Matching Media Contents with User Profiles by means of the Dempster-Shafer Theory

Luigi Troiano
University of Sannio
Department of Engineering
Benevento, Italy
Email: troiano@unisannio.it

Irene Díaz
Oviedo University
Computer Science Department
Gijón, Spain
Email: sirene@uniovi.es

Ciro Gaglione
Sky Italia
Interactive Tv Lab
Milan, Italy
Email: ciro.gaglione@skytv.it

***Abstract*—The media industry is increasingly personalizing the offering of contents in attempt to better target the audience. This requires to analyze the relationships that goes established between users and content they enjoy, looking at one side to the content characteristics and on the other to the user profile, in order to find the best match between the two. In this paper we suggest to build that relationship using the Dempster-Shafer's Theory of Evidence, proposing a reference model and illustrating its properties by means of a toy example. Finally we suggest possible applications of the model for tasks that are common in the modern media industry.**

I. Introduction

Digital technologies are radically changing the way of performing business in media industry, with new possibilities of tailoring the catalog so that everybody has the chance of enjoying contents that best fit his/her interests, often on demand, at the time that is most appropriate for each user. Such a change is requiring to reformulate the way of building the content offering. Data collected from customers regarding their profile and preferences become central, so models able to interpret and to reason about data.

These models aims to discover and exploit the relationship that stands between users and media contents they enjoy. Here the problem is not to ask directly the user what are his/her interests and preferences, but to infer them by looking at those contents they access and to the feedback they provide about them. The ultimate goal is to learn a model from data able to link user to the vast catalog of contents made available by a large media company.

Looking at past interactions is useful to help users to discover contents that they would appreciate as valuable part of the product they paid for. This means to improve the customer retention and foster their upgrade towards more profitable products. The benefits coming from the implementation and use of these models go beyond existing contents and customers. They also help to propose new contents to existing customers, and on the other way to support new customers in discovering existing contents. Soon, new contents and new customers become part of the model, enriching the dataset of new entities, along a self-growing process. Predictiveness of models make them also suitable to support the acquisition of new contents and customers.

These models are at the core logic of recommender systems (RS), that obtained large attention once Netflix showed potentiality of algorithms in developing and supporting their streaming platform [1]. Recommender systems gained large application because of the e-commerce diffusion. They are generally grouped in different types, including Content-based recommenders [2], Collaborative recommenders [3], Demographic recommenders [4], and Hybrid recommenders [5].

The purpose of a recommender system is to provide a suggestion, regarding available alternatives, by scoring and ranking them according to the user preferences. In order to accomplish its task, a recommender system requires information regarding the user profile and habits with respect to the different alternatives that can be proposed to him. This information can be acquired explicitly by asking the users to rate items or implicitly by monitoring users' behavior (booked hotels or heard songs). RS can also use other kinds of information as demographic features (e.g, age, gender) or social information. The research related to RS has been focused on movies, music and books [6], being music recommendations the most studied topic, although later it has been applied to other e-commerce domains [7].

Similar to RS, we need data about user likings regarding catalog items such as movies, series and shows. Such information can be gathered by asking the user to rate the items, e.g., by using stars or likes, or implicitly by monitoring the customer behavior, e.g., which item enjoyed fully an which partially, how often they accessed the content description, etc. In addition we need other information regarding demographics such age, gender, family members, job, etc. The objective is to relate user profiles to content descriptors. Different techniques have been experimented in order to discover and exploit this relationship. Most of them take the form of information fusion.

Following the idea explored by [8], and more concretely the model developed in [9], we aim to build a relationship model based on the Dempster-Shafer's Theory of Evidence (D-S theory) [10], [11] and to use it to make inference regarding the relationship between users and contents. The reminder of this paper is organized as follows: Section II provides some

DISCLAIMER. This article was prepared or accomplished by Ciro Gaglione in his personal capacity. The opinions expressed in this paper are the authors' own and do not reflect the view of Sky Italia.

preliminaries regarding D-S Theory; Section III describes the model; Section IV outlines some examples of application; Section V draws conclusions and future directions.

II. Preliminaries

The Dempster-Shafer theory, also known as the Theory of Evidence [10], [11], is used as basis for the preference model presented in [9]. In D-S theory, basic probabilities are allocated to subsets, instead of elements, according to the following definitions.

Definition 1. A function $m : 2^{\Omega} \longrightarrow [0, 1]$ over a set Ω is called a *basic probability assignment* if

$$m(\emptyset) = 0 \; and \sum_{A \in 2^{\Omega}} m(A) = 1$$

Definition 2. Let Ω be a set, then $A \subseteq \Omega$ is a focal element if $m(A) > 0$. In addition, $F(\Omega) \subset 2^{\Omega}$ represents the set of focal elements induced by m.

Definition 3. Let m be a basic probability assignment function over a set Ω. The Belief of $A \subseteq \Omega$ induced by m is defined as follows

$$Bel(A) = \sum_{B \subseteq A} m(B) \tag{1}$$

Definition 4. Let m be a basic probability assignment function over a set Ω. The Plausibility of $A \subseteq \Omega$ induced by m is defined as follows

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B) \tag{2}$$

The relationship between Plausibility and Belief is given by the following equation:

$$Pl(A) = 1 - Bel(\overline{A}) \tag{3}$$

where $\overline{A}$ is the complement of A to Ω.

When the probability basic assignments are given by different sources, it is possible to combine them. The first and most common combination method is known as the Dempster's rule, that is defined as follows:

Definition 5. Let m_1 and m_2 be two basic probability assignments, the *joint basic probability assignment* is computed as

$$m_{1,2}(A) = \frac{1}{1 - Z} \sum_{B \cap C = A} m_1(B) \cdot m_2(C) \tag{4}$$

where

$$Z = \sum_{B \cap C = \emptyset} m_1(B) \cdot m_2(C) \tag{5}$$

is a measure of *conflict* between the two basic probability assignment sets. In addition, it is assumed $m_{1,2}(\emptyset) = 0$.

Belief and Plausibility are monotonic functions with respect to inclusion. This means that if we consider the lattice of Ω subsets, as shown in Fig. 1, Belief and Plausibility will increase from bottom ($Bel(\emptyset) = Pl(\emptyset) = 0$) to top ($Bel(\Omega) = Pl(\Omega) = 1$). In particular Belief and Plausibility will be kept

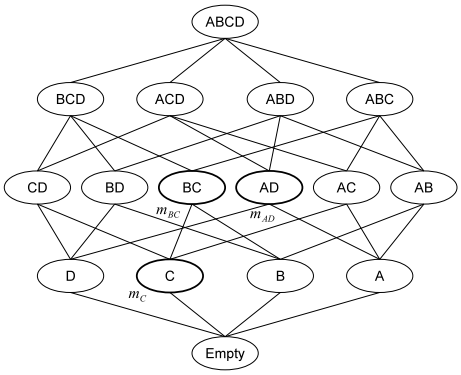

Fig. 1. The Boolean lattice of item subsets from $\Omega = \{A, B, C, D\}$ with focal elements $F(\Omega) = \{C, BC, AD\}$ [9]

constant as far as we move to nodes that do not a probability mass assigned to them. As consequence of this property, we can identify regions of connected nodes, each assuming a specific value of Belief or Plausibility, as illustrated by Fig. 2.

In this example, focal elements are C, BC and AD with the associated basic probability assignments m_C, m_{BC} and m_{AD} (assuming $m_C + m_{BC} + m_{AD} = 1$). This leads to identify 8 groups in the lattice, each with Belief and Plausibility depending from a focal subset of $F(\Omega)$. Fig. 2 outlines these regions for both Belief and Plausibility. we can observe how all portions of lattice associated to a given value of Belief or Plausibility are *connected*.

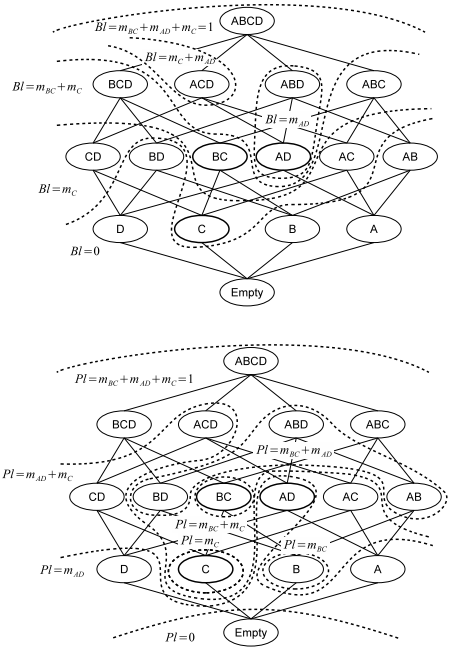

Fig. 2. Belief (top) and Plausibility (bottom) regions induced by $F(\Omega) = \{C, BC, AD\}$ [9]

If we sort the Belief (or Plausibility) values in ascending order, we get a sequence of levels, each grouping the nodes into those that are below the level and over the level. For

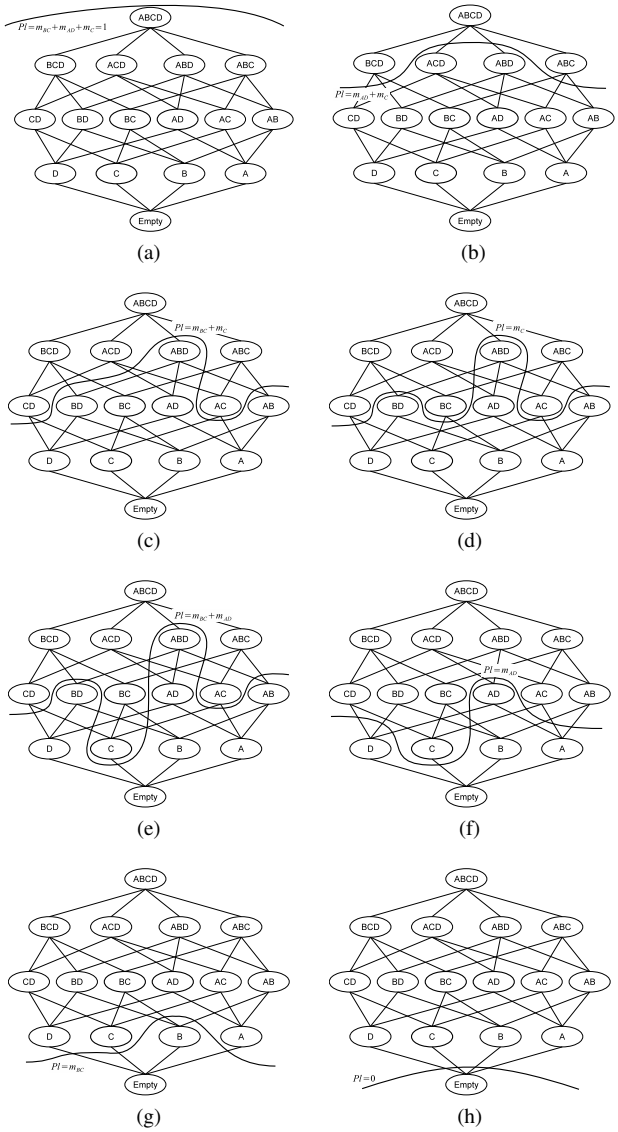

Fig. 3. Plausibility levels

instance, if we assume

$$0 \leq m_{BC} \leq m_{AD} \leq m_{BC} + m_{AD} \leq m_C$$
$$\leq m_{BC} + m_C \leq m_{AD} + m_C \leq m_{BC} + m_{AD} + m_C = 1$$

we get the situation depicted by Fig. 3 with respect to Plausibility. The following definitions enable the concept of *classes of equivalence* among the subsets with respect to Belief or Plausibility and to identify those elements that are most representative of the class.

Definition 6 (Core)**.** Given a subset $A \subseteq \Omega$, the set of focal elements included in A, *core* of A, is defined as

$$Cr(A) \stackrel{\text{def}}{=\!=} \{B \in F(\Omega) | B \subseteq A\} \tag{6}$$

Definition 7 (Support)**.** Given a subset $A \subseteq \Omega$ and the set of focal elements (even partially in A), *support* of A, is defined as

$$Su(A) \stackrel{\text{def}}{=\!=} \{B \in F(\Omega) | B \cap A \neq \emptyset\} \tag{7}$$

For instance, according to the example in Fig. 1 $F(\Omega) = \{C, BC, AD\}$, we have $Cr(BCD) = \{C, BC\} = Cr(BC)$ and $Su(ABD) = Su(BD) = Su(AB) = \{BC, AD\}$. It is straightforward that $Cr(A) \subseteq Su(A)$, for all $A \subseteq \Omega$. The core and support represent the basis for computing respectively the Belief and the Plausibility of A. The core and the support are able to group the subsets of Ω into classes of equivalence as the following definition states.

Definition 8 ($Cr-$ and $Su-$ Equivalence)**.** Two sets A and B are said to be Cr-equivalent if and only if $Cr(A) = Cr(B) = Cr$. A Cr-equivalence class is defined as the collection

$$E_{Cr} \stackrel{\text{def}}{=\!=} \{A \subseteq \Omega \mid Cr(A) = Cr\} \tag{8}$$

In addition, A and B are Su-equivalent if and only if $Su(A) = Su(B) = Su$. The Su-equivalence class obtained from this relation. is defined as

$$E_{Su} \stackrel{\text{def}}{=\!=} \{A \subseteq \Omega \mid Su(A) = Su\} \tag{9}$$

Fig. 4(a) provides an example of Cr-equivalence class assuming as core $Cr = \{BC, C\}$. Fig. 4(b) shows the Su-equivalence class for the support $Su = \{BC, AD\}$.

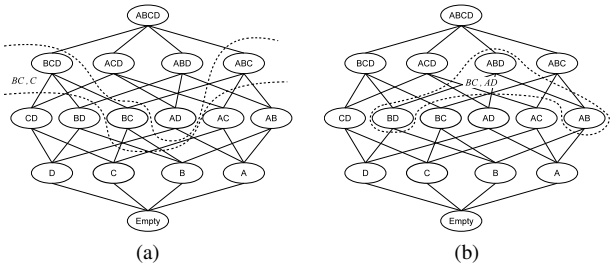

Fig. 4. Examples of Cr-equivalence (a) and Su-equivalence (b) classes

As an immediate consequence, if A and B are Cr-equivalent, then $Bel(A) = Bel(B)$, while if they are Su-equivalent, $Pl(A) = Pl(B)$.

$Cr-$ and $Su-$ equivalence classes perform a partitioning of 2^{Ω}. Thus, each subset $X \subset \Omega$ can belong only to one equivalence class. Grouping subsets in $Cr-$ and $Su-$equivalence classes allows (i) to explore the lattice by moving across classes, instead of exploring the whole item subset space, and (ii) to choose a representative of each class, so that the list of recommended items is shorter. For instance, we might be interested in using the smallest subset within a Cr-equivalence class.

As representative of a $Cr-$equivalence class we can assume the smallest subset. We call this set $Cr-$*minimal*. For instance, for the class $\{BC, ABC, BCD\}$, the core is $\{C, BC\}$ and the $Cr-$minimal is BC. It is possible to prove that each $Cr-$equivalence class as one single $Cr-$minimal. Conversely, for $Su-$equivalence classes we assume as representative the largest subset, that we call $Su-$*maximal*. Similarly to $Cr-$equivalence classes, it is possible to prove that any $Su-$equivalence class has one single $Su-$maximal. For example, the class $\{C, AD\}$, whose support is $\{ACD, AC, AD\}$, as ACD as maximal.

III. Model

In the context of our interest we assume $I = \{I_1, \ldots, I_m\}$ as the set of items belonging to the content catalog, while $U = \{U_1, U_2, \ldots, U_n\}$ as the set of users.

Both sets are projected on two feature spaces, respectively made of p and q dimensions. The first is referred to the set of characteristics describing the items in I, $C = \{C_1, \ldots, C_p\}$, while the second to the user profiling $P = \{P_1, \ldots, P_q\}$. Both spaces are discrete, so that each C_i and P_j can assume a finite number of values.

The relationship between items and users is expressed by a choice matrix, as that shown in Tab. I. The choice matrix is places side by side to the item characteristics matrix (left side) and to the profile matrix (top).

TABLE I
Structure of dataset assumed by the model

				P_q	$p_{1,q}$	$p_{2,q}$	$p_{3,q}$	$\ldots$	$p_{n,q}$
				P_q	$p_{1,q}$	$p_{2,q}$	$p_{3,q}$	$\ldots$	$p_{n,q}$
				P_q	$p_{1,q}$	$p_{2,q}$	$p_{3,q}$	$\ldots$	$p_{n,q}$
				$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
				P_q	$p_{1,q}$	$p_{2,q}$	$p_{3,q}$	$\ldots$	$p_{n,q}$
C_1	C_2	$\ldots$	C_p	I \ U	1	2	3	$\ldots$	n
$c_{1,1}$	$c_{1,2}$	$\ldots$	$c_{1,p}$	1	✓	✓		$\ldots$	
$c_{2,1}$	$c_{2,2}$	$\ldots$	$c_{2,p}$	2	✓		✓	$\ldots$	
$c_{3,1}$	$c_{3,2}$	$\ldots$	$c_{3,p}$	3		✓	✓	$\ldots$	✓
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$c_{m,1}$	$c_{m,2}$	$\ldots$	$c_{m,p}$	m	✓	✓		$\ldots$	

In general, data points $c_{i,h}$ and $p_{j,k}$ are multi-valued, meaning that they are represented by sets of values. For instance if C_h is representing the movie cast, $c_{i,h}$ is represented by the list of actors that are featuring in the movie I_i. Similarly, if P_k is "interests", $p_{j,k}$ will list what the user U_j is interested in. In other cases they are single-valued, such as in the case of characteristics such as "director" and "year" or in the case of profiling features such as "age" or "location". An example of this matrix is given in Tab.II.

Let us denote with $\Phi(C_h)$ the overall set of values assumed over the item characteristic C_h, and with $\Phi(P_k)$ the overall set of values for the user profiling feature P_k. They are respectively given in Tab.III and Tab.IV.

Since here we are interested to use both information regarding the item characteristics and the user profiles, we compute for any

$$m(K) = \frac{|L(K)|}{|L|} \tag{10}$$

where

- $K \subseteq \Phi$, with Φ being the overall set of a given characteristic C_h or a profiling feature P_k.
- $L \subseteq I \times U$ is the set of preferences ("likes")
- $L(K) \subseteq L$ is the subset of preferences referred to K

It is easy to prove that $m(\emptyset) = 0$ and $m(\Phi) = 1$. Assuming that in our example $|L| = 15$, some example of masses assigned to characteristics are given below.

- Stars: $m(De\ Niro, Bacon, Pitt) = \frac{2}{15}$
- Director: $m(Boyle) = \frac{4}{15}$
- Year: $m(1996) = \frac{5}{15}$
- Genre: $m(Drama) = \frac{3}{15}$

If we refer to profiling features, some examples are the following:

- Age: $m(30s) = \frac{8}{15}$
- Gender: $m(F) = \frac{5}{15}$
- Location: $m(IT) = \frac{11}{15}$
- Interests: $m(Movies, Books) = \frac{3}{15}$

We notice that focal elements of each dimension are given by its unique values, i.e. by rows after removing duplicates. For instance, for the "Director" and "Year" dimensions, focal elements are given by the set of director names, i.e., Boyle, Levinson, Scorsese, Howard, Zemeckis, Edwards and Scott, and by years, i.e., 1985, 1990, 1994, 1995, 1996, 2000, 2015, 2016. Similarly for "Age", i.e., 20s, 30s, 40s, "Location", i.e., IT, SP, and "Gender", i.e., M, F. They are all single-value dimensions. For them, focal elements are singletons. In this case the model becomes additive. For instance,

- $Bel(Scorsese, Boyle) = m(Scorsese) + m(Boyle)$
- $Pl(Scorsese, Boyle) = m(Scorsese) + m(Boyle)$

Instead, the dimensions "Genre" and "Stars" are multi-value, so their focal elements are not singletons. For instance, Drama, Comedy-Drama and Adventure-Drama-History are three focal elements of "Genre". Similarly, Movies-Books, Books, Sport, Music-Sport are focal elements of "Interests" among the profiling features. For multi-value dimensions the model is not additive. As an example, let us consider the belief of Adventure-Comedy-Sci-Fi-Drama. We have,

$$\begin{aligned} Bel(Adventure, Comedy, Sci-Fi, Drama) = \\ m(Adventure, Drama, Sci-Fi) + \\ m(Adventure, Sci-Fi) + \\ m(Comedy, Drama) + \\ m(Drama) = \\ \frac{1}{15} + \frac{3}{15} + \frac{1}{15} + \frac{3}{15} = \frac{8}{15} \end{aligned}$$

Conversely, we have

$$Pl(Adventure, Comedy, Sci-Fi, Drama) = 1$$

because all focal elements of "Genre" are involved in its computation.

So far, we considered each dimension in isolation. They provide a range for probability $Pr(K) = [Bel(K), Pl(K)]$, with $K \subseteq \Phi$, that is a measure of likelihood that a content in I characterized by K will be enjoyed by the set of users in U, if Φ is referred to some item characteristic C_h. Or, if we look at K as referred to some profiling feature P_k, it is the likelihood that a user in U will enjoy the catalog of contents offered by means of I.

TABLE II
THE DATASET USED AS EXAMPLE.

				Age	30s	30s	20s	40s
				Gender	M	F	M	M
				Location	IT	IT	SP	IT
				Interests	Movies Books	Sport	Books	Music Sport
Director	Year	Stars	Genre	I \ U	1	2	3	4
Boyle	1996	Ewan McGregor, Ewen Bremner	Drama	0	✓	✓	✓	
Levinson	1996	Robert De Niro, Kevin Bacon, Brad Pitt	Crime, Drama, Thriller	1	✓			✓
Scorsese	2015	Robert De Niro, Leonardo DiCaprio, Brad Pitt	Short, Comedy	2		✓		
Scorsese	1990	Robert De Niro, Ray Liotta, Joe Pesci	Biography, Crime, Drama	3			✓	
Boyle	2000	Leonardo DiCaprio	Adventure, Drama, Romance	4			✓	
Howard	1995	Tom Hanks, Kevin Bacon	Adventure, Drama, History	5		✓		✓
Zemeckis	1994	Tom Hanks	Comedy, Drama	6	✓			
Zemeckis	1985	Michael J. Fox, Christopher Lloyd	Adventure, Sci-Fi	7				✓
Edwards	2016	Felicity Jones, Diego Luna	Adventure, Sci-Fi	8		✓	✓	
Scott	2015	Matt Damon	Adventure, Drama, Sci-Fi	9		✓		

TABLE III
OVERALL SETS OF ITEM CHARACTERISTICS

Director	Year	Actors	Genre
Boyle	1996	Ewan McGregor	Crime
Levinson	2015	Ewen Bremner	Drama
Scorsese	1990	Ray Liotta	Thriller
Howard	2000	Robert De Niro	Short
Zemeckis	1995	Kevin Bacon	Comedy
Edwards	1994	Brad Pitt	Biography
Scott	1985	Leonardo DiCaprio	Adventure
	2016	Joe Pesci	Romance
		Ray Liotta	History
		Tom Hanks	Sci-Fi
		Michael J. Fox	
		Christopher Lloyd	
		Felicity Jones	
		Diego Luna	
		Matt Damon	

TABLE IV
OVERALL SETS OF USER PROFILING FEATURES

Age	Gender	Location	Interests
20s	M	IT	Books
30s	F	SP	Movies
40s		SP	Sport
			Music

If we would like to look at multiple dimensions we are not allowed to use the Dempter's combination rule as described in the section above. The main issue is that dimensions belong to different domains, so that the information fusion given by Eq.(4) cannot be performed over comparable sets. This problem can be solved when we look at focal elements as representative of preferences over the matrix L. Let K_1 and K_2 two features defined over different dimensions. We can combine the two by means of conjunction or disjunctions, depending on the semantics we associate to the operation.

Thus, in order to perform a combination of K_1 and K_2 we need to look at $L(K_1)$ and $L(K_2)$. In the case of conjunction $K = K_1 \odot K_2$, we have $L(K) = L(K_1) \cap L(K_2)$, so that

$$m(K) = \frac{|L(K_1) \cap L(K_2)|}{|L|} \tag{11}$$

For instance, if K_1 is Zemeckis and K_2 is Adventure-Sci-Fi, we have that $L(K1)$ is made of preferences at rows 6 and 7, while $L(K_2)$ at rows 7 and 8, so that $L(K_1) \cap L(K_2)$ is made only of row 7, and $m(K) = \frac{1}{15}$. Once we have focal elements for the conjunction of the "Director" and "Genre", we can compute the belief and plausibility over the conjunction of the two. For instance,

$$Bel(Zemeckis \odot Drama) = m(Zemeckis \odot Drama) = \frac{1}{15}$$

and

$$Pl(Zemeckis \odot Drama) = m(Zemeckis \odot Drama) = \frac{1}{15}$$

The meaning of $Pr(Zemeckis \odot Drama)$ is the likelihood that a drama directed by Zemeckis will be enjoyed given L and users in U, that is exactly 1 over 15.

The other way of combining two dimensions is by means of disjunction. In this case $K = K_1 \oplus K_2$ and

$$m(K) = \frac{|L(K_1) \cup L(K_2)|}{|L|} \tag{12}$$

For instance, with regard to profiling features, if K_1 is 20s and K_2 is Sport, we have

$$Bel(Sport \odot 20s) = \frac{7}{15}$$

as the conjunction of the two collect rows 0,2,3,4,5,8,9. In this case, both belief and plausibility are larger.

IV. Applications in Media Industry

The model presented so far can be employed for different tasks. We briefly outline some of them below.

Recommendations. The model can be used to suggest a content to a user according to each dimension. For instance, chosen the dimension of "Director", the system might suggest directors that are most likely be of interest for the user. It is also possible to combine different dimensions. For example, "Genre" and "Year". In any case, the inference of preferences is performed by looking at users indistinguishably, meaning that profile information is not taken into account.

Audience targeting. In this case, given a single content we are interested to find user profiles that might be interested to it. For example, given a new movie, the model might estimate how likely could be of interest for each range of age. Also in this case it is possible to combine multiple dimensions that are user profiling features. For instance, considering multiple age ranges, taking into account the different genders.

Content bundling. This application is aimed to propose a bundle of contents to a group of users, possibly with different profiles. This result can be suggested by the model through a combination of dimensions among characteristics and profiling features. The process can be led by two different perspectives. The first moves from the bundle of contents and it is aimed at identifying a group of users that might be interested in. For instance, given all drama movies in 90s, which users could be interested to such an offer. But it is also possible to move the other way round: selected a group of users, what is the bundle of contents that might be of their interest. With respect to our example, given users in the 20s that are interested to books, what is the bundle of contents that could be likely of their interest.

Segmentation. This is a generalization of the problem above. In this case both users and contents are objective of the analysis. We are interested to find clusters of users, contents and user/contents that maximize the likelihood of preferences within the group and minimize the likelihood of preferences between groups. For instance, by looking at our example, we could be interested to see if there are users with different profiles that are likely to enjoy the same contents, or if there are contents are have similar likelihood to be enjoyed by the audience, or if there groups of users that are likely to enjoy the same group of contents, besides the others.

In all tasks above, it plays a key role the possibility of comparing and ranking alternatives. However, the D-S theory provides only an imprecise probability that ranges between the lower bound given by the degree of belief and the upper bound given by the degree of plausibility. This issue can be addressed by different approaches.

The first approach is to use a degree that is representative of a range, such as the middle point between belief and plausibility. Another possibility is to use only belief degrees (conservative approach) or plausibility (challenging approach). Another approach could be to randomly choose n pairs from both ranges and to use the majority or pairwise comparisons in order to decide the order of two alternatives. It is also possible to choose randomly an alternative when the two cannot be sorted. Finally, it is possible to look at other solutions investigated in the field of partial order theory.

V. Conclusions

In this paper, we further investigated a preference model based on the Dempster-Shafer theory and its application to media industry. This work is an evolution of what has been done so far by introducing some elements of novelty. Among them the possibility of including the user profile as part of the inference, instead of being considered neutrally with respect to different applications and problems that have been discussed in the section before. There are still some issue to address. The most important is referred to scalability of the model. Indeed, the nature of the D-S theory is inherently combinatorial, so that the search space is exploding by including more elements within the dimension overall sets Φ. The possibility of defining equivalence classes in terms of belief and plausibility is a way to reduce complexity, but still work has to be done to make this solution feasible in practice. In addition, the model presented here requires to be validated. This can be done by looking at correspondences between the probability ranges and the frequency of positive voted that are after recorded. In the future we aim to develop further the model in order to include more complex queries and to solve issues regarding the application of the model in practice with respect to large catalogs and audience.

References

[1] C. A. Gomez-Uribe and N. Hunt, "The netflix recommender system: Algorithms, business value, and innovation," *ACM Trans. Manage. Inf. Syst.*, vol. 6, no. 4, pp. 13:1–13:19, Dec. 2015. doi: 10.1145/2843948. [Online]. Available: http://doi.acm.org/10.1145/2843948

[2] J. Salter and N. Antonopoulos, "Cinemascreen recommender agent: Combining collaborative and content-based filtering," *IEEE Intelligent Systems*, vol. 21, no. 1, pp. 35–41, Jan. 2006. doi: 10.1109/MIS.2006.4. [Online]. Available: http://dx.doi.org/10.1109/MIS.2006.4

[3] L. Candillier, F. Meyer, and M. Boullé, "Comparing state-of-the-art collaborative filtering systems," pp. 548–562, 2007.

[4] M. J. Pazzani, "A framework for collaborative, content-based and demographic filtering," *Artif. Intell. Rev.*, vol. 13, no. 5-6, pp. 393–408, Dec. 1999. doi: 10.1023/A:1006544522159. [Online]. Available: http://dx.doi.org/10.1023/A:1006544522159

[5] R. Burke, "Knowledge-based recommender systems," in *Encyclopedia of Library and Information Science, vol. 69*, A. Kent, Ed. Taylor and Francis, 2000, pp. 180–201.

[6] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," *Knowledge-Based Systems*, vol. 46, no. 0, pp. 109 – 132, 2013. doi: 10.1016/j.knosys.2013.03.012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0950705113001044

[7] J. J. Castro-Schez, R. Miguel, D. Vallejo, and L. M. López-López, "A highly adaptive recommender system based on fuzzy logic for {B2C} e-commerce portals," *Expert Systems with Applications*, vol. 38, no. 3, pp. 2441 – 2454, 2011. doi: 10.1016/j.eswa.2010.08.033

[8] K. Zhang and H. Li, "Fusion-based recommender system," in *Information Fusion (FUSION), 2010 13th Conference on*, July 2010. doi: 10.1109/ICIF.2010.5712091 pp. 1–7.

[9] L. Troiano, L. J. Rodríuez-Muñiz, and I. Díaz, "Discovering user preferences using dempster-shafer theory," *Fuzzy Sets and Systems*, vol. 278, pp. 98 – 117, 2015. doi: http://dx.doi.org/10.1016/j.fss.2015.06.004

[10] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *Annals of Mathematical Statistics*, vol. 38, pp. 325–339, 1967.

[11] G. Shafer, *A Mathematical Theory of Evidence*. Princeton: Princeton University Press, 1976.