

Segmentation comportementale à l'aide des réseaux communautaires

Gaël Bardury*, Teddy Boula**

*Laboratoire LAMIA, Université des Antilles

gbardury@gmail.com

**teddy.boula@gmail.com

Résumé. La mise en place d'actions marketing efficaces passe par la segmentation de la clientèle. C'est-à-dire que les clients sont regroupés en ensembles homogènes en fonction de leurs habitudes de consommation, ce qui rend possible les actions ciblées. Ces dernières, en personnalisant l'offre permettent d'obtenir des taux de transformation plus importants et de meilleures ventes.

Dans cet article, une méthode originale de segmentation comportementale de la clientèle est présentée. Elle permet de visualiser les segments de clients à travers des réseaux de communautés et de déceler aisément des mutations soudaines ou graduelles dans les comportements de quelques individus ou d'un ensemble plus important. L'analyste bénéficie alors d'une meilleure visibilité et peut adapter l'offre à tout moment.

1 Introduction

La segmentation est un élément clé en marketing, et une étape essentielle avant de pouvoir cibler la clientèle. Elle consiste à former des groupes homogènes à partir d'une population hétérogène, en se basant sur des critères comme le sexe, l'âge, la catégorie socio-professionnelle et le comportement. De sorte que la qualité d'une segmentation peut généralement se mesurer par le degré d'homogénéité dans un segment, et le degré d'hétérogénéité entre segments (Wedel et Kamakura, 2000).

Cet article se focalisera sur la segmentation comportementale en prenant en compte les habitudes de consommation d'un ensemble de clients de la grande distribution. Pour cela nous analysons les ventes d'une grande surface qui propose dans ses rayons en moyenne 20 000 produits par an, et qui peut identifier les transactions de près de 10 000 clients uniques grâce à la mise en place d'une carte de fidélité.

Les habitudes de consommation sont alors extraites en déterminant des connexions, c'est-à-dire, des groupes de produits que des clients vont chercher à acheter et parfois utiliser ensemble. Ainsi, en associant chaque connexion aux clients qui les réalisent, il est possible de regrouper les individus ayant des habitudes similaires dans un même segment comportemental. Un changement d'habitude de consommation est donc défini comme étant le déplacement d'un individu d'un groupe à un autre entre deux périodes, ce qui entraîne la modification de l'aspect de ces groupes.

De ce fait, la représentation et la visualisation des segments permet à l'analyste de détecter

aisément des changements de comportements chez les clients, d'appréhender immédiatement la relation entre les groupes, et d'identifier des phénomènes intéressants. Ce qui lui permet, en s'appuyant sur son expérience, son expertise ainsi que son intuition, de prendre des décisions, de mener des actions ciblées et de réagir rapidement en cas de mutation des comportements. Dans la section 2 de cet article, nous expliquerons brièvement comment les connexions et les réseaux de produits sont établis. Puis, dans la section 3, nous montrerons comment notre méthode associe et représente les clients en réseaux de communautés correspondant à des segments comportementaux. Ce sont ces segments que l'on utilisera alors afin de visualiser les habitudes de consommation, et de détecter les possibles changements de comportement entre deux périodes. Enfin, dans la conclusion nous ferons une synthèse de nos observations et nous présenterons les prolongements possibles de ces travaux.

2 Modélisation du panier de la ménagère

Afin de segmenter les clients en fonction de leur comportement d'achat, il est nécessaire de modéliser le panier de la ménagère de façon concrète et intuitive. Nous nous appuyons, dans cette étude sur la proposition de Bardury et Symphor (2013) qui modélise le panier de la ménagère par un réseau complexe où les articles du panier sont les noeuds du réseau. Les connexions sont, quant à elles, régies par un indice de similarité de Jaccard. Si l'indice dépasse un seuil fixé par l'utilisateur, un lien existera entre les deux produits testés.

Dans cette étude, nous utiliserons plutôt l'indice de Sorensen à cause de la cinétique des articles qui n'est pas assez prise en compte par Jaccard. En effet, il est entendu que les articles dans une enseigne ne s'achète pas sur le même rythme quand bien même leur forte complémentarité. Prenons l'exemple du pain et de la confiture. Le pain se consommant plus vite, il sera potentiellement présent plus fréquemment dans les transactions que la confiture, qui a une vitesse de consommation plus lente. Doit-on pour cela remettre en cause leur complémentarité en se basant sur leur indice de Jaccard faible. Pour palier ce défaut, dans ce contexte, il faut donner plus de poids à la présence simultanée des produits d'où le choix de Sorensen.

La modélisation du panier décrit, celui-ci, sur des périodes dont la longueur est définie par l'utilisateur. Elle peut être quotidienne, hebdomadaire ou encore mensuelle. L'utilisateur est à même de suivre l'évolution du panier à un rythme choisi. Nous nous appuyons sur des données produites selon la périodicité voulue. Les données transactionnelles correspondent aux transactions des clients dans l'enseigne avec une identification possible des dits clients.

Nous produirons donc une modélisation semaine après semaine en conservant les informations relatives aux clients ayant effectué les achats présents dans le panier.

La modélisation du panier de la ménagère ainsi faite avec des liens aussi robustes que pertinents, nous amène vers l'objectif principal de notre contribution. Grâce à ces liens, nous sommes en mesure de générer des profils de comportements clients. Dans les sections suivantes, nous allons nous employer à décrire la méthodologie utilisée afin de construire une segmentation clientèle basée sur notre réseau de produits.

3 Segmentation comportementale de la clientèle

D'après Rigby et Ledingham (2004), le processus de la relation client inclut nécessairement une segmentation préalable de la population. Elle permet notamment de faire du marketing direct et donc viser certains segments de clientèle avec des offres spécifiques. Les taux de réponse des offres faites en marketing direct sont bien meilleurs que lors d'un ciblage aléatoire. De nombreuses études portant sur la segmentation clientèle ont utilisé des attributs généraux comme donnée de base. La collecte des données pour ce genre d'attributs se fait couramment via des sondages auprès des clients. Les méthodes de segmentation utilisées sur les données collectées, sont généralement les réseaux de neurones ou les K-Means (Bloom, 2005), (Liao et al., 2011). Ce genre d'études ne reprend pas les comportements transactionnels réels des clients ou alors ils sont collectés via des questions du sondage. En tout état de cause, ce genre d'analyse présente un manque en matière de pertinence et surtout de cohérence des données en majorité incomplètes. Concernant la question du comportement, des articles ont utilisé les attributs de la RFM (Récence-Fréquence-Montant) couplés avec les méthodes de segmentation largement connues (Cheng et Chen, 2009; Chang et Tsai, 2011). Coussement et al. (2014) cherchent à évaluer la validité de la RFM selon les contextes d'utilisation. En effet, la RFM possède des attributs macroscopiques qui ne révèlent pas la réalité des goûts des clients, critères qui sont importants en marketing direct. La tendance dans ce domaine, s'est portée sur des segmentations basées sur des données transactionnelles. Les graphes bi-partis entre clients et produits sont une piste proposée pour répondre à cette problématique (Videla-Cavieres et Ríos, 2014), cependant se pose de nouveau la question de la lisibilité de la production. Dans cette section, nous allons montrer comment en partant d'un réseau de produits, nous parvenons à effectuer une segmentation comportementale des clients et à l'agréger de façon aisée.

3.1 Méthodologie de construction du réseau

La méthode de construction se rapproche de celle du réseau de produits. Ici, les sommets du graphe seront les clients. Nous rajoutons trois subtilités par rapport à celle ci.

- Premièrement, le jeu de données utilisé est différent. Ce jeu de données est issu de la modélisation du panier. Lors de la validation de la connexion entre deux produits, un processus de recensement des clients est effectué. Il est alors possible de transposer les données transactionnelles (*Client X Articles*) vers des données du type du tableau 1. La connexion entre deux produits est conservée grâce à un code unique e_1 qui sera

| Connexion | Période | ID Clients |
|-----------|---------|--------------------|
| e_1 | t | c_1, c_2, c_{10} |

TAB. 1 – Identifiants clients de la connexion e_1 pour la période t

réutilisé si cette connexion est validée sur une autre période. Les identifiants des clients sont aussi conservés pour la connexion e_1 pour la période t ce qui nous permet de suivre cette connexion dans le temps, et aussi les clients l'ayant réalisée. Ceci s'avérera utile pour la détection des changements de comportements.

- La période d'analyse cyclique pour le suivi du panier de la ménagère peut être définie par l'utilisateur comme nous l'avons déjà évoquée. Nous préconisons une période heb-

domadaire car un constat empirique montre que c'est le rythme moyen d'achats de la majorité des clients. Cette information, nous amène au deuxième point. Pour construire ce réseau, nous nous basons sur des regroupements de quatre semaines consécutives du réseau de produits. En effet, généralement, les segmentations clientèles s'appuient sur des périodes plus longues qu'une semaine et tiennent compte de plus de données afin de bâtir un modèle plus robuste et plus complet. De plus, cette période plus longue nous permet d'intégrer plus de clients lors du traitement.

- Enfin troisièmement, les connexions sont toujours générées en fonction d'un indice de similarité. Néanmoins, pour ce type de graphe, nous utiliserons l'indice de Jaccard. En effet, la fréquence n'étant pas prise en compte, nous n'avons plus besoin d'utiliser l'indice de Sorensen, c'est à dire de valoriser la cooccurrence. Il faut retenir que deux clients connectés dans le réseau, font sensiblement les mêmes connexions de produits.

3.2 Communautés de clients

Les connexions entre clients ainsi faites, nous pouvons générer un réseau de clients. L'idée est maintenant de pouvoir effectuer une segmentation des comportements. Les clusters non connexes d'articles nous révèlent qu'il y a des groupes de clients qui ont un fonctionnement non similaire aux autres clients. De plus, il serait intéressant de savoir si à l'intérieur des clusters, les comportements sont homogènes. Nous allons pour cela tenter de détecter des structures communautaires au sein du réseau de clients. Basiquement, les structures communautaires se trouvent en maximisant le nombre de liens à l'intérieur d'un groupe de sommets et en minimisant le nombre de liens entre les groupes. Pour partitionner notre réseau, nous utiliserons la méthode de Louvain (Blondel et al., 2008)

La résultante de la détection de communautés conduit à une forte densité de liens intracommunautaires et une faible densité de liens inter-communautaires. Nous pouvons établir que, dans un même cluster, si il y a plusieurs communautés connexes, leur ressemblance sera faible mais existera quand même à cause des liens inter-communautaires. Pour analyser de façon aisée les connexités entre communautés, il conviendra de regrouper tous les clients des groupes, dans un noeud pour une vue macroscopique.

3.3 Cas pratique

La figure 1 nous montre un exemple de réseau, qualifiée avec la structure communautaire, de clients créé sur des données de ventes réelles. A première vue sur ce réseau, nous constatons qu'il y a beaucoup de parties non connexes. Il y a un cluster de clients au centre qui regroupe 75% de ceux-ci. De façon globale, sur les 8200 clients qui ont eus des transactions sur les 4 semaines d'analyse, seuls 5300 apparaissent dans le réseau. Cela est dû au seuil de similarité très élevé que nous avons fixé. Les clients n'ayant pas eu de comportements d'achat similaires à au moins un autre client n'ont pas été retenus, de ce fait, ils n'apparaissent pas dans le réseau.

Nous allons, dans ce cas pratique, nous intéresser à la communauté de couleur violette. Celle-ci après analyse correspond à un groupe de clients ayant acheté des produits exclusivement de marque distributeur. Ces produits sont connus pour être moins cher que les produits de grande marque de qualité équivalente. D'un point de vue commercial, ces produits sont aussi générateurs de marge importante pour le distributeur. Afin de développer ce type de clientèle

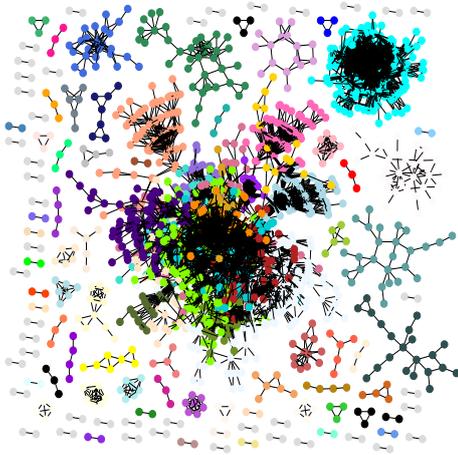


FIG. 1 – Exemple de réseau de clients

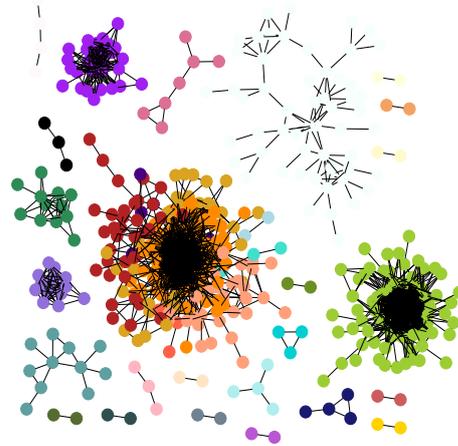


FIG. 2 – Exemple de réseau de clients

le distributeur pourrait cibler à l'aide une campagne marketing des clients de communautés connexes pour les faire migrer vers la communauté violette .

La communauté turquoise en haut à droite correspond à des restaurateurs venus acheter des boissons en masse pour leur restaurant. Ce comportement est lié à un profil particulier ce qui explique l'isolation de ce groupe.

3.4 Détection de changements

En analysant le réseau de clients et la structure communautaire de façon suivie et périodique, il est possible de voir des changements de comportements. De façon globale, nous pouvons analyser la structure du réseau et sa connectivité. Le nombre de composantes connexes peut augmenter ou diminuer. Si le nombre de composantes diminuent et se regroupent, on peut imaginer que le comportement des individus s'homogénéise. Entre la figure 1 et la figure 2 qui représentent respectivement les périodes $t - 1$ et t , on remarque que le nombre de clients a augmenté entre ces deux périodes, cela signifie, qu'en t plus de clients ont des comportements similaires. D'ailleurs, il est visible que la taille et la densité des réseaux ont augmenté. Parallèlement, il y a eu un développement du nombre de petites composantes indépendantes .

4 Conclusion

Nous avons montré dans cet article, comment à partir de données de transactions usuelles d'achat de clients, nous parvenons à proposer une méthode de segmentation comportementale basée sur la détection de communauté. En cherchant à découvrir des structures communautaires au sein de ce réseau, nous obtenons des segments comportementaux qui sont essentiels à toutes actions marketing personnalisées. De plus, la visualisation et la détection de changements dans les comportements permettent à l'analyste de connaître à tout moment le marché et d'être réactif. Il est intéressant de noter que toute donnée transactionnelle identifiée peut être

le sujet de notre méthode peu importe le contexte. Pour aller plus loin, nous envisageons de fouiller la question de la modélisation du panier du consommateur en rajoutant par exemple, des attributs informatifs. Ces attributs pourraient être la notion de promotion ou encore des identifiants intrinsèques à l'article comme sa nomenclature ou son type. Concernant le réseau de clients, des attributs intrinsèques à ceux-ci peuvent être aussi intéressants à incorporer.

Références

- Bardury, G. et J. Symphor (2013). Détection précoce de tendances produits dans le cadre des activités commerciales de la grande distribution. In *Extraction et gestion des connaissances (EGC'2013), Actes, 29 janvier - 01 février 2013, Toulouse, France*, pp. 235–240.
- Blondel, V., J. Guillaume, R. Lambiotte, et E. Lefebvre (2008). Fast unfolding of communities in large network. *Journal of Statistical Mechanics : Theory and Experiment 10008*, 1–12.
- Bloom, J. Z. (2005). MARKET SEGMENTATION : A Neural Network Application. *Annals of Tourism Research 32*(1), 93–111.
- Chang, H.-C. et H.-P. Tsai (2011). Group {RFM} analysis as a novel framework to discover better customer consumption behavior. *Expert Systems with Applications 38*(12), 14499 – 14513.
- Cheng, C.-H. et Y.-S. Chen (2009). Classifying the segmentation of customer value via {RFM} model and {RS} theory. *Expert Systems with Applications 36*(3, Part 1), 4176 – 4184.
- Coussement, K., F. A. V. den Bossche, et K. W. D. Bock (2014). Data accuracy's impact on segmentation performance : Benchmarking {RFM} analysis, logistic regression, and decision trees. *Journal of Business Research 67*(1), 2751 – 2758.
- Liao, S.-H., Y. ju Chen, et H. hua Hsieh (2011). Mining customer knowledge for direct selling and marketing. *Expert Systems with Applications 38*(5), 6059 – 6069.
- Rigby, D. K. et D. Ledingham (2004). Crm done right. *Harvard Business Review November 2004*.
- Videla-Cavieres, I. F. et S. A. Ríos (2014). Extending market basket analysis with graph mining techniques : A real case. *Expert Systems with Applications 41*(4, Part 2), 1928 – 1936.
- Wedel, M. et W. Kamakura (2000). *Market Segmentation : Conceptual and Methodological Foundations (International Series in Quantitative Marketing)*. Kluwer Academic Publishers.

Summary

The customer segmentation is needed to perform an effective marketing action. Customers are then grouped into homogeneous groups according to their consumption patterns, which makes targeted actions possible. Those actions, by customizing the offer for the customers, result in a greater conversion rate and better sales. In this paper we present an original method of customer segmentation, which allows to visualize customer segments through community networks and to easily detect sudden or gradual changes in the behavior of a few or a greater number of individuals. Therefore, the analyst has a better visibility on the market and can adapt the offers at any time.