## Introduction

Le MBTI est un questionnaire qui permet de classifier des individus selon quatre dimensions binaires. Chaque dimension représente deux préférences opposées tel qu’un individu n’en possède qu’une seule en même temps. Le MBTI contient plus d’une centaine de questions. Le but de ce projet est de proposer une solution en utilisant des réseaux bayésiens pour minimiser le nombre de questions posées.

## Réseaux Bayésiens

Un réseau bayésien modélise les relations entre les différentes variables d’un système. Ces variables peuvent avoir plusieurs états qui apparaissent plus ou moins souvent que d’autre. Un réseau bayésien est un graph acyclique où chaque nœud représente une variable et chaque arc représente la dépendance conditionnelle entre deux variables. Les réseaux bayésiens permettent donc de modeler les relations conditionnelles entre différentes variables d’un système. Un réseau bayésien aléatoire (RBA) est un ensemble de réseaux bayésiens qui sont construits aléatoirement. Les RBA sont utilisé pour avoir des résultats plus uniformes lors de l’utilisation de réseaux bayésiens. En effet, les réseaux bayésiens sont des modèles biaisé qui ne représente que certain aspect d’un système. Un RBA contient plusieurs variations de réseaux bayésiens et représente donc mieux le modèle original.

Dans mon approche, la création d’un réseau bayésien est définit par 5 paramètres :

|  |  |
| --- | --- |
| Nom | Description |
| Dimensions | Nombre d’attributs dans le modèle |
| Minimum par niveau | Le nombre minimum de nœuds par niveau |
| Maximum par niveau | Le nombre maximum de nœuds par niveau |
| Minimum de parents | Le nombre minimum de parents par nœud |
| Maximum de parents | Le nombre maximum de parents par nœud |

L’algorithme consiste à créé le réseau niveau par niveau. Chaque niveau contient un ensemble de variables choisis où chacune est représentée par un seul nœud. L’ordonnément des niveaux et la sélection de leurs nœuds est complétement aléatoire. Dans le cadre de ce projet, chaque question du MBTI est une variable aléatoire. La création du réseau est aléatoire à cause du manque de connaissances concrètes sur les différentes variables. En effet, dans le MBTI il est difficile de dire quelles questions ont une influence sur d’autres questions. Vu que la création du réseau est aléatoire, il est facile de généré un RBA; celui-ci est un ensemble de réseaux bayésiens aléatoires.

Le RBA est utilisé pour minimiser la quantité de questions à répondre sur le MBTI. En effet, lorsqu’un utilisateur a déjà répondu à quelques questions, le RBA peut deviner les réponses à d’autres questions et, dans certains cas, peut même trouver directement le résultat du MBTI. Dans la prochaine section, je présente les résultats liés à cette tâche. Soit X l’ensemble de questions répondu par un utilisateur et soit Y l’ensemble de questions non répondu, incluant le résultat final. En utilisant un réseau bayésiens il est possible de trouver une ou plusieurs questions parmi Y qui ont une très forte probabilité d’avoir une valeur donnée. En répétant ce processus récursivement, il est possible de compléter un sous-ensemble de Y. Toute question devinée n’a pas besoin d’être répondu par l’utilisateur.

Il est aussi possible de minimiser le nombre de question à répondre avant même que l’utilisateur commence à remplir le questionnaire. Pour ce faire, il est possible d’analyser le ou les réseaux bayésiens pour identifier quelle variables n’ont aucun impact sur le résultat du test ou celle qui sont prédictible à l’aide d’autres variables. Cette tâche de minimisation n’est pas forcément facile ni approprié pour les réseaux bayésiens à cause de la représentation des données et n’ai donc pas abordée dans ce projet.

## Résultats

## Références