



**HAL**  
open science

## Pourquoi le client doit il être parfait ?

Arnaud Martin

► **To cite this version:**

Arnaud Martin. Pourquoi le client doit il être parfait ?. Survey Magazine, 2019, pp.18-19. hal-02409321

**HAL Id: hal-02409321**

**<https://hal.science/hal-02409321v1>**

Submitted on 13 Dec 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## **Pourquoi le client doit il être parfait ?**

**Arnaud Martin, Univ Rennes, CNRS, IRISA, DRUID**

Nous sommes sans cesse sollicités pour répondre à des enquêtes de satisfaction. Si ces enquêtes peuvent être utiles aux entreprises, pouvons nous répondre sincèrement et comment les entreprises peuvent elles réellement utiliser nos données ?

### **Quelles données clientèles ?**

Les données des clients ou futures clients ont toujours intéressés les entreprises pour mieux connaître et ainsi cibler leur offre. Ces dernières années les évolutions technologiques ont permis de stocker de très grandes quantités de données. Le problème ne réside donc plus dans le stockage des données mais dans le traitement de ces données.

Depuis les années 60, les méthodes d'analyse de données permettent d'extraire des informations pertinentes sur des données issues d'enquêtes par exemple. Cependant, ces approches étaient essentiellement adaptées à des données quantitatives ou bien qualitatives ordinales ou nominales, mais toujours avec un nombre réduit de catégories. Si ces méthodes de traitement d'enquêtes ont évolué, notamment pour traiter de grandes quantités de données, elles ne permettent pas encore de bien tenir compte des questions ouvertes. En général, les questions ouvertes ne servent qu'à satisfaire le client et ne sont que peu traitées. Des méthodes de traitement automatique de la langue naturelle permettent tout de même d'extraire certaines informations, essentiellement à partir de mots clés à définir au préalable.

Les approches d'analyse de données, qu'on appelle aussi maintenant fouille de données, peuvent être non supervisées (c'est-à-dire sans aucun a priori sur ce qui est à découvrir dans les données) ou supervisées (c'est-à-dire certains profils sont attendus, nommés les classes). Ces dernières années, avec la résurgence des approches des réseaux de neurones des années 80, aujourd'hui appelées deep learning, les méthodes supervisées ont montré leur efficacité lorsqu'une très grande quantité de données est disponible pour l'apprentissage de ces approches.

Si l'évolution des technologies pour faire parler les données est constante, il ne sera possible d'en tirer des conclusions que si les données de base sont bien représentatives de la réalité. Qui peut dire que son opinion ou ses réponses sur une enquête de satisfaction est d'une part immuable mais surtout certaine et précise ? Comme nous l'avons vu précédemment, les méthodes de traitement ne sont pas adaptées à des questions ouvertes. Par exemple, si on vous demande le nombre de paquets de céréales que vous achetez par semaine, le questionnaire peut prévoir que vous n'achetez jamais de céréales, 1 paquet, 2 paquets, 3 paquets, ou plus. Une réponse ouverte est ainsi évitée, mais comment préciser, sur ce type de question, que certaines semaines vous n'en achetez pas et d'autres vous avez une consommation plus importante et vous achetez 2 paquets ? Comment préciser que vous n'êtes pas sûr de la réponse que vous donnez, par exemple parce que vous n'allez pas toujours faire les courses, voire même comment dire que vous ne savez pas répondre à cette question ?

Derrière ces questions, trois notions importantes apparaissent : l'imprécision, l'incertitude et l'ignorance. Forcer une personne à répondre à une question sans lui permettre d'exprimer son imperfection dans sa réponse ne peut entraîner que des erreurs dans le traitement de ces données.

## **Comment modéliser les imperfections des clients ?**

Nous avons vu qu'il est important de pouvoir laisser le client s'exprimer au mieux et le plus complètement sur les questions d'une enquête et pas seulement sur la question ouverte. Pour se faire, les questions doivent être conçues de façon explicite et compréhensible pour les clients. Il s'agit ici d'adapter les questions en fonction de la thématique abordée et de la cible des répondants. Nous ne détaillerons donc pas davantage cet aspect dans cet article qui dépend directement du problème à traiter.

Une fois les données recueillies, il faut les modéliser finement afin d'avoir la meilleure représentation possible des réponses des clients. Il est important de pouvoir modéliser à la fois le caractère incertain et le caractère imprécis des réponses. La théorie des probabilités permet de bien modéliser les incertitudes, la probabilité étant une mesure d'incertitude. La théorie des sous-ensemble flous permet de modéliser les imprécisions, via une représentation imprécise des connaissances. La théorie des fonctions de croyance permet de modéliser à la fois le caractère incertain d'une information mais aussi son caractère imprécis.

Le principe de la théorie des fonctions de croyance repose sur la croyance élémentaire que le client peut porter sur une assertion. Comme pour une probabilité, cette croyance élémentaire peut s'exprimer par une valeur entre 0 et 1 qui peut être traduite sur un questionnaire par une échelle ordonnée ou un curseur pour les questionnaires en ligne. A la différence des probabilités, la normalisation ne se fait pas sur l'ensemble possible des réponses (le cadre de discernement), mais sur toutes les disjonctions du cadre de discernement. Ainsi une croyance élémentaire de 1 sur un élément représente une certitude ; une croyance élémentaire de 1 sur un ensemble d'éléments représente une imprécision ; une croyance élémentaire de 1 sur l'ensemble de tous les éléments représente l'ignorance.

Si nous reprenons l'exemple de la question sur le nombre de paquets de céréales acheté par semaine, il est possible de dire que vous achetez 1 ou 2 paquets avec une certitude de 0.8. De plus, ceci induit que vous n'êtes pas sûrs avec un niveau de 0.2 que vous avez répondu correctement à la question.

Certaines études ont déjà montré l'intérêt de ce type de réponse pour les personnes interrogées car ces réponses permettent de s'exprimer d'une façon plus proche d'un langage naturel tout en étant suffisamment fermées. Il est alors nécessaire de définir des approches pour extraire l'information contenue dans ce type de réponse.

## **Comment traiter les imperfections des clients ?**

La modélisation fine des imperfections au travers la théorie des fonctions de croyance va permettre de développer des méthodes tenant compte de ces imperfections. En effet, mieux vaut une information imparfaite mais fournie honnêtement qu'une information erronée.

Ainsi, dans le cadre de cette théorie, plusieurs approches de classification supervisée et non supervisée ont été développées. Ces méthodes permettent de caractériser les données avec ou sans a priori, c'est-à-dire apprentissage. Sans entrer dans les détails techniques de ces approches, il est important de préciser qu'elle permettent de prendre des décisions comme les approches classiques mais elles peuvent également proposer des décisions imprécises et fournir des indicateurs sur la certitudes des décisions à prendre.

Un autre aspect qu'il peut être intéressant d'évoquer est la capacité de ces approches à fusionner les informations issues de plusieurs personnes afin de tirer des conclusions. Cet aspect peut être particulièrement pertinent dans le cas d'enquête d'opinions, pour les élections par exemple. Demandons par exemple leur avis à des personnes sur leur candidat préféré à la prochaine élection. Elles peuvent répondre de façon incertaine et imprécise par exemple en disant qu'elles voteront plutôt pour le candidat « Un Tel ». Si beaucoup de personnes pensent voter pour le candidat « Un Tel », sur la base de leurs opinions diverses et indépendantes, on peut penser qu'il y a plus de chance que le candidat « Un Tel » soit élu, même si les personnes initialement étaient incertaines. Ainsi, plusieurs stratégies de combinaison de ces informations dans le cadre de la théorie des fonctions de croyance ont été proposées. Au cœur de ces approches, réside la notion de conflit car lorsqu'on laisse la possibilité à plusieurs personnes indépendantes de s'exprimer, inévitablement certaines opinions seront contradictoires. Le conflit peut donc être vu comme une conséquence inhérente à la diversité des réponses et donc des opinions. Atteindre un consensus est l'objectif de la fusion d'informations. C'est en effet à partir d'un consensus qu'une décision peut être prise que ce soit lors d'une assemblée de personnes ou par un algorithme. Si le consensus est mou, la décision sera moins fiable et parfois peu pertinente.

Il paraît donc essentiel de proposer des questions menant à des réponses aussi naturelles que possible pour les enquêtes de satisfaction. Laisser la possibilité aux répondants d'être incertain et imprécis dans leur réponse permet d'avoir des réponses plus justes et donc plus exploitables. Les outils de mobilisation et de traitement de ces informations existent, il reste à les utiliser.