## 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.

La création d’un réseau bayésien n’est qu’une partie du problème posé pour ce projet, l’utilisation du réseau s’avère être une tâche complexe avec une multitude de possibilité. Dans mon cas, 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 deux expériences liées à cette tâche.

## Experiences

Dans cette section, je présente deux expériences qui permettent d’évaluer les performances de mon RBA. Le but ultime de ce projet et de présenter une solution au problème de minimisation de questions, mais l’évaluation de cette solution doit être faite avec une tache différente. En effet, un RBA peut être utilisées de maintes façons et les deux expériences présentées utilisent le même processus de construction pour le RBA mais divergent dans leurs utilisations.

La première expérience consiste à évaluer les performances du réseau bayésien à compléter les questions non répondu du MBTI. Les données sont divisées aléatoirement en deux sous-ensembles; les données d’entrainement et celle de teste. Le RBA est construit avec les données d’entrainement. Chaque vecteur de réponse du sous-ensemble de teste est aléatoirement modifié; une portion du vecteur est enlevée. Dans cette expérience, le RBA est utilisé pour deviner la valeur originale des données enlevées. Les résultats suivant sont présenté en termes de précision, de couverture et du nombre moyen de question bien deviné par individu.

La seconde expérience consiste à calculer le nombre minimal de questions nécessaire pour trouver le résultat du MBTI pour un individu. Comme pour l’expérience précédente, les données sont divisées aléatoirement et le RBA est construit de la même façon. Pour chaque vecteur X dans les données d’entrainements, je calcule Prob(X|Ci) où Ci représente la

## Références