Polarities, Axiallities and Marketability of Items

Dan A. Simovici[1], Paul Fomenky[1], and Werner Kunz[2]

[1] University of Massachusetts Boston,
Department of Computer Science, Boston, MA 02125, USA
[2] University of Massachusetts Boston,
Department of Marketing, College of Management, Boston, MA 02125, USA

Abstract. We apply polarities, axiallities and the notion of entropy to the task of identifying marketable items and the customers that should be approached in a marketing campaign. An algorithm that computes the criteria for identifying marketable items and the corresponding experimental work is also included.

Keywords: polarity, axiallity, closure operator, entropy

1 Introduction

Recommender systems (RS) aim to help users deal with the immensity of offers in electronic commerce by customizing the most adequate offerings for their specific needs. After almost two decades, this discipline is quite established following the initial publications [11, 10] and [13]. The general framework of RSs involves a bipartite graph whose set of vertices is partitioned into customers and items (see Figure 1). Edges of the form (c, t), where c is a customer and t is an item are marked by numerical ratings $r(c, t)$. Recommender systems that use ratings generate item recommendations for customers or identify sets of customers suitable for sets of items and fit into one of the following broad three approaches: content-based, collaborative, or hybrid. An excellent survey of developments in recommender systems can be found in [3].

Further, more sophisticated types of RSs extend the user/item paradigm to take into account temporal and other contextual characteristics of customers and items (see [2]). All existing approaches have in common the use of additional data (rating, survey information) and many are based on hard-to-implement complex inference statistical approaches [1, 7, 9]. Furthermore, most of then aim to obtain optimal recommendations for the consumer and ignore the business perspective.

Formal concept analysis which makes use of the notion of polarity was applied in the study of RSs in [6] in the quest of simplifying the task of finding similar users or similar items without loss of accuracy or coverage. The purpose of this paper is distinct from the main topics of the area recommender systems. We investigate possibilities to identify items that should be the object of marketing campaigns and we intend to extend this approach to sets of items that can be co-marketed. Thus, we approach RSs from the position of the seller rather than from the prospective of the users. Instead of providing recommendations to the users we propose to suggest items or sets of items to retailers which could stimulate sales.

We use entropy (and some of its generalizations) as measures of scattering of partition blocks for partitions of finite sets. Namely, if $\pi = \{B_1, \ldots, B_n\}$ is a partition of a finite set S, its generalized b-entropy (see [5, 12]) is defined as

$$\mathcal{H}_b = \frac{1}{1 - 2^{1-b}} \left(1 - \sum_{i=1}^{n} \left(\frac{|B_i|}{|S|} \right)^b \right),$$

where $b > 1$.

Two special cases of generalized entropy are particularly interesting. For $b = 2$ we have the Gini index of π given by

$$\mathsf{gini}(\pi) = 2 \left(1 - \sum_{i=1}^{n} \left(\frac{|B_i|}{|S|} \right)^2 \right).$$

The largest value of $\mathsf{gini}(\pi)$ is obtained when all blocks have equal size (and this, the elements of S are uniformly scattered in the blocks of π); in this case we have

$$\mathsf{gini}(\pi) = 2 \left(1 - \frac{1}{n} \right).$$

The lowest value, $\mathsf{gini}(\pi) = 0$, is obtained when π consists of one block.

The other interesting case of generalized entropy is obtained when b tends to 1. In this case

$$\lim_{b \to 1} \mathcal{H}_b(\pi) = - \sum_{i=1}^{n} \frac{|B_i|}{|S|} \log_2 \frac{|B_i|}{|S|},$$

which recaptures the well-known Shannon entropy.

The article is structured as follows. Section 2 introduces the notions of polarity and axiallity in the context of recommender systems. In Section 3 we define marketable item sets and formulate an algorithm for identifying these sets. Experimental work is described in Section 4.

2 Polarities, Axiallities and Recommender Systems

Let C and T be two finite sets, referred to as the *set of customers* and the *set of items*, respectively and let $\rho \subseteq C \times T$. Following the terminology of concept lattices [8] we shall refer to the triple $\mathcal{C} = (C, T, \rho)$ as a *recommendation context*. The fact that $(c, t) \in \rho$ means that the customer c has purchased the item t. Sets of customers will be denoted by letters from the beginning of the alphabet $D, E, K, \ldots$; sets of items will be denoted by letters from the end of the alphabet $U, V, W, \ldots$.

For a set S, the set of subsets of S is denoted by $\mathcal{P}(S)$. If S, T are sets, a function $f : \mathcal{P}(S) \longrightarrow \mathcal{P}(T)$ is *monotonic* if for every $X, Y \in \mathcal{P}(S)$, $X \subseteq Y$ implies $f(X) \subseteq f(Y)$; f is *anti-monotonic* if $X \subseteq Y$ implies $f(X) \supseteq f(Y)$.

Consider the mappings $\phi_\rho : \mathcal{P}(C) \longrightarrow \mathcal{P}(T)$, and $\psi_\rho : \mathcal{P}(T) \longrightarrow \mathcal{P}(C)$ given by

$$\phi_\rho(D) = \{t \in T \mid (\forall d \in D)(d,t) \in \rho\},$$
$$\psi_\rho(U) = \{c \in C \mid (\forall u \in U)(c,u) \in \rho\}$$

for $D \in \mathcal{P}(C)$ and $U \in \mathcal{P}(T)$. In other words, $\phi_\rho(D)$ consists of items that were bought by all customers in D and $\psi_\rho(U)$ consists of customers who bought all items of U. As shown in [4] (Chapter V), the mappings ϕ_ρ and ψ_ρ are anti-monotonic. The pair $\mathsf{Pol}_\rho = (\phi_\rho, \psi_\rho)$ is the *polarity* of ρ,

Let $\overline{\rho} = (C \times T) - \rho$ be the relation that consists of all pairs (c,t) that are not in ρ. Another pair of functions defined by ρ is $(\alpha_\rho, \beta_\rho)$, where $\alpha_\rho : \mathcal{P}(C) \longrightarrow \mathcal{P}(T)$ and $\beta_\rho : \mathcal{P}(T) \longrightarrow \mathcal{P}(C)$ are given by

$$\alpha_\rho(D) = \phi_{\overline{\rho}}(\bar{D}) \text{ and } \beta_\rho(U) = \overline{\psi_{\overline{\rho}}(U)}$$

for $D \in \mathcal{P}(C)$ and $U \in \mathcal{P}(T)$, where $\bar{D} = C - D$.

By applying the definition of $\phi_{\overline{\rho}}$ we have

$$\begin{aligned}\alpha_\rho(D) &= \phi_{\overline{\rho}}(\bar{D}) = \phi_{\overline{\rho}}(C - D)\\ &= \{t \in T \mid (\forall d \in C - D)(d,t) \notin \rho\}.\end{aligned}$$

In other words, $\alpha_\rho(D)$ consists of those items $t \in T$ which were not bought by any customer who does not belong to D.

Similarly, by applying the definition of $\psi_{\overline{\rho}}$ we have

$$\begin{aligned}\beta_\rho(U) &= \overline{\psi_{\overline{\rho}}(U)} = C - \psi_{\overline{\rho}}(U)\\ &= C - \{c \in C \mid (\forall u \in U)(c,u) \notin \rho\}\\ &= \{c \in C \mid (\exists u \in U)(c,u) \in \rho\},\end{aligned}$$

which shows that $\beta_\rho(U)$ consists of customers who bought some items in U.

It is immediate that the functions α_ρ, β_ρ are monotonic. In other words, we have

$$D_1 \subseteq D_2 \Rightarrow \alpha_\rho(D_1) \subseteq \alpha_\rho(D_2),$$
$$U_1 \subseteq U_2 \Rightarrow \beta_\rho(U_1) \subseteq \beta_\rho(U_2),$$

for $D_1, D_2 \in \mathcal{P}(C)$ and $U_1, U_2 \in \mathcal{P}(T)$.

The pair $\mathsf{Axl}_\rho = (\alpha_\rho, \beta_\rho)$ is the *axiallity* of ρ.

For the polarity mappings we have

$$D \subseteq \psi_\rho(\phi_\rho(D)) \text{ and } U \subseteq \phi_\rho(\psi_\rho(U))$$

for any set of customers $D \in \mathcal{P}(C)$ and every set of items $U \in \mathcal{P}(T)$. The mappings $\psi_\rho\phi_\rho$ and $\phi_\rho\psi_\rho$ are closure operators on $\mathcal{P}(C)$ and $\mathcal{P}(T)$, respectively.

For the axiallity mappings we have

$$\begin{aligned}\beta_\rho(\alpha_\rho(D)) &= \overline{\psi_{\overline{\rho}}(\alpha_\rho(D))}\\ &= \overline{\psi_{\overline{\rho}}(\phi_{\overline{\rho}}(\overline{D}))} \subseteq D,\end{aligned}$$

and

$$\begin{aligned}\alpha_\rho(\beta_\rho(U)) &= \alpha_\rho(\overline{\psi_{\overline{\rho}}(U)}) \\ &= \phi_{\overline{\rho}}(\psi_{\overline{\rho}}(U)) \supseteq U,\end{aligned}$$

which allows us to conclude that $\beta_\rho\alpha_\rho$ is an interior operator on sets of customers and $\alpha_\rho\beta_\rho$ is a closure operator on sets of items.

Let ITEMS_c be the set of items acquired by customer c and let CUST_t be the set of customers who bought item t. Also, define the sets P_c and R_t by

$$P_c = \{c\} \times \mathsf{ITEMS}_c \text{ and } R_t = \mathsf{CUST}_t \times \{t\}$$

for each customer c and item t (see Figure 1).

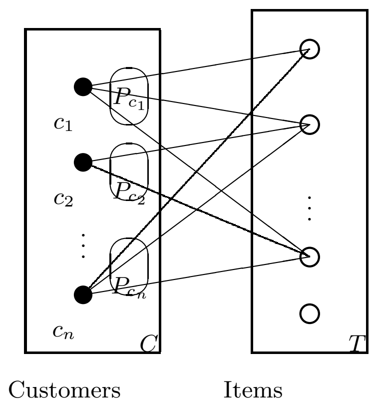

Fig. 1. Partitions of the set of purchases generated by customers

Both collections $\{P_c \mid c \in C\}$ and $\{R_t \mid t \in T\}$ are partitions of the set of purchases ρ. Also, observe that $P_c \cap R_t = \{(c,t)\}$ for $c \in C$ and $t \in T$.

For a set of customers $D \subseteq C$, the set of purchases is $\mathsf{pur}(D) = \bigcup_{c\in D} P_c$ and the collection of sets $\{P_c \mid c \in D, P_c \neq \emptyset\}$ is a partition of $\mathsf{pur}(D)$.

The entropy (or the Gini index) of the partition $\pi_D = \{P_c \mid c \in D, P_c \neq \emptyset\}$ captures the diversity of purchasing patterns for the customers in D. Note that

$$\begin{aligned}\mathcal{H}_1\left(\{P_c \mid c \in D, P_c \neq \emptyset\}\right) &= -\sum_{c\in C} \frac{|P_c|}{|\mathsf{pur}D|} \log_2 \frac{|P_c|}{|\mathsf{pur}D|} \\ &= \log_2 |\mathsf{pur}(D)| - \frac{1}{|\mathsf{pur}(D)|} \sum_{c\in D} |P_c| \log_2 |P_c|.\end{aligned}$$

.

We use the *specific entropy* $h_1(D)$ of a set of customers D defined as the ratio between the entropy of the purchases of customers in D and the size of D

$$h_1(D) = \frac{\mathcal{H}_1\left(\{P_c \mid c \in D, P_c \neq \emptyset\}\right)}{|D|}.$$

The specific entropy is intended to compensate the growth of the entropy of the partition $\{P_c \mid c \in D\}$ due to an increase in the size of the customer population D, and seems to be a better indicator of the diversity of the purchasing patterns of the population in D than the entropy of the partition π_D.

3 Marketable Items

We examine criteria for choosing items that should be the object of a marketing campaign. The reason for starting a marketing campaign involving an item t is that the set of users who purchased t is non-empty but small; in other words, $|\psi_\rho(\{t\})|$ does not exceed a threshold θ. Using the notions of polarity and axiallity that can be defined starting from the purchasing relation defined as a binary relation on the sets of customers and items we focus on items that satisfy several conditions:

1. the closure $\phi_\rho(\psi_\rho(\{t\})$, which consists of items that were bought by customers who bought t must be sufficiently large;
2. since β_ρ is a monotonic mapping, the set of customers who bought some item in the previously mentioned set of items, $\beta_\rho(\phi_\rho(\psi_\rho(\{t\}))$ will, in turn be large, and
3. the purchasing patterns of these customers must be sufficiently diverse, to ensure a reasonable chance that they will decide to buy t.

The target of the marketing campaign is the set of customers $\beta_\rho(\phi_\rho(\psi_\rho(\{t\})) - \psi_\rho(\{t\})$. These criteria are summarized in Algorithm 1.

The SQL procedure that implements the algorithm and includes the computation of the entropy of purchases is given next.

```
create procedure market1(item1 integer)
    begin
        # cleaning up tables for intermediate results
        call cleanup();

        # custforitem(userid) contains customers who bought item1

        insert into custforitem
            select userid from pur where item = item1;


        # items bought by every customer in custforitem
        # are stored in itemsbyallcust(item)

        insert into itemsbyallcust
            select distinct item from pur r where
                not exists(select * from custforitem where
                        not exists(select *
```

Data: A table pur of purchases, a minimum and a maximum number of purchases minpurand maxpur, respectively
Result: a set of customers targeted for the marketing campaign
Place in table selitems items from pur bought by at least minpur customers but not more that maxpurcustomers ;
foreach *item t in selitems* **do**
  retrieve in table custforitem customers who bought t, $\psi_\rho(\{t\}) = \beta_\rho(\{t\}) \rightarrow$ custforitem;
  retrieve in table itemsbyallcust items bought by every customer in custforitem, $\phi_\rho(\psi_\rho(\{t\}) \rightarrow$ itemsbyallcust;
  retrieve in table custwhoboughtsome customers who bought some item in itemsbyallcust, $\beta_\rho(\phi_\rho(\psi_\rho(\{t\})) \rightarrow$ custwhoboughtsome;
  retrieve in targetitem customers targeted for marketing $\beta_\rho(\phi_\rho(\psi_\rho(\{t\})) - \psi_\rho(\{t\}) \rightarrow$ targetitem;
  compute the entropy (or the Gini index) for the purchases made by customers targeted for marketing;
end

Algorithm 1: Algorithm for computing the target set of a marketing campaign

```
                from pur where
                userid = custforitem.userid
                and item = r.item));

# custwhoboughtsome(userid) contains customers who bought
# some item in itemsbyallcust

insert into custwhoboughtsome
    select distinct userid from pur
    where item in (select item from itemsbyallcust);

# targetitem(userid) contains customers targeted for marketing

insert into targetitem
    select userid from custwhoboughtsome where
        not exists(select * from custforitem
                        where userid = custwhoboughtsome.userid);

select 'Customers targeted for marketing';

select userid from targetitem;

# calculation of entropy for the customers in targetitem
# follows
```

```
    insert into purcust(userid, noitem)
        select userid, numitemcust(userid) from targetitem;

    insert into custsq
        select userid,noitem*log(2,noitem) from purcust;

    insert into results
        select sum(c),(select sum(noitem) from purcust)
            from custsq;


    select log(2,ct) - (1/ct) *s from results;
end
```

4 Experimental Study

We used the MovieLens data set that contains 100,000 anonymous ratings of 1,682 movies made by 943 customers (referred to as users in the documentation of the data set). This data set was obtained from the University of Minnesota GroupLens Research `www.movielens.org` and was processed using `mySQL`. The main characteristics of the attributes of this data set are specified below.

- UserIDs are integers;
- MovieIDs range between 1 and 1,682;
- Ratings are made on a 5-star scale (whole-star ratings only);
- Timestamp is represented in seconds;
- Each user has at least 20 ratings.

The relation pur was extracted by projecting the data set on the attributes UserId and MovieId (referred to as item). The customer sets for the analyzed items consisted between 5 and 10 individuals, as shown in the second column of Table 1. D_t is the set of customers targeted for the marketing campaign for t,

$$D_t = \beta_\rho(\phi_\rho(\psi_\rho(\{t\})) - \psi_\rho(\{t\}).$$

The customer population targeted for these campaigns varied between 116 and 880 individuals, as shown in Table 1. It is not a surprise that the size of the targeted customer population D_t has a strong positive correlation with the entropy of the partition of purchases of this set of customers. For example, for the sample of movies we experimented the correlation coefficient is 0.91.

As it can be seen from Figure 2, the larger the targeted population of customers, the larger the entropy is and therefore, we have a better chance that some of these customers will buy the item t, which is the focus of the marketing campaign. This explains the strong positive correlation between the size of the targeted population and the entropy.

Table 1. Size and Entropy for several items viewed by five to ten users.

item t	cust_for_item $\|\psi_\rho(t)\|$	cust_targeted $\|D\|$	entropy $\mathcal{H}_1(\pi_{D_t})$	spec_entropy $h(D_t))$
34	7	880	9.28	19.50
37	8	731	9.09	22.45
74	7	661	8.98	24.11
75	5	714	9.06	22.58
104	5	684	9	22.53
113	9	116	6.44	16.40
247	5	546	8.68	22.02
296	6	670	8.87	18.12
314	5	798	9.11	18.18
390	10	600	8.82	22.57
437	5	556	8.96	57.37
438	6	655	8.96	23.66
439	5	556	8.96	57.37
446	9	447	8.51	29.93

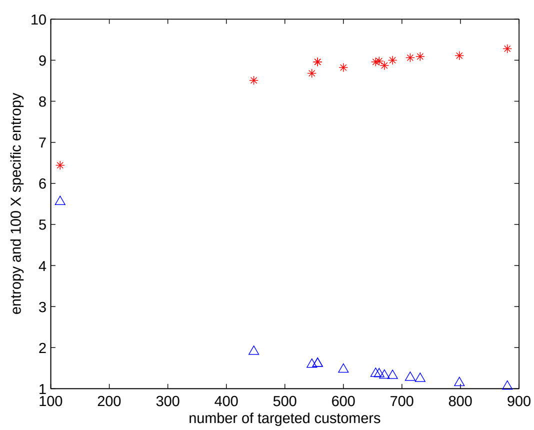

Fig. 2. Entropy ($*$) and Specific Entropy per Customer ($\triangle$) vs. Size of Set of Targeted Customers

For a targeted population D_t the maximum entropy of the partition of purchases is $\log_2 |D_t|$. The specific entropy defined as

$$h(D_t) = \frac{\log_2 |D_t|}{\log_2 |D_t| - \mathcal{H}_1(\pi_{D_t})}$$

is a better indicator of the diversity of purchases because it takes into account the relative size of the customer population targeted. Thus, the best targets for a marketing campaign are the items 437 and 439 for which $h_1(D_t)$ has a relatively high value (57.37).

5 Further Work

In this paper we introduced an approach that can be used by companies to market item sets to customer groups that possess a very likely high preference for these products. The algorithm is easy to implement and use in a daily managerial life. As input data, solely the past purchase data of all customers are needed, which is usually available in a company today. Our proposed approach can be used towards two directions. It can be used to segment the customers according to their preference fit for an individual item (set) as well as to build groups of items and rank them according their productivity for the existing customer base. In future research more simulation studies are needed to show that the proposed approach is not only easier to implement but is also equally (or better) suited to forecast customer product fit as well as optimize profit for the implementing company.

The productivity $\mathsf{prod}(t)$ of a marketing campaign for an item t can be measured by the ratio between the size of the target population of customers and the size of the set of customers who bought t:

$$\mathsf{prod}(t) = \frac{|\beta_\rho(\psi_\rho(\phi_\rho(\{t\})))|}{\phi_\rho(\{t\})}$$

This function can be extended to a set of items U by defining

$$\mathsf{prod}(U) = \frac{|\beta_\rho(\psi_\rho(\phi_\rho(U)))|}{\phi_\rho(U)}$$

for $U \subseteq T$. Note that this function is monotonic with respect to U; in other words, $U_1 \subseteq U_2$ implies $\mathsf{prod}(U_1) \leq \mathsf{prod}(U_2)$. We intend to explore criteria for marketing jointly sets of items using both productivity and the entropy of the set of customer purchases.

Incorporating the effect of the ratings of items by users will also be investigated in the future.

Acknowledgment

D. A. Simovici was supported by a S&T grant from the President's Office of the University of Massachusetts.

References

1. A. Ansari, S. Essegaier, and R. Kohli. Internet Recommendation Systems. *Journal of Marketing Research*, 37:363–377, 2000.
2. G. Adomavicius, R. Sankaranarayanan, S. Sen, and A. Tuzhilin. Incorporating contextual information in recommender systems using a multidimensional approach. *ACM Transactions on Information Systems*, 23:103–145, 2005.
3. G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749, 2005.
4. G. Birkhoff. *Lattice Theory*. American Mathematical Society, Providence, RI, third (with corrections) edition, 1995.
5. Z. Daróczy. Generalized information functions. *Information and Control*, 16:36–51, 1970.
6. P. du Boucher-Ryan and D. Bridge. Collaborative recommending using formal concept analysis. *Knowledge-Based Systems*, 19:309–315, 2006.
7. R. Decker and M. Trusov. Estimating aggregate consumer preferences from online product reviews. *International Journal of Research Marketing*, 27:293–307, 2010.
8. K. Denecke, M. Erné, and S. L. Wismuth. *Galois Connection and Applications*. Kluwer Academic Publishers, Dordrecht, 2004.
9. P. S. Fader and B. G. S. Hardie. Modeling consumer choice among SKUs. *Journal of Marketing Research*, 33:442–452, 1996.
10. W. Hill, L. Stead, M. Rosenstein, and G. Furnas. Recommending and evaluating choices in a virtual community of use. In *Proceedings of the Conference on Human Factors in Computer Systems*, pages 194–201, New York, 1995. ACM.
11. P. Resnick, N. Iakovou, M. Sushak, P. Bergstrom, and J. Riedl. Grouplens: An open architecture for collaborative filtering of netnews. In *Proceedings of Computer Supported Cooperative Work*, pages 175–186, New York, 1994. ACM.
12. D. A. Simovici and C. Djeraba. *Mathematical Tools for Data Mining*. Springer Verlag, New York, 2008.
13. U. Shardanand and P. Maes. Social information filtering: Algorithms for automating word of mouth. In *Proceedings of the Conference on Human Factors in Computer Systems*, pages 210–217, New York, 1995. ACM.