each extracted aspect, AS_i . Therefore the final ranking of products is based on an aspect weighted sentiment score aggregation.

3.1 Time context

New and improved product aspects grow over time. While there are product aspects that are continuously improving, others stabilise when a majority of the products possess them. We can observe such trends on a sample of data collected from Amazon between 2008 and April 2014 (see Table 1). Here we summarise the statistics of aspect frequency and show the top 10 most frequently mentioned aspects. In 2008, *resolution* was most frequent, however, this aspect's importance diminished during the following years finally disappearing from the top 10 in 2010. On the other hand, aspects like *picture* or *battery* remain in the top 10 list. *Resolution* might have been an important aspect for a camera in 2008, but it is no longer a determinant as the majority of the SLR cameras are now equipped with high resolution. On the contrary, users keep seeking for better pictures or longer battery life in SLR cameras.

Another interesting observation is that the number of different aspects grows with time (see Table 1). This is not surprising as manufacturers introduce new product fea-

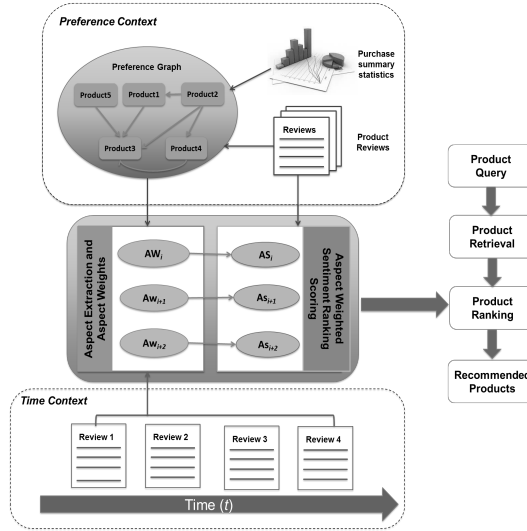


Fig. 1: Social recommendations with temporal dynamics

tures every year (e.g. HD video, creative image function etc.) to attract customers. This situation also explains the top aspect weights presented in Table 1 decreasing over time: a higher number of aspects per year means a lower aspect average frequency. Accordingly such situations challenge existing recommendation systems, calling for adaptive extraction and weighting algorithms that can leverage temporal information from product reviews for product recommendation.

Table 1: Top 10 aspects and #aspects ordered by frequency between years 2008 until April 2014

TOP 10 ASPECTS						
2008	2009	2010	2011	2012	2013	2014 until April
<i>resolution</i>	photographer	photographer	photographer	photographer	photographer	<i>picture</i>
photographer	<i>picture</i>	<i>picture</i>	<i>picture</i>	<i>picture</i>	<i>picture</i>	feature
software	feature	feature	feature	feature	feature	photographer
feature	battery	setting	battery	battery	setting	battery
<i>picture</i>	setting	battery	setting	setting	battery	setting
battery	reason	result	result	result	result	photography
setting	<i>resolution</i>	photography	photography	photography	photography	image
reason	result	reason	video	video	image	result
noise	photography	image	capability	image	video	time
result	software	video	image	time	time	quality
# ASPECTS						
315	557	672	721	866	934	896

3.2 Preference Context

Like time, the product preference behaviour of users also paints a picture about what aspects are likely to be important when making purchase decisions. For instance if we know which product was preferred over which other product then by comparing the product aspect differences we can infer a degree of aspect importance. To do this

we need to establish preference knowledge and thereafter quantify product aspect differences on the basis of sentiment.

We acquire preference knowledge from preference graph generated from viewed and purchased product pairs. The weight of an aspect is derived by comparing the sentiment difference between node pairs in the graph. A preference relation between a pair of products denotes the preference of one product over the other through the analysis of viewed and purchased product relationships. A preference graph, $G = (\mathcal{P}, \mathcal{E})$, is generated from such product pairs (see Figure 2). The set of nodes, $p_i \in \mathcal{P}$, represent products, and the set of directed edges, \mathcal{E} , are preference relations, $p_j \succ p_i$, such that a directed edge from product p_i to p_j with $i \neq j$ represents that, for some users, p_j is preferred over product p_i . For any p_i , we use \mathcal{E}^i to denote in-coming and \mathcal{E}_i for outgoing product sets.

Figure 2 illustrate a preference graph generated from a sample of Amazon data on *Digital SLR Camera*. The number of reviews/questions for a product is shown below each product node. It is not surprising that such products appear in Amazon’s *Best Seller* ranking (e.g. *B003ZYF3LO* is amongst Amazon’s top 10 list). In our recent work [1], we observed that the higher the number of incoming edges (quantity) from preferred products (quality), the more preferred, is for product, p_i . However we also observed that while our assumption is true with most studied products, it is not always the case that a product with higher number of incoming edges will also have a higher rank in Amazon’s *Best Seller*. This motivates the need to leverage further dimensions of knowledge sources such as sentiment from product reviews.

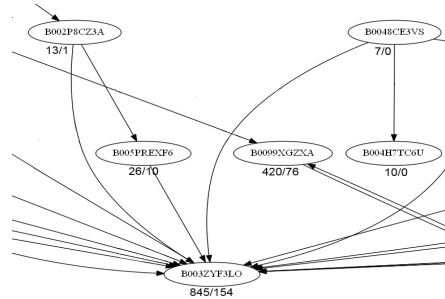


Fig. 2: Preference graph for Amazon SLR Camera

4 Aspect Weighted Sentiment-based Ranking

User generated product reviews contain user opinion in the form of positive and negative sentiment. Strength of sentiment expresses the intensity with which an opinion is stated with reference to a product [17]. We exploit this information as a means to rank our products, such that products ranked higher denote higher positive sentiment. *ProdSenti* of a product, p_i , given a set of related reviews \mathcal{R}^i , a weighted summation of sentiment expressed at the aspect level is computed as follows:

$$ProdSenti(p_i) = \frac{\sum_{j=1}^{|\mathcal{A}^i|} AW(a_j, t) * AS(p_i, a_j)}{|\mathcal{A}^i|} \quad (1)$$

$$AS(p_i, a_j) = \frac{\sum_{m=1}^{|\mathcal{R}_j^i|} SentiScore(r_m)}{|\mathcal{R}_j^i|} * (1 - Gini) \quad (2)$$

Where \mathcal{R}_j^i is a set of reviews for product p_i related to aspect a_j , $r_m \in \mathcal{R}_j^i$ and AW is a function of a_j 's weight over time t . $SentiScores$ are generated by the SMARTSA system [15]. The negative and positive strength is expressed as a value in the range [-1:1]. It allows the sentiment of product, p_i , to be associated with individual aspects $a_j \in \mathcal{A}^i$ where $\mathcal{A}^i \subseteq \mathcal{A}$. The aspects of a product are extracted by using Algorithm 1. These sentiment instantiated values for product aspects allows the comparison of product pairs in terms of sentiment. We exploit this further to derive aspects weights based on preference and sentiment knowledge. We use Gini index [20] to acknowledge higher sentiment scores to an aspect when there is consensus on the distribution of the sentiment and otherwise penalised accordingly.

4.1 Aspect Extraction

Grammatical extraction rules [14] are used to identify a set of candidate aspect phrases from sentences. These rules operate on dependency relations in parsed sentences¹. Rules involving negation are not included because our SMARTSA system already takes this into consideration when generating sentiment scores.

Algorithm 1 : Aspect Selection by Dependency Patterns

```

1: INPUT:  $S$  = sentences
2: for all  $s_j$  in  $S$  do
3:    $g = \text{grammaticalRelations}(s_j)$ 
4:    $candidateAspects = \{\}$ 
5:   for  $dp_i \in DP$  and  $1 \leq i \leq 4$  do
6:     if  $g.matches(dp_i) \wedge g.contains(nm)$  then
7:        $candidateAspects. \leftarrow g.apply(\{dp_5, dp_6\})$  ▷ Apply rules  $dp_5, dp_6$ 
8:     end if
9:   end for
10:   $aspects \leftarrow candidateAspects.select(N, NN)$  ▷ select nouns and compound nouns
11: end for
12:  $filterByFrequency(aspects)$  ▷ ignore low frequency aspects
13: return  $aspects$ 

```

Consider the sample sentence "The camera lens is good". According to Algorithm 1 applies to rule three: $cop(good, is) + nsubj(good, lens) \rightarrow \langle lens, good \rangle$.

¹ Sentences can be parsed using the Stanford Dependency parser [11]

Next, if a Noun Compound Modifier (*nm*) exists in the sentence, rules five and six apply; and in this example rule five applies resulting in the following candidate aspects: $(lens, good) + nm(lens, camera) \rightarrow \langle camera\ lens, good \rangle$. In this way given a set of reviews a set of candidate phrases are extracted. For each candidate, non noun (*N*) words are eliminated. Thereafter frequency of each candidate is calculated according to its *N* and *NN* phrase; retaining only those candidates above a frequency cut-off. See [1, 14] for the detailed definition of grammatical relations and its application.

4.2 Time-dependent Aspect Weight

The first of our weighting schemes assumes that an aspect is deemed important when it is frequently mentioned by the authors of the reviews. This allows us to monitor aspect trends in straight forward manner as reviews can be fairly easily grouped into seasonal time windows. Based on this principle, an aspect weight is derived by the relative aspect frequency at time window *t*.

$$AW(a_j, t) = \frac{Freq(a_j, t)}{\sum_{a_i \in \mathcal{A}} Freq(a_i, t)} \quad (3)$$

Where *Freq* returns the frequency of an aspect *a* in a given time window *t* into which reviews have been grouped. Frequency here is the number of times a term appears in a specified group of reviews. Table 2 shows an example of aspect weight calculations in two different time windows i.e. 2008 and 2011. We observe that although the aspect frequency of *megapixel* has increased overtime, its relative weight compared to all other aspects decreased significantly. Therefore, we suggest that the importance of the aspect *megapixel* has dropped over the years. Whilst frequency of aspects over time allows us to infer global trends about aspect usage, it does so without considering the relationship between aspects from preferred products. Therefore an alternative approach is to compare aspects shared between preferred products.

Table 2: Time-dependent *megapixel* aspect weight in 2008 - 2011

Year	Freq (<i>megapixel</i> , Year)	$\sum_{a_i \in \mathcal{A}} \text{Freq}(a_i, \text{Year})$	$AW(\text{megapixel}, \text{Year})$
2008	434	$1.4 \cdot 10^6$	$3 \cdot 10^{-5}$
2011	504	$4.5 \cdot 10^6$	$1.1 \cdot 10^{-5}$

4.3 Preference Aspect Weight

A product purchase choice is a preference made on the basis of one or more aspects. The notion of aspect importance arises when the same set of aspects contribute to similar purchase decisions. Using this principle, aspects weights are derived by comparing the aspect sentiment score differences between purchased and viewed product pairs in which $(p_x, p_y) \in \{(p_x, p_y)\}_{x,y=1 \wedge x \neq y}^d$

$$AW(a_j) = \frac{\sum_{x=1}^{|\mathcal{A}|} \sum_{y=1}^{|\mathcal{E}|} \delta(a_j, p_x, p_y)}{|d \in \mathcal{E}|} \quad (4)$$

where either $p_x \succ p_y$ or $p_y \succ p_x$ or both and d is the set of product preference pairs containing aspect a_j . The preference difference between any pairs of products is computed as:

$$\delta(a_j, p_x, p_y) = |L_{min}(\mathcal{A}, \mathcal{E})| + \delta'(a_j, p_x, p_y) \quad (5)$$

$$\delta'(a_j, p_x, p_y) = AS(a_j, p_x) - AS(a_j, p_y) \quad (6)$$

Here $|L_{min}(\mathcal{A}, \mathcal{E})|$ is the least minimum preference difference scores obtained over all aspects and product preference pairs. $AS(a_j, p_x)$ is the sentiment associated to aspect a_j of product p_x . Higher values of δ contribute to higher weights. Since δ' computation can result in negative scores, we use $|L_{min}(\mathcal{A}, \mathcal{E})|$ to bound the lowest value to zero. Thereafter we normalise $\delta(a_j, p_x, p_y)$ such that it is in range $[0,1]$.

5 Evaluation

In this section we evaluate our proposed integration of time and preference guided aspect weight extraction applied to product recommendation. We first conduct a pilot study to evaluate the quality of aspects extracted by our algorithm with the state-of-the-art. Thereafter, we evaluate how well the recommendation system works in practice on Amazon and Cnet data using two derived benchmark rankings.

5.1 Comparative Pilot Study - Aspect Extraction Analysis

We use a public dataset on product reviews containing manually marked-up product aspects [8, 3]. For this study we use phone category products with at least hundred reviews. Precision, recall and F measure is used to compare manually labeled aspects with extracted ones. We advocate higher precision because this would mean we are able to identify true aspects of a product and this will lead to better learning of users preference. Therefore, $F_{0.5}$ measure was used in this study. We compare our proposed approach with the following alternative extraction algorithms:

- FQITEMSETS uses Apriori algorithm to identify candidate aspects that are then pruned using a frequency cut-off threshold [8].
- FQPOS uses Part-of-Speech(POS) extraction patterns that are then pruned using sentiment informed frequency cut-off threshold [5]
- DPRULES uses the dependency extraction rules in [14].
- FQDPRULES same as DPRULES but prunes candidate aspects using a frequency cut-off (See Algorithm 1).

Precision of all frequency based extraction approaches are significantly better compared to DPRULES (see Table 3). As expected best results are achieved with FQDPRULES when deep NLP semantics is combined with frequency pruning. Here

Table 3: Preliminary results for aspect extraction approaches

Approach	Precision	Recall	$F_{0.5}$ measure
FQITEMSETS	0.51	0.24	0.41
FQPOS	0.71	0.11	0.22
DPRULES	0.28	0.66	0.32
FQDPRULES	0.76	0.25	0.54

we observe a 49% and 7% improvement in precision over FQITEMSETS and FQPOS respectively. We also observe that FQDPRULES has the highest $F_{0.5}$ value (0.54). Recall trends suggest that FQDPRULES must have many false negatives and so missed extraction opportunities compared to DPRULES. However, in the context of our study a lower precision is more damaging as it is likely to introduce aspect sparsity problems which have detrimental effect on sentiment difference computations. Therefore, we confirm that FQDPRULES provides a better result in the context of our study. On the basis of these results we use FQDPRULES to extract product aspects for sentiment analysis in social recommender experiments in the next sections.

5.2 Amazon Dataset and Ranking Strategies

We crawled 2264 Amazon products during April 2014. From this we use the *Digital SLR Camera* category containing more than 20,000 user generated reviews. Since we are not focusing on the cold-start problem, newer products and those without many user reviews are removed. Here we use 1st January 2008 and less than 15 reviews as the pruning factor for products. Finally, any synonymous products are united leaving us data for 80 products. The FQDPRULES algorithm extracted 981 unique aspects and on average 128 different aspects for each product. Importantly more than 50% of the products shared at least 70 different aspects, while 30% shared more than 90 aspects on average. The fact that there are many shared aspects is reassuring for product comparison when applying Equation 1.

The retrieval set of a query product consists of products that share a similar number of aspects. This retrieval set is ranked using the following sentiment-based recommendation strategies considering only the k top shared aspects between the retrieved and the query product. The comparative weighting schemes used in our experiments are as follows:

- *BASE*: recommend using aspect sentiment analysis without considering aspect weights presented in Equation 2;
- *PrefW*: same as *BASE* but with the additional preference aspect weighting component from Equation 4;
- *TimeW_t*: same as *PrefW* but considering the time context for aspect weighting component (instead of preference) presented in Equation 3.

It is worth noting that Amazon only provides the current snap shot of preferences. Therefore we are unable to explore the impact of time on preference-based aspect weight extraction. We will present the *TimeW_t* strategy considering all reviews created between three different time windows: 2008-2014, 2011-2014 and 2014.

5.3 Evaluation Metrics

In the absence of ground truth data and individual user specific purchase trails, we generated two different benchmark lists according to the following dimensions:

- *Popular*: Derived from Amazon’s reviews, questions and timeline data. Products are ranked based on its popularity computed by means of Equation 7.

$$Popular(p) = \frac{nReviews + nQuestions}{days_online} \quad (7)$$

where $nReviews$ and $nQuestions$ refer to the number of reviews and questions of a product respectively, and $days_online$ is the number of days the product has been on Amazon’s website. We found that this formula has some correlation with the top 100 Amazon Best Seller ranking (*Spearman correlation* of -0.4381). Unlike Amazon’s benchmark this allows us to experiment with query products that may not necessarily be in the Amazon top 100². 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.

$$\%RankGain = \frac{\sum_{i=1}^{n=3} benchmark(P_q) - benchmark(P_i)}{n * |\mathcal{P} - 1|} \quad (8)$$

where n is the size of the retrieved set and *benchmark* returns the position on the benchmark list. The greater the gain over the query product the better.

- *Precision*: Derived from Cnet.com expert recommendations for DSLRs cameras³. This groundtruth is divided in three subcategories (*entry-level DSLRs*, *midrange DSLRs* and *professional DSLRs*), each containing a list of cameras recommended by Cnet experts.

$$Precision = \sum_{i=1}^n \frac{TopN_{cat} \cap Cnet_{cat}}{n} \quad (9)$$

In the absence of a defined Cnet ranking per category, we use a leave-one-out methodology to evaluate the capacity of our strategies to recommend the expert-selected cameras in each category. We compute the precision by means of Equation 9, where $TopN_{cat}$ is the list of the top n recommended products for category cat , and $Cnet_{cat}$ is the list of Cnet expert recommended cameras for that category.

5.4 Results - Amazon

Here we present results from our exploration of aspects trends in terms of weights over time followed by a comparison of the two weighting schemes.

² <http://www.amazon.co.uk/Best-Sellers-Electronics-Digital-SLR-Cameras>

³ <http://www.cnet.com/uk/topics/cameras/best-digital-cameras/>

5.4.1 Importance Of Time on Aspect Weighting

Figure 3 shows the weight of the aspects *megapixel*, *autofocus* and *battery* computed by using strategy $TimeW_t$ for years between 2008 and 2014. We observe that *megapixel* was an important aspect in 2008 with a frequency weight of close to 0.009. However, its importance decreased dramatically during the following years, reducing its weight up to five times in 2011. In contrast, with *autofocus*, we see an increasing trend. A different trend can be observed for *battery* in Figure 3. Here it is interesting to note that the aspect weight is maintained over the years. Whilst there is a negligible loss in the raw score this is explained by the difference in number of unique aspects in the time period. For example, in 2008 we found approximately 250 aspects whilst in 2014 this had increased to 900.

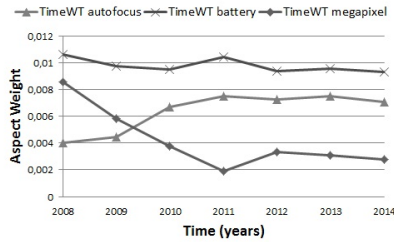


Fig. 3: Aspects weight over time (in years)

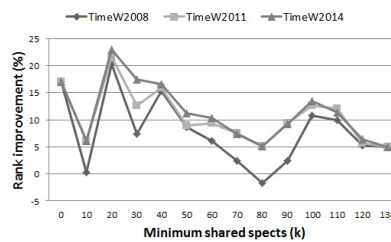


Fig. 4: ProdSenti on Popular benchmark

In Figure 4, we use $TimeW_t$ with $t = 2008-2014$, $2011-2014$ and 2014 to rank the recommendations for increasing number of shared aspects (k) on benchmark *Popular*. In general, we observe that weights learned using $TimeW_{2014}$ perform up to a 15% better for $k = 30$ when recommending actual products. $TimeW_{2011}$ falls close to the recommendations made by $TimeW_{2014}$, being the weights learned by $TimeW_{2008}$ the ones that perform worst. These results indicate that considering the most recent time frame for computing the aspect weights improves the recommendations made by the system, and that aspect frequency over time is a good indicator of what users care most when considering cameras.

5.4.2 Time vs Preference Weighting

In Figure 5 we compare the three strategies, $TimeW_{2014}$, $PrefW$ and BASE using the *Popular* benchmark. We include the strategy agnostic of aspects weights, BASE, in order to compare the impact that weights have on the recommendations while considering preferences and time weighted aspects. As we can observe, $TimeW_{2014}$ and $PrefW$ strategies outperform BASE by more than 10% on average. On the other hand, we observe that $PrefW$ outperforms $TimeW_{2014}$ for all values of k comprised between 30 and 100. This suggests that preference weights used by $PrefW$ are able to recommend better products for the *Popular* benchmark since they represent the most recent snapshot of the current users preferences. We also observe that there seems to be a sweet spot in terms of the number of shared aspects (k), with $PrefW$

achieving best results with $k=30$, and a gradual decrease thereafter. However, the rank improvement obtained by considering time in $TimeW_{2014}$ should not be ignored as it performs 15% better than $PrefW$ with smaller subsets of shared aspects (e.g. $k = 20$) and obtains a similar rank improvement with increased numbers of shared aspects (e.g. $k \geq 100$). Close examination of $TimeW_{2014}$'s performance suggests that the retrieval set consisted of a high number of products (85%) with $k = 20$ shared aspects. This is in contrast to its poorer performance with higher values of k . This suggests that with less frequently occurring aspects the frequency based weight computation of $TimeW_{2014}$ is likely to be less reliable compared to $PrefW$.

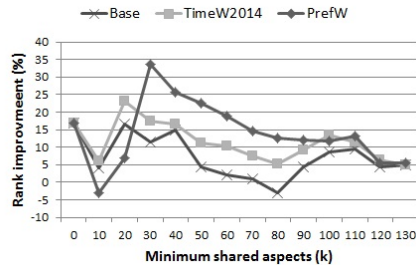


Fig. 5: Comparison of different strategies

Table 4: Top 10 for $TimeW_{2014}$ and $PrefW$

Top 10 aspects for $TimeW_{2014}$		Top 10 aspects for $PrefW$	
Aspect	Weight	Aspect	Weight
picture	.00982	shutter	.00229
feature	.00974	photography	.00181
photographer	.00956	point	.00179
battery	.00948	system	.00176
setting	.00942	video	.00166
photography	.00877	setting	.00165
image	.00857	picture	.00158
result	.00847	advantage	.00152
time	.00843	sensor	.00150
quality	.00842	manual	.00150

Table 4 presents the top 10 aspects extracted by means of $TimeW_{2014}$ and $PrefW$ and its correspondent weight. Here the lists and weights of the top aspects obtained by the two strategies are different except for aspects *picture*, *setting* and *photography*. Although occupy different ranking positions and are weighted differently, both strategies seem to agree on their importance. Also, the weight distribution of each strategies is different. For example, $TimeW_{2014}$ gives almost five times more weight to *picture* (0.00982) than $PrefW$ does to its top aspect *shutter* (0.00229). We also notice that there are several semantically related aspects that appear in the top 10: *image* represents a similar concept to *picture* and *photography*; similarly *system* and *settings*. It is likely that mapping such related aspects to a common concept is able to generate more accurate aspect weights.

5.5 Results - Cnet

Next we compare our recommendations using weights extracted from Amazon sources against expert advice. For this purpose we use Cnet expert advice on DSLR camera recommendations. We divided all cameras from our dataset in three subsets, corresponding to *entry-level*, *midrange* and *professional* categories presented on the website, based on price (see Table 5). Table 5 also shows the number of products included in every subset and the number of Cnet products used as the gold standard.

Table 6 shows the average precision of $TimeW_{2014}$, $PrefW$ and BASE for every Cnet category and different recommendation set sizes ($n=1$ and $n=3$) computed by means of Equation 9. We included a strategy that randomly recommends products, *Random*, to facilitate understanding of the results. As we observe, strategies that

Table 5: Cnet dataset classification for Entry-level, Mid-range and Professional

	Entry-level	Mid-range	Professional
Price range (\$)	0-1k	1k-2.2k	2.2k-7k
$\#products_{cat}$	60	14	6
$\#products_{cat} \cap Cnet_{cat}$	7	8	3

are aware of aspect sentiments are able to improve precision in every category. In *Popular* dataset, *PreFW* is the strategy that performs better: for the *Cnet Entry-level* subset, it is able to recommend a Cnet top product 37% of the time for $n=1$ and a 24% for $n=3$ on average. These results are promising considering that the probability of recommending a top Cnet product amongst the entire subset in this category is 20% and 7% respectively. Precision results for all three strategies are much higher when applied to smaller Cnet datasets; *Mid-range* and *Professional*, achieving a precision of close to 85% and 80% for $n=1$ respectively and doubling the precision of the random recommender. Overall, since our system recommendations closely match with Cnet expert’s recommendations, we conclude that the aspect weights learned from Amazon are likely to correspond with criteria that the Cnet expert’s might have implicitly used. However, we cannot accurately verify this without manually demanding user trial. Nevertheless it is interesting that consensus knowledge discovered from social media seem to closely echo views of the domain experts.

Table 6: Precision for different retrieved set sizes n in Cnet.

Strategy	Entry-level		Mid-range		Professional	
	$n=1$	$n=3$	$n=1$	$n=3$	$n=1$	$n=3$
Random	0.203	0.077	0.461	0.376	0.2	0.277
BASE	0.220	0.146	0.769	0.423	0.4	0.377
<i>TimeW</i> ₂₀₁₄	0.254	0.128	0.846	0.444	0.8	0.4
<i>PreFW</i>	0.372	0.243	0.846	0.461	0.6	0.422

6 Conclusions

Tracking users preference over time raises unique challenges for recommendation systems. Every product potentially goes through a series of changes which typically involves functional improvements resulting in a broader range of aspects which in turn will be echoed in changes to user preferences. Our previous findings suggested that preference knowledge allows us to identify aspects that are important to users but lacks the capability to trace aspect importance trends.

In this paper, we formalise a novel aspect-based sentiment ranking that utilise both time and preference contexts. The benefits are demonstrated in a realistic recommendation setting using benchmarks generated from Amazon and Cnet. We show that monitoring aspect frequency in product reviews allows to capture changes to aspect importance over time. Importantly, we confirm that time context can be conveniently exploited by using the recent time frame to improve recommendations.

We compare the performance of aspect-based sentiment ranking in the context of time and preference. We observed that both contexts perform well in different number of shared aspects, further work is needed to study the benefit of integrating

both contexts in ranking algorithm. Our results show that similar aspects were mentioned using different terms, further work is needed to study how sparsity problems might impact the difference calculations. Finally, it would be interesting to integrate trending information within the aspect weight computation to infer their importance.

Acknowledgments.

This research has been partially supported by AGAUR Scholarship (2013FI-B00034), ACIA mobility scholarship and Project Cognitio TIN2012-38450-C03-03.

References

1. Chen, Y., Ferrer, X., Wiratunga, N., Plaza, E.: Sentiment and preference guided social recommendation. In: Inter. Conf. on Case-Based Reasoning. Accepted in ICCBR '14 (2014)
2. Cho, Y., Cho, Y., Kim, S.: Mining changes in customer buying behavior for collaborative recommendations. *Expert Systems with Applications* **28**, 359–369 (2005)
3. Ding, X., Liu, B., Yu, P.: A holistic lexicon-based approach to opinion mining. In: Proc. Int. Conf. on Web Search and Data Mining (2008)
4. Ding, Y., Li, X.: Time weight collaborative filtering. In: Proc. Inter. Conf. on Information and Knowledge Management, CIKM '05, pp. 485–492 (2005)
5. Dong, R., Schaal, M., O'Mahony, M., McCarthy, K., Smyth, B.: Opinionated product recommendation. In: Inter. Conf. on Case-Based Reasoning (2013)
6. Esuli, A., Sebastiani, F.: Sentiwordnet: A publicly available lexical resource for opinion mining. In: Proc. Language Resources and Evaluation Conference, pp. 417–422 (2006)
7. Hong, W., Li, L., Li, T.: Product recommendation with temporal dynamics. *Expert Syst. Appl.* **39**, 12,398–12,406 (2012)
8. Hu, M., Liu, B.: Mining and summarising customer reviews. In: Proc. of ACM SIGKDD Inter. Conf. on Knowledge Discovery and Data Mining, pp. 168–177 (2004)
9. Kolter, J., Maloof, M.: Dynamic weighted majority: A new ensemble method for tracking concept drift. In: Inter. Conf. on Data Mining, pp. 123–130 (2003)
10. Koren, Y.: Collaborative filtering with temporal dynamics. *Communications of Association for Computing Machinery* **53**, 89–97 (2010)
11. Marneffe, M., MacCartney, B., Manning, C.: Generating typed dependency parses from phrase structure parses. In: Proc. of Language Resources and Evaluation Conf. (2006)
12. McCarthy, K., Salem, Y., Smyth, B.: Experience-based critiquing: reusing experiences to improve conversational recommendation. In: Inter. Conf. on Case-Based Reasoning (2010)
13. Moghaddam, S., Ester, M.: Opinion digger: An unsupervised opinion miner from unstructured product reviews. In: Proc. Inter. Conf. on Information and Knowledge Management (2010)
14. Moghaddam, S., Ester, M.: On the design of lda models for aspect-based opinion mining. In: Proc. Inter. Conf. on Information and Knowledge Management (2012)
15. Muhammad, A., Wiratunga, N., Lothian, R., Glassey, R.: Contextual sentiment analysis in social media using high-coverage lexicon. In: *Research and Development in Intelligent* (2013)
16. Popescu, A., Etzioni, O.: Extracting product features and opinions from reviews. In: *Natural language processing and text mining*, pp. 9–28 (2007)
17. Turney, P.: Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: Proc. Annual Meeting for Computational Linguistics (2002)
18. Vasudevan, S., Chakraborti, S.: Mining user trails in critiquing based recommenders. In: Proc. Inter. Conf. on World Wide Web Companion, pp. 777–780 (2014)
19. Xiang, L., Yuan, Q., Zhao, S., Chen, L., Zhang, X., Yang, Q., Sun, J.: Temporal recommendation on graphs via long- and short-term preference fusion. In: Proc. Inter. Conf. on Knowledge Discovery and Data Mining, pp. 723–732 (2010)
20. Yitzhaki, S.: Relative deprivation and the Gini coefficient. *The Quarterly Journal of Economics* pp. 321–324 (1979)