A Fuzzy Linguistic RFM Model Applied to Campaign Management

Ramón Alberto Carrasco[1], María Francisca Blasco[1], Jesús García-Madariaga[1], Enrique Herrera-Viedma[2] *

[1] Department of Management and Marketing, Complutense University, Madrid 28223 (Spain)
[2] Department of Computer Science and Artificial Intelligence, University of Granada, Granada 18071 (Spain)

Received 12 November 2017 | Accepted 11 February 2018 | Published 23 March 2018

Abstract

In the literature there are some proposals for integrated schemes for campaign management based on segmentation from the results of the RFM model. RFM is a technique used to analyze customer behavior by means of three variables: Recency, Frequency and Monetary value. It is s very much in use in the business world due to its simplicity of use, implementation and interpretability of its results. However, RFM applications to campaign management present known limitations like the lack of precision because the scores of these variables are expressed by an ordinal scale. In this paper, we propose to link customer segmentation methods with campaign activities in a more effective way incorporating the 2–tuple model both to the RFM calculation process and to its subsequent exploitation by means of segmentation algorithms, specifically, k-means. This yields a greater interpretability of these results and also allows computing these values without loss of information. Therefore, marketers can effectively develop more effective marketing strategy.

Keywords

2-Tuple Model, Campaign Management, Relational Strategy, RFM.

DOI: 10.9781/ijimai.2018.03.003

I. Introduction

Most marketers have difficulty in identifying the right customers to engage in successful campaigns. Customer segmentation is a popular method that is used for selecting appropriate customers for a launch campaign. In order to link customer segmentation methods with campaign activities Chan [1] presents an approach that combines customer targeting and customer segmentation for campaign strategies based on RFM model.

RFM (Recency, Frequency and Monetary) is a model used to analyze customer behavior proposed by Hughes in 1994 [2]. "Recency" represents the length of a time period since the last purchase, while "Frequency" denotes the number of purchases within a specified time period and "Monetary value" means the amount of money spent in this specified time period. The RFM models were developed as a logical step in the evolution of marketing segmentation techniques. When the shotgun approaches (marketing everything to everyone) proved inefficient in terms of returns, the marketing campaigns started separating customers in segments based on socio-demographics attributes [3]. Commonly, RFM methods have been used to measure the Customer Lifetime Value (CLV), i.e., the predicted value a customer is going to generate in his entire lifetime [4]-[6].

With RFM analysis, organizations could discover these most valuable customers easily by observing their past behaviours [7]. In fact, these three variables belong to behavioral variables and can be used to make predictions based on the behavior in the transactional database [8]. Therefore, in a RFM process, the goal is to obtain the customer purchase behavior (the most loyal customers, dormant customers…) from these transactional data to proactively trigger appropriate direct marketing actions (retention, reactivation campaigns…) [9], [10].

* Corresponding author.

E-mail address: viedma@decsai.ugr.es

In recent years, more sophisticated statistical and data-mining techniques have been employed in direct marketing field: chi-squared automatic interaction detection (CHAID), logistic regression, neural network models, etc. Despite the deployment of these methods, marketers continue to employ RFM models. There are several reasons for the popularity of RFM among which the following are worth mentioning [11]: It is easy to use; it can generally be implemented very quickly; and is a method that managers and decision makers can understand.

McCarty and Hastak [11] have compared RFM, CHAID, and logistic regression as analytical methods for direct marketing segmentation, using two different datasets. It turns out that CHAID tends to be superior to RFM when the response rate to a mailing is low and the mailing would be send to a relatively small portion of the database. However, RFM is an acceptable procedure in other circumstances.

RFM approaches present known limitations like the lack of precision. Indeed, the scores of these RFM variables are expressed by an ordinal scale. The most common scale is the set {1,...,5} that refer to the customer contributions to revenue for enterprises. The 5 refers to the most customer contribution to revenue and 1 refers to the least contribution to revenue [12].

On the other hand, the fuzzy linguistic approach is a tool intended for modeling qualitative information in a problem. It is based on the concept of linguistic variable [13] and has been satisfactorily used in many problems [14]-[17]. The 2-tuple fuzzy linguistic approach is a model of information representation that carries out processes of "computing with words" without the loss of information [18] that has been widely used in many business and management applications [19]-[25].

In this paper, we propose to link customer segmentation methods with campaign activities in a more effective way incorporating the 2–tuple model both to the RFM calculation process and to its subsequent exploitation by means of segmentation algorithms, specifically, k-means. This yields a greater interpretability of these results and also allows computing these values without loss of information. Therefore, interpreting these linguistic results, decision makers can effectively identify valuable customers and consequently develop more effective marketing strategy. Additionally, we present an IBM SPSS Modeler [26] implementation of this model. This enables us to be more applicable at the practical level and not remain solely confined to the theoretical one.

The rest of the paper is organized as follows: Section II revises the preliminary concepts, i.e., the integrated scheme of customer segmentation with campaign activities using the RFM model and 2-tuple model. In Section III we propose to modify this integration scheme by incorporating the 2-tuple model in two directions: in the RFM scores computation and the subsequent segmentation algorithm. Additionally we show an implementation and use case of this new model using IBM SPSS Modeler comparing it with the previous one. Finally, we point out some concluding remarks and future work.

II. Preliminaries

In this section we present the basic elements needed to understand our new proposal: an integrated scheme of customer segmentation with campaign activities based on the RFM model and the 2-tuple fuzzy linguistic approach.

A. Integrated Scheme of Customer Segmentation with Campaign Activities

The RFM analytic approach is a common model that identifies customer purchase behavior, i.e., that differentiates important customers from large data by three variables [9]:

- **Recency** (R): The time (in units such as days, months, years…) since the most recent purchase transaction or shopping visit.
- **Frequency** (F): The total number of purchase transactions or shopping visits in the period examined.
- **Monetary** value (M): The total value of the purchases within the period examined.

In order to link customer segmentation methods with campaign activities [1] the following scheme that integrates the RFM model (Fig. 1):

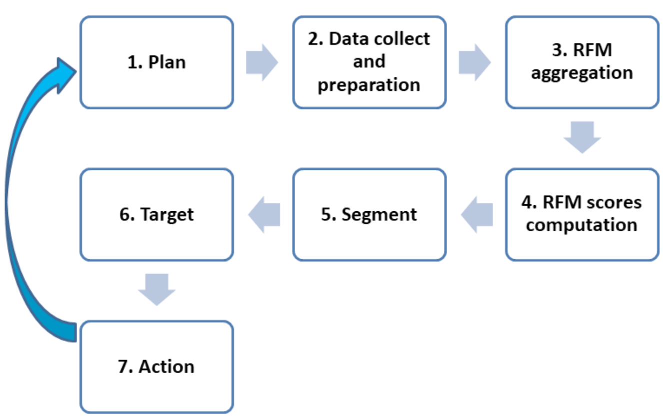

Fig. 1. Integrated scheme of customer segmentation with campaign activities (based on [1]).

1. **Plan**. Conducting a campaign requires first determining series of sequential marketing plans to establish the corresponding relational strategy (Kang 2015), for example, aimed at promoting customer loyalty or preventing customer churn. We can formalize this plan using the following table of campaigns: *Campaigns (CampaignID, RelationalStrategyDES)*, where *CampaignID* contains the campaign code and *RelationalStrategyDES* is the description of the relational strategy of that campaign.
2. **Data collect and preparation**. The second step is collecting related customer information. Once we have chosen the period to analyze, the customers are selected if they have at least purchased during this period. Transactional data on these customers must be retrieved, audited, cleaned, and prepared for subsequent operations. Let *Transactions (CustomerID, Date, Amount)* be the table where these transactional data on purchases are included. The customer code is stored in *CustomerID*. *Date* and *Amount* are the corresponding date and amount spent by the customer in the purchase.
3. **RFM aggregation**. Transactional data is aggregated at a customer level, i. e., on the *CustomerID* attribute. Thus, we obtain the table *CustomerTransactions (CustomerID, Recency, Frequency, Monetary)* with the RFM information summarized for each customer identified by *CustomerID*. *Recency* would be the days since the last purchase of such customer (using a later fixed reference date for all customer purchases). *Frequency* is the number of times the customer has purchased. *Monetary* contains the total amount of those purchases.
4. **RFM scores computation**. Customers are sorted according to the respective RFM measure and are grouped in classes of equal size, typically quintiles. Customers are sorted independently according to each of the individual RFM components and then binned into five groups of 20 per cent. This result is included into the table *CustomerRFM (CustomerID, RecencyScore, FrequencyScore, MonetaryScore, RFMScore)* with *RecencyScore, FrequencyScore, MonetaryScore* ∈ {1,…, 5}. Therefore the RFM measures are transformed into ordinal scores such that the value 1 includes the 20% of customers with the worst values and the 5 the 20% of customers with the best values in the corresponding measure. Especially for the *Recency* attribute, the scale of the derived ordinal score, *RecencyScore*, should be reversed so that larger scores represent the most recent buyers. Sometimes it can be useful to have a unique measure, *RFMScore*, which characterizes together the RFM scores. In order to provide this continuous RFM score, the R, F, and M bins are summed, with appropriate user-defined weights, i.e., w_R, w_F, w_M. The RFM score is the weighted average of its individual components and is calculated as follows:

$$RFMScore = RecencyScore \times w_R + FrequencyScore \times w_F + MonetaryScore \times w_M. \quad (1)$$

5. **Segment**. Once the results of the previous step are validated, the marketers of the enterprise apply this RFM knowledge in order to search the most suitable customer group for each plan campaign. For this, segmentation or clustering techniques are especially useful. Clustering or segmentation is the process of grouping a set of objects into groups of similar objects. In this way, clustering based on RFM scores of the table *CustomerRFM* provides more behavioral knowledge of customers' actual marketing levels than other cluster analyses [27].

K-means is one of the well-known algorithms for clustering [28], [29] of which various modifications have been proposed including fuzzy logic [30]. In this algorithm each cluster is characterized by its center point i.e. centroid. K-means is a partitioning cluster algorithm by grouping n vectors (customers in our case) based on attributes into k partitions, where $k < n$, according to some measure, usually Euclidean distance. The name comes from the fact that k clusters are determined and the center of a cluster is the mean of all vectors within this cluster. The algorithm starts with k initial centroids, then assigns vectors to the nearest centroid using Euclidean distance and re-computes the new centroids as means of the assigned data vectors. This process is repeated over and over again until vectors no longer changed clusters between iterations [27], [31]. Thus, using a k-means algorithm, the centroid results of this algorithm are: $v_s = (v_{s1}, v_{s2}, v_{s3})$, with $s = 1..k$, one for each cluster. These centroids are quite interpretable from the point of view of business as explained in the previous stage (5 best values and 1 the worst).

6. **Target**. It is necessary to identify the most profitable groups of customers for each campaign plan. Thus, once the segmentation is concluded and validated, marketers should determine the targeted clusters that can be associated with the subsequent campaign and then get the clients that belong to those groups which are stored in the table *CustomerTarget (CustomerID, CampaignID)*.
7. **Action**. The last step is to implement effective campaign management oriented selected target.

Much of aforementioned approach can be solved with several data science or data mining tools. In Fig. 2 we show an example using IBM SPSS Modeler [26].

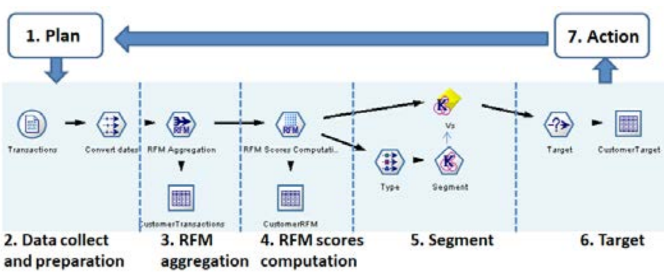

Fig. 2. Integrated scheme of customer segmentation with campaign activities with IBM SPSS Modeler.

Following, we explain each stage of this stream:

1. **Plan**. Based on Ref. [1], in Table I we present the set of campaigns to be carried out.

TABLE I. Campaign Marketing Plan Included in Campaigns Table

CampaignID	*RelationalStrategyDES*
Best	They are the most valuable clients for the company. The relational strategy will be aimed at managing the relationship in order to maintain the value of the clients. For example, we could offer them free services.
New/ Reactivates	They are new clients or ex-clients that have been reactivated in the period analyzed. Welcome gifts, bonus... could be offered associated with the next purchase.
Growing	They are a low-value client which it is necessary to carry out a growth strategy by up-selling or cross-selling.
Churn	The customers who are possible to leave or turn to other competitors. The strategy could be retention by means special discounts…
Worse	They are clients with a minimum degree of relationship with the company so they can be considered ex-clients. The relational strategy will be similar to that of leads, that is, acquisition. Free trial could be offered for example.

2. **Data collect and preparation**. Transactional data of the two last years were retrieved, audited, cleaned, and prepared (casting to date type) for subsequent stages. Inactive customers with no purchases during this period were not included into the *Transactions* table. We have based this example on a file obtained based on data referenced in [32] with 69659 purchase transactions, corresponding to $n = 23570$ distinct customers.
3. **RFM aggregation**. For this step we use the RFM *Aggregate* node (labeled as RFM *Aggregation* in Fig. 2) that simplifies the computation of this stage. We only have to designate the required transaction fields (*CustomerID*, *Date* and *Amount*) and the fixed date to compute *Recency* as the time of difference (days, hours, minutes or seconds) between *Date* and this date (2018-01-01), see Fig. 3.

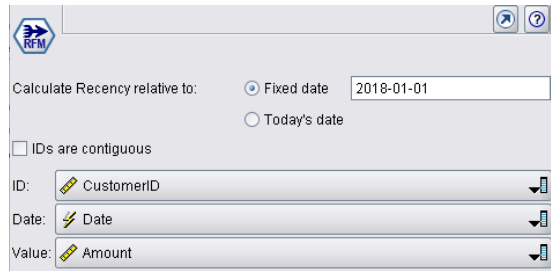

Fig. 3. Detail of the *RFM Aggregate* node settings.

4. **RFM scores computation**. IBM SPSS Modeler also offers a node named *RFM Analysis* (called *RFM scores computation* in Fig. 2) that can directly group the R, F, and M measures into the selected number of quantiles (five in our case). This node also computes the *RFMScore* using the Eq. (1) (with w_R=1/3, w_F=1/3 and w_M= 1/3), see Fig. 4.

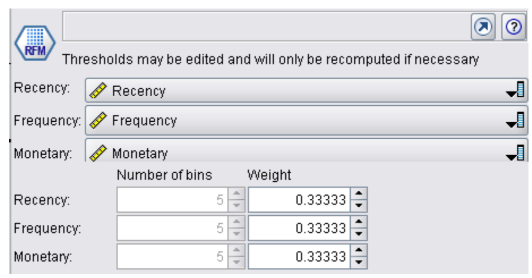

Fig. 4. Detail of the *RFM Analysis* node settings.

Before performing segmentation using scores may address certain plan campaigns, for example, in order to identify the most valuable customers (campaign identified by *Best* in Table I). This could be solved with the RFM model sorting in descending order by the field *RFMScore*. As can be seen in Table II, there are many clients with equal score because when grouping customers in quintiles the procedure results in a total of 5 × 5 × 5 = 125 distinct values as much of *RFMScore*. This lack of precision can be a problem when selecting customers for the different campaigns.

TABLE II.
Detail of the *CustomerRFM* Table Ordered by *RFM Score* Descending (the First 2640 Clients Have the Same *RFMScore* = 5)

	CustomerID	Recency	Frequency	Monetary	Recency Score	Frequency Score	Monetary Score	RFM Score
1	22524	7153	15	185.830	5	5	5	5.000
2	8957	7243	26	345.100	5	5	5	5.000
3	10399	7157	16	228.870	5	5	5	5.000
4	8969	7160	42	630.540	5	5	5	5.000
5	8970	7166	24	400.380	5	5	5	5.000
6	8975	7239	16	182.440	5	5	5	5.000
7	8976	7216	12	175.300	5	5	5	5.000
8	8986	7193	24	277.410	5	5	5	5.000
9	8989	7185	17	255.140	5	5	5	5.000
10	5195	7221	11	137.460	5	5	5	5.000
…								
2640	17142	7167	23	388.310	5	5	5	5.000
2641	12635	7380	13	179.530	4	5	5	4.667
2642	19377	7434	12	253.510	4	5	5	4.667

5. **Segment**. IBM SPSS Modeler provides the appropriate module for k-means. We specify in its settings the number of groups to obtain, this is, i.e. $k = 5$. Before using this node, you must specify with a *Type* node which variables are going to be used in the segmentation, i.e. the RFM scores. We proceed to analyze the different clusters obtained, associating them, if possible, to the campaigns of the specified plan. In Table III, we show the results of this process of clustering and the association to plan.

TABLE III. Results of the K-means Clustering

S	C_s	Recency Score v_{s1}	Frequency Score v_{s2}	Monetary Score v_{s3}	RFM Patter	Campaign ID
cluster-1	2440	3.72	1.97	2.25	R↑F↓M↓	New/ Reactivates
cluster-2	4067	2.74	2.85	3.11	R↕F↕M↕	Growing
cluster-3	7054	4.58	4.56	4.60	R↑F↑M↑	Best
cluster-4	1921	2.11	4.03	4.19	R↓F↑M↑	Churn
cluster-5	8088	1.73	1.23	1.48	R↓F↓M↓	Worse

6. **Target**. Simply, we select those clients that belong to the chosen cluster.
7. **Action**. Based on the description of the strategy that has been made in Table I, the action of each of the proposed campaigns would be implemented for each customer.

B. The 2-Tuple Model

The *2-tuple fuzzy linguistic approach* [18] is a continuous model of information representation that has been used in many business and management applications [19]-[25]. This model carries out processes of "computing with words" without the loss of information which are typical of other fuzzy linguistic approaches. Following, we explain the basic notations and operational laws to understand our proposal:

Let $S = \{s_0,\ldots,s_T\}$ be a linguistic term set with odd cardinality, where the mid-term represents a indifference value and the rest of terms are symmetric with respect to it. We assume that the semantics of labels are given by means of triangular membership functions and consider all terms distributed on a scale on which a total order is defined, i.e. $s_i \leq s_j \Leftrightarrow i < j$. In this fuzzy linguistic context, if a symbolic method aggregating linguistic information obtains a value $b \in [0,T]$, and $b \notin \{0,\ldots,T\}$, then an approximation function is used to express the result in S.

Definition 1 [18]. *Let b be the result of an aggregation of the indexes of a set of labels assessed in a linguistic term set S, i.e. the result of a symbolic aggregation operation, $b \in [0,T]$. Let $i = round(b)$ and $\alpha = b-i$ be two values, such that $i \in [0,T]$ and $\alpha \in [-0.5,0.5)$, then α is called a Symbolic Translation.*

The 2-tuple fuzzy linguistic approach [18] is developed from the concept of symbolic translation by representing the linguistic information by means of 2-tuple (s_i, α_i), $s_i \in S$ and $\alpha_i \in [-0.5, 0.5)$, where s_i represents the information linguistic label, and α_i is a numerical value expressing the value of the translation from the original result b to the closest index label, i, in the linguistic term set S. The value (s_i, α_i) also can be represented as $s_i \pm \alpha_i$ (+ or - depending on the sign of α_i).

This model defines a set of transformation functions between numeric values and 2-tuple:

Definition 2 [18]. *Let $S = \{s_1,\ldots,s_T\}$ be a linguistic term set and $b \in [0,T]$ a value representing the result of a symbolic aggregation operation, then the 2-tuple that expresses the equivalent information to b is obtained with the following function:*

$$\Delta: [0,T] \rightarrow S \times [-0.5,0.5)$$

$$\Delta(b) = (s_i, \alpha), \textit{ with } \begin{cases} s_i, & i = round(b) \\ \alpha = b - i, & \alpha \in [-0.5,0.5). \end{cases} \tag{2}$$

where round(·) is the usual round *operation, s_i has the closest index label to b and α is the value of the symbolic translation.*

For all Δ, there exists Δ^{-1}, defined as:

$$\Delta^{-1}(s_i, \alpha) = i + \alpha. \tag{3}$$

The negation operator is defined as:

$$neg((s_i, \alpha)) = \Delta(T - (\Delta^{-1}(s_i, \alpha))). \tag{4}$$

Information aggregation consists of obtaining a value that summarizes a set of values. Hence, the result of the aggregation of a set of 2-tuples must be a 2-tuple. Using the functions Δ and Δ^{-1} that transform numerical values into linguistic 2-tuples and vice versa without loss of information, any of the existing aggregation operators can be easily extended for dealing with linguistic 2-tuples. Below, we describe the aggregation operators which we use in our model:

Definition 3. *Let $A = \{(l_1, \alpha_1),\ldots, (l_n, \alpha_n)\}$ be a set of linguistic 2-tuple and $W = \{w_1,\ldots, w_n\}$ be their associated weights. The 2-tuple weighted average $\bar{A}^w$ is:*

$$\bar{A}^w[(l_1, \alpha_1),\ldots, (l_n, \alpha_n)] = \Delta\left(\frac{\sum_{i=1}^{n} \beta_i \cdot w_i}{\sum_{i=1}^{n} w_i}\right) \tag{5}$$

Definition 4. *Let $A = \{(l_1, \alpha_1),\ldots, (l_n, \alpha_n)\}$ be a set of linguistic 2-tuple. The 2-tuple average $\bar{A}$ is:*

$$\bar{A}[(l_1, \alpha_1),\ldots, (l_n, \alpha_n)] = \Delta\left(\frac{\sum_{i=1}^{n} \beta_i}{n}\right) \tag{6}$$

III. Applying the 2-Tuple Approach for Campaign Management

As explained in Section II.A, although *RFM* analysis is a very *useful tool for campaign management*, it has its limitations such as its lack of precision in the calculation of scores. This is due to the representation as an ordinal number of these RFM scores (for example, see the Table II where you cannot identify which are really the best customers). In this section, we propose to incorporate the 2-tuple model in order to improve this campaign management. This is possible because by incorporating the 2-tuple model we will get results from the RFM model with more linguistic interpretability and above all with more precision.

The campaign management scheme followed is the same as shown in Fig. 1 and explained in Section II.A but changing stages 3 and 4 as explained in the following two sub sections in which we also show the implementation in SPSS Modeler of the model:

A. RFM Scores Computation

The basic idea is to compute and store the scores included into the output table of this step (*CustomerRFM*), i.e., *RecencyScore, FrequencyScore, MonetaryScore* and *RFMScore* using the 2-tuple model [25].

First, we need to define the symmetric and uniformly distributed domain S using five linguistic labels. These labels have a semantic meaning for these four variables of the RFM model referred to the degree of agreement on the goodness of the variable:

Let $S = \{s_0,\ldots, s_T\}$, $T = 4$: $s_0 =$ *Strongly Disagree = SD*, $s_1 =$ *Disagree = D*, $s_2 =$ *Neutral = N*, $s_3 =$ *Agree = A*, and $s_4 =$ *Strongly Agree = SA*, with the definition showed in Fig. 5.

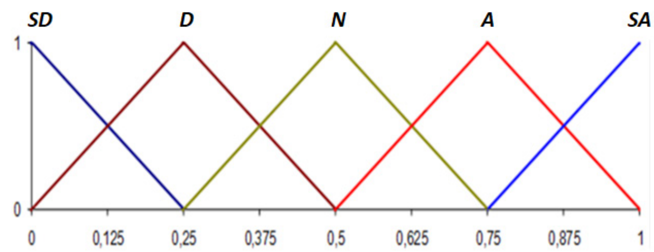

Fig. 5. Definition of the set S.

Therefore, we have the variables to calculate: *RecencyScore, FrequencyScore, MonetaryScore, RFMScore* $\in S \times [-0.5, 0.5)$.

For each customer $i = 1,\ldots, n$, we obtain $A_i = (A_{i1}, A_{i2}, A_{i3})$ with $A_{i1} = RecencyScore_i$, $A_{i2} = FrequencyScore_i$ and $A_{i3} = MonetaryScore_i$. Firstly, customers are sorted in ascending order according to each of the individual RFM components $B_i = (B_{i1}, B_{i2}, B_{i3})$, with $B_{i1} = Recency_i$, $B_{i2} = Frequency_i$ and $B_{i3} = Monetary_i$, stored in *CustomerTransactions* (obtained as explained in phase 3 of Section II.A). Now, we define $rank_{ij} \in \{1,\ldots, n\}$ as the ranking of each client respect to each of these variables:

$$percent\_rank_{ij} = (rank_{ij}-1) / (n-1)$$

with $percent\_rank_{ij} \in [0, 1]$, $i = 1,\ldots, n$, $j = 1,\ldots, 3$ and $n > 1$. The final 2-tuple score A_{ij} is obtained as following:

$$A_{ij} = \begin{cases} \Delta((percent\_rank_{ij}), \text{ if } j \neq 1 \\ \\ neg(\Delta(percent\_rank_{ij})), \text{ if } j = 1 \end{cases} \qquad (7)$$

where $\Delta(\cdot)$ and $neg(\cdot)$ have been defined in Section II.B (Eq. 2 and 4 respectively). We use the negation function on *Recency* because the larger scores represent the most recent buyers.

The 2-tuple $RFMScore_i$, which characterizes together the RFM scores, is calculated for each i-customer using the Eq. (5) as follows:

$$RFMScore_i = \bar{A}^w\, [A_{ij}] \qquad (8)$$

with the user-defined weights $W = \{w_R, w_F, w_M\}$.

In a previous paper [24] the authors have proposed both a representation data type 2-tuple as the implementation of the functions Δ and Δ^{-1} using IBM SPSS Modeler. Using these tools, the 2-tuple approach proposed in this paper has been implemented. Thus, the stream to solve the example of the Section II.A is showed in the Fig. 6.

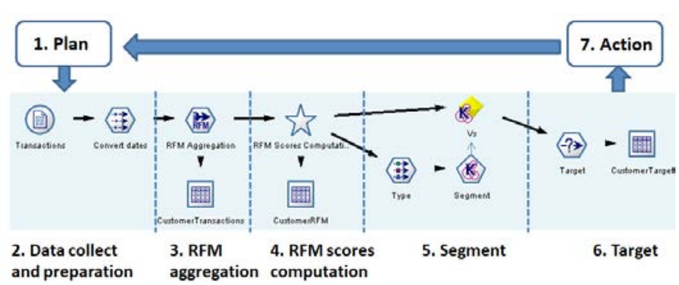

Fig. 6. Integrated scheme of customer segmentation with campaign activities with IBM SPSS Modeler using the 2-tuple model.

This stream is very similar to the presented in the Fig. 2 being its main difference the implementation of the fourth stage. For this purpose we have created a super node (symbolized by ☆) named *RFM Scores Computation* (shown in Fig. 7). A super node is similar to a procedure with inputs (labeled with *From Stream*) and/or outputs values (labeled with *To Stream*). This super node computes de 2-tuple RFM scores (using the Eq. 7 and 8) and it also computes the corresponding Δ-1 function (Eq. 3) on these values to apply conventional numerical operations necessary in the fifth stage.

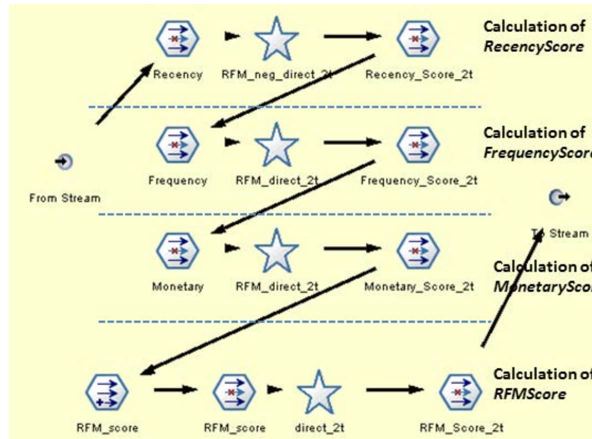

Fig. 7. IBM SPSS super node for RFM scores computation.

We execute the new version of the stream (until 4th phase in Fig. 6) on the same input data and the same user-defined weights (w_R, w_F, w_M) used on the conventional stream (Fig. 2). The selection of the best customers, i.e., the highest Δ^{-1}*(RFMScore)* is presented in the Table IV. Also in this table we show the RFM scores that were obtained according to the conventional process. In the 2-tuple implementation the interpretability of the scores is easier as they are expressed by linguistic labels instead of ordinal numbers. Also the accuracy of such scores is greater, owing the 2-tuple model, allowing a better prioritization (selection) of the best customer in order to identify the most valuable customers (campaign *Best* in Table I).

Table IV.
Results of the Conventional RFM Process Vs 2-Tuple RFM Process Ordered by Vs 2-Tuple RFM Score Descending (Top 20 Customers)

					conventional RFM process				2-tuples RFM process			
	CustomerID	Recency	Frequency	Monetary	Recency Score	Frequency Score	Monetary Score	RFM Score	Recency_Score	Frecuency_Score	Monetary_Score	RFM_Score
1	14048	185	1033	8976.330	5	5	5	5.000	SA-0.001188	SA	SA-0.000042	SA-0.000210
2	7592	186	917	13990.930	5	5	5	5.000	SA-0.003479	SA-0.000042	SA	SA-0.000526
3	3049	186	301	4262.850	5	5	5	5.000	SA-0.003479	SA-0.000339	SA-0.000339	SA-0.000810
4	2908	186	188	2914.870	5	5	5	5.000	SA-0.003479	SA-0.000721	SA-0.000849	SA-0.001230
5	8250	187	161	2582.740	5	5	5	5.000	SA-0.005622	SA-0.001188	SA-0.001018	SA-0.001726
6	3901	187	174	2411.530	5	5	5	5.000	SA-0.005622	SA-0.000976	SA-0.001188	SA-0.001832
7	7983	190	536	6973.070	5	5	5	5.000	SA-0.011880	SA-0.000085	SA-0.000085	SA-0.001854
8	1965	187	122	2010.980	5	5	5	5.000	SA-0.005622	SA-0.002143	SA-0.001739	SA-0.002362
9	16517	191	299	4185.100	5	5	5	5.000	SA-0.013407	SA-0.000382	SA-0.000424	SA-0.002367
10	2484	186	128	1784.360	5	5	5	5.000	SA-0.003479	SA-0.001994	SA-0.002291	SA-0.002440
11	21950	190	204	2982.800	5	5	5	5.000	SA-0.011880	SA-0.000636	SA-0.000806	SA-0.002450
12	6569	192	322	4968.000	5	5	5	5.000	SA-0.015189	SA-0.000297	SA-0.000212	SA-0.002487
13	12352	188	152	2102.920	5	5	5	5.000	SA-0.008040	SA-0.001315	SA-0.001612	SA-0.002547
14	20250	185	107	1592.760	5	5	5	5.000	SA-0.001188	SA-0.003034	SA-0.002970	SA-0.002709
15	10902	188	149	1966.710	5	5	5	5.000	SA-0.008040	SA-0.001379	SA-0.001824	SA-0.002712
16	710	190	130	2502.300	5	5	5	5.000	SA-0.011880	SA-0.001782	SA-0.001061	SA-0.002756
17	398	185	118	1569.330	5	5	5	5.000	SA-0.001188	SA-0.002397	SA-0.003140	SA-0.002773
18	22348	190	195	2356.890	5	5	5	5.000	SA-0.011880	SA-0.000679	SA-0.001315	SA-0.002836
19	8956	189	138	2131.020	5	5	5	5.000	SA-0.010289	SA-0.001506	SA-0.001527	SA-0.002839
20	499	194	328	4378.550	5	5	5	5.000	SA-0.019177	SA-0.000255	SA-0.000297	SA-0.003125

B. Segment

The main problem with the previous approach (4th phase in Section II.A) is the lack of precision in the representation of each individual customer (RFM scores). Consequently the results (centroids) obtained in the next stage are also imprecise. In this section, we propose applying the 2-tuple fuzzy linguistic RFM approach to customer segmentation (using k-means) to obtain more accurate results. On the other hand, it will also increase the interpretability of these centroids as we use linguistic values.

The scores obtained in the previous step (Section III.A) are 2-tuple values: $A_i = (A_{i1}, A_{i2}, A_{i3})$ with $A_{i1} = RecencyScore_i$, $A_{i2} = FrequencyScore_i$

and A_{i3} =$MonetaryScore_i$.

The objective is to *obtain the centroids* $v_s = (v_{s1}, v_{s2}, v_{s3})$ with $s = 1..k$, one for each cluster. The values of these centroids will be expressed using model 2-tuple model, thus we get a better linguistic interpretability. In order to apply the algorithm, we need to get the distance between customers and these centroids. We propose to use the Euclidean distance d_E following:

$$d_E(i,s) = \sqrt{\begin{aligned}&(\Delta^{-1}(A_{i1}) - \Delta^{-1}(v_{s1}))^2 + \\ &(\Delta^{-1}(A_{i2}) - \Delta^{-1}(v_{s2}))^2 + \\ &(\Delta^{-1}(A_{i3}) - \Delta^{-1}(v_{s3}))^2\end{aligned}}$$

for each customer $i = 1..n$ and for each cluster $s = 1..k$.

In each step of the k-mean algorithm, we recalculate the new cluster center *vs* using the Eq. 6:

$$v_s = (\bar{A}\,[A_{r1}], \bar{A}\,[A_{r2}], \bar{A}\,[A_{r3}])$$

with $r = 1..\ c_s$, that symbolizes the *r*-customer such that belongs to the *s*-cluster.

Following the example used previously, we show the result of our 2-tuple model in Table V after executing the proposed algorithm. With our model the linguistic interpretability of the clusters (centroids) is better (see the linguistic labels included in Fig. 5). But the main advantage is that these results are also more accurate as we have already commented. We can see how the distribution of groups with our model is more equitable than the conventional process, where the largest groups are those that contain the worst and best customers (see Table III).

TABLE V. Results of the 2-tuple K-means Clustering

S	C_s	Recency Score v_{s1}	Frequency Score v_{s2}	Monetary Score v_{s3}	RFM Patter	Campaign ID
cluster-1	3139	A-0.015419	N+0.018895	N+0.004251	R↑ F↕M↕	New/ Reactivates
cluster-2	5868	D+0.078401	D+0.092712	D+0.114528	R↓F↓M↓	Growing
cluster-2	5913	A+0.088279	A+0.111631	A+0.109542	R↑F↑M↑	Best
cluster-3	3276	N-0.078436	A-0.065937	A-0.060768	R↕F↑M↑	Churn
cluster-1	5374	D-0.024133	D-0.099303	D-0.115425	R↓F↓M↓	Worse

Therefore, our model could get a more appropriate and effective campaign plan.

IV. Concluding Remarks and Future Work

RFM [2] is a technique widely used a lot more now in marketing due to its simplicity of use, implementation and interpretability of its results. Even its results are better, in a practical level, than other more sophisticated techniques as CHAID and logistic regression in specific circumstances [11]. However, RFM applications to direct marketing present known limitations like the lack of precision.

In this context, we have presented an integrated relational campaign management scheme based on RFM analytic process that incorporates the 2-tuple model in order to obtain a higher precision and an easier linguistic interpretability of the RFM model results and the subsequent segmentation, in order to develop a more effective campaign plan.

Additionally, we have presented an IBM SPSS Modeler implementation of this model. In such a way, our proposal could be widely applied at a practical level on several marketing problems of this type. As an example, we have applied the implemented model on a well-known data set verifying the advantages of the new model regarding the conventional campaign management scheme.

We are currently focusing on the use of this model to several marketing problems, especially in banking industry.

Acknowledgment

This paper has been elaborated with the financing of FEDER funds in the Spanish National research project (TIN2013-40658-P), Spanish Department for Economy and Competitiveness project (TIN2016-75850-R).

References

[1] C. C. H. Chan, "Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer," *Expert systems with applications*, vol. 34, no. 4, pp. 2754-2762, 2008.

[2] A. M. Hughes, "Strategic database marketing," Chicago: Probus Publishing Company, 1994.

[3] J. Bernabé-Moreno, A. Tejeda-Lorente, C. Porcel, E. Herrera-Viedma, "A new model to quantify the impact of a topic in a location over time with Social Media," *Expert Systems with Applications*, vol. 42, no. 7, pp. 3381–3395, 2015.

[4] D. R. Liu, Y. Y. Shih, "Integrating AHP and data mining for product recommendation based on customer lifetime value," *Information & Management* vol. 42, no. 3, pp. 387-400, 2005.

[5] M. Khajvand, K. Zolfaghar, S. Ashoori, S. Alizadeh, "Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study," *Procedia Computer Science*, vol. 3, pp. 57–63, 2011.

[6] B. Sohrabi, A. Khanlari, "Customer lifetime value (CLV) measurement based on RFM model," *Iranian Accounting & Auditing Review*, vol. 14, no. 47, pp. 7–20, 2007.

[7] Y. M. Li, C. H. Lin, C. Y. Lai, "Identifying influential reviewers for word-of-mouth marketing," *Electronic Commerce Research and Applications*, vol. 9, no. 4, pp. 294–304, 2010.

[8] W. Jo-Ting, L. Shih-Yen, W. Hsin-Hung, "A review of the application of RFM model," *African Journal of Business Management*, vol. 4, no. 19, pp. 4199–4206, 2010.

[9] K. Tsiptsis, A. Chorianopoulos, "Data mining techniques in CRM: inside customer segmentation,", John Wiley & Sons, 2011.

[10] W. Buckinx, D. Van den Poel, "Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting," *European Journal of Operational Research*, vol. 164, no. 1, pp. 252–268, 2005.

[11] J. A. McCarty, M. Hastak, "Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression," *Journal of business research*, vol. 60, no. 6, pp. 656–662, 2007.

[12] C. H. Cheng, Y. S. Chen, "Classifying the segmentation of customer value via RFM model and RS theory," *Expert systems with applications*, vol. 36, no. 3, pp. 4176–4184, 2009.

[13] L. A. Zadeh, "The concept of a linguistic variable and its applications to approximate reasoning," *Inf Sci*, vol. 8, pp. 199–249, 1975.

[14] F. J. Cabrerizo, R. Al-Hmouz, A. Morfeq, A. S. Balamash, M. A. Martínez, E. Herrera-Viedma, "Soft consensus measures in group decision making using unbalanced fuzzy linguistic information," *Soft Computing*, vol. 21, no. 11, pp. 3037-3050, 2017.

[15] J. A. Morente-Molinera, J. Mezei, C. Carlsson, E. Herrera-Viedma, "Improving supervised learning classification methods using multi-granular linguistic modelling and fuzzy entropy," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 5, pp. 1078-1089, 2017.

[16] S. Alonso, I. J. Pérez, F. J. Cabrerizo, E. Herrera-Viedma, "A Linguistic Consensus Model for Web 2.0 Communities," *Applied Soft Computing*, vol. 13, no. 1, pp. 149-157, 2013.

[17] S. Massanet, J. V. Riera, J. Torrens, E. Herrera-Viedma, "A new linguistic computational model based on discrete fuzzy numbers for computing with words," *Information Sciences*, vol. 258, pp. 277-290, 2014.

[18] F. Herrera, L. Martínez, "A 2-tuple fuzzy linguistic representation model

for computing with words," *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 6, pp. 746–752, 2000.

[19] A. Cid-López, M. J. Hornos, R. A. Carrasco, E. Herrera-Viedma, "Applying a linguistic multi-criteria decision-making model to the analysis of ICT suppliers' offers," *Expert Systems with Applications*, vol. 57, pp. 127-138, 2016.

[20] A. Cid-López, M. J. Hornos, R. A. Carrasco, E. Herrera-Viedma, F. Chiclana, "Linguistic multi-criteria decision-making model with output variable expressive richness," *Expert Systems with Applications*, vol. 83, pp. 350-362, 2017.

[21] A. Cid-López, M. J. Hornos, R. A. Carrasco, E. Herrera-Viedma, "A hybrid model for decision-making in the Information and Communications Technology sector," *Technological and Economic Development of Economy*, vol. 21, no. 5, pp. 720–737, 2015.

[22] R. A. Carrasco, F. Muñoz-Leiva, J. Sánchez-Fernández, F. J. Liébana-Cabanillas, "A model for the integration of e-financial services questionnaires with SERVQUAL scales under fuzzy linguistic modeling," *Expert Systems with Applications*, vol. 39, no. 14, pp. 11535–11547, 2012.

[23] R. A. Carrasco, J. Sánchez-Fernández, F. Muñoz-Leiva, M. F. Blasco, E. Herrera-Viedma, "Evaluation of the hotels e-services quality under the user's experience," *Soft Computing*, vol. 21, no. 4, pp. 995-1011, 2017.

[24] R. A. Carrasco, M. F. Blasco, E. Herrera-Viedma, "An Implementation of a Linguistic Multi-Criteria Decision Making Model: An Application to Tourism," *Lecture Notes in Artificial Intelligence*, vol. 8536, pp. 232–239, 2014.

[25] R. A. Carrasco, M. F. Blasco, E. Herrera-Viedma, "A 2-tuple fuzzy linguistic RFM model and its implementation," *Procedia Computer Science*, vol. 55, pp. 1340-1347, 2015.

[26] IBM SPSS Modeler. URL (consulted 2018): http://www.ibm.com/software/analytics/spss/products/modeler/

[27] D. Birant, "Data Mining Using RFM Analysis," in *Knowledge Oriented Applications in Data Mining*, K. Funatsu and Hasegawa, K., Eds., Rijeka, Croatia: InTech , pp. 91–108, 2011.

[28] N. Settouti, M. E. A. Bechar, M. A. Chikh, "Statistical comparisons of the top 10 algorithms in data mining for classification task," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no. 1, pp. 46-5, 2016.

[29] A. M. Jalil, I. Hafidi, L. Alami, E. Khouribga,"Comparative study of clustering algorithms in text mining context," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 3, no. 7, pp. 42-45, 2016.

[30] B. S. Harish, B. S. Kumar, "Anomaly based Intrusion Detection using Modified Fuzzy Clustering," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no. 6, pp. 54-59, 2017.

[31] C. H. Cheng, Y. S. Chen, "Classifying the segmentation of customer value via RFM model and RS theory," *Expert Systems with Applications*, vol. 36, no. 3, pp. 4176-4184, 2009.

[32] P. S. Fader, B. G. Hardie, "Creating an RFM Summary Using Excel," http://www.brucehardie.com/notes/022/RFM_summary_in_Excel.pdf, 2008.

Ramón Alberto Carrasco

Ramón Alberto Carrasco has worked in the financial sector for over 24 years and has extensive experience acting as a leader of several departments related to Analytical Marketing and Business Intelligence. He completed his PhD in data mining at the University of Granada. His main field of research is focused on Data Science applied to Marketing (Data Mining, Multicriteria Decision-Making, Sentiment Analysis, Recommender Systems, etc) on which he has written several textbooks as well as around 60 research papers published in international journals and conferences specialized on these topics. He is currently involved in various (European, national, regional and local) Research, Development and Innovation (R&D&I) Projects. Dr Carrasco is a member of the SCI2S and SECABA Research Groups and is also serving as reviewer for several international journals and conferences. He has been an assistant professor at the University of Granada, Granada, Spain (2004-2013) and currently at the Complutense University, Madrid, Spain (from 2013).

María Francisca Blasco

Professor and Vice-Dean of the Faculty of Commerce and Tourism of Complutense University, Ph.D. in Comunication Sciences by Complutense University. She is chief editor of Journal Cuadernos de Estudios Empresariales; associate editor of International Academy of Management and Bussines Journal (IAMB); Journal and member of the scientific committee of Esic Market Journal. Her work focuses on several issues related to Data Sciences in Marketing and Tourism, Customer Behaviour, Customer Relationship Management, Category Management among others. Her research has been published in journals like Soft Computing, BRQ Business Research Quaterly, Procedia Computer Science, Tourism Review, Universia Business Review, Business Research Quaterly, etc. She works for some research competitive projects like RETO program (Horizon 2020). She is co-authored of books like Marketing del Turismo Cultural, Nuevas Tendencias en Comunicación Estratégica, Principios de Marketing (ESIC Ed.) and Diccionario de Administración de Empresas y Marketing (EcoBook).

Jesús García-Madariaga

Jesús García-Madariaga is currently Associate Professor in Management and Marketing at Complutense University and chair of the University Research team "Markco2". His present research interests encompass Tourism, Customer Relationship Management, Data Mining and Corporate Social Responsibility. His research has been published in refereed international journals, including Journal of Retailing and Consumer Services, Journal of Marketing Channels, Spanish Journal of Marketing ESIC, Journal of Database Marketing and Customer Strategy Management, Journal of International Technology an Information Management, Innovar Journal of Administrative and Social Sciences, Academy of Marketing Science News and Journal of Business Ethics among others. Jesús is member of the Doctoral Committee and Senior Lecturer in graduate courses and Master's programs at the Economics and Business School at Universidad Complutense de Madrid. He is Associate Editor at Associate Editor at International Journal of Internet Marketing and Advertising (IJIMA) and member of the Scientific Committee of some others.

Enrique Herrera-Viedma

Enrique Herrera-Viedma is Professor in Computer Science and A.I in University of Granada and currently the new Vice-President for Research and Knowlegde Transfer. His current research interests include group decision making, consensus models, linguistic preference modeling, aggregation of information, information retrieval, bibliometrics, digital libraries, web quality evaluation, recommender systems, social networks, and social media. In these topics he has published more than 200 papers in ISI journals and coordinated more than 20 research projects. Dr. Herrera-Viedma is member of the gobernment board of the IEEE SMC Society and an Associate Editor of several international journals such as the IEEE Trans. On Syst. Man, and Cyb.: Systems, Knowledge Based Systems, Soft Computing, Fuzzy Optimization and Decision Making, Applied Soft Computing, Journal of Intelligent and Fuzzy Systems, and Information Sciences.