

Introduction à la Transparence des Algorithmes pour la Décision

Brice Mayag

Chapter 1

Plan

- 1 Models
- 2 Decision theory and Decision analysis
- 3 Main steps of developing a decision model
- 4 Algorithms & Transparency
- 5 Our program

Current definitions of a model

- a standard or example for imitation or comparison.
- a representation, generally in miniature, to show the construction or appearance of something.
- an image in clay, wax, or the like, to be reproduced in more durable material.
- a person or thing that serves as a subject for an artist, sculptor, writer, etc.
- a person whose profession is posing for artists or photographers.
- a person employed to wear clothing or pose with a product for purposes of display and advertising.
- a style or design of a particular product: His car is last year's model.
- a pattern or mode of structure or formation.
- a typical form or style.
- a simplified representation of a system or phenomenon, as in the sciences or economics, with any hypotheses required to describe the system or explain the phenomenon, often mathematically.
- Zoology: an animal that is mimicked in form or color by another.

What is a model?

- \Rightarrow Representation of reality
- **More Precisely:** A model refers to some form of symbolic representation of our assumptions about reality

Why do we use models?

- Enhance our understanding of the world to improve our decision making
- Elaborate a scientific method to solve a problem
 - Duplicable (repeatable)
 - Reduce bias

Types of models

① Deterministic models

- outcomes are precisely determined through known relationships among states and events
- in such models, a given input will always produce the same output
Ex: Resources to make a PC are the same every time
- Domains: Multi-Attribute Decision Making; Linear programming; ...

② Probabilistic (stochastic) models

- Not all data is known with certainty
- Ex: College acceptance, being above average increases likelihood of acceptance but does not make it certain
- Domains: Queuing; Simulation; ...

Models are fed by data

① Qualitative data

- measured by quality
- Expert opinions
- Ex: class atmosphere, ...

② Quantitative data

- Easily measured by numbers
- Ex: Numbers of tv programs a day; number of applications in a phone; ...

Models are used every day

- A Scenario: Driving to school

- At what time do you need to leave home to be at school on time?
- $\text{Distance} = \text{Rate} \times \text{Time}$
- $\text{Time} = \text{Distance}/\text{Rate}$

Formal models vs Informal models

- A formal model is a precise statement of components to be used and the relationships among them.
- Formal models are **usually stated via mathematics, often equations.**
- Formal models can be precisely communicated because they are well-defined.
- Formal models give replicable results. This is the simple meaning of “mathematical proof”.

Formal models vs Informal models

- Formal models are not reality: you must choose the model.
- Formal models may not correspond to reality: the prediction will turn out to be false.
- An informal model is one in which the symbols are mental, verbal, or pictorial: e.g. we toss a coin, we ask an oracle, we visit an astrologer, we consult an expert, we think
- Informal models simply have some lack of precision. Some relationships may not be stated as equations

Plan

- 1 Models
- 2 Decision theory and Decision analysis
- 3 Main steps of developing a decision model
- 4 Algorithms & Transparency
- 5 Our program

A definition of Decision

- The act or process of deciding; determination, as of a question or doubt, by making a judgment:

Ex: They must make a decision between these two contestants.

- The act of or need for making up one's mind:

This is a difficult decision.

- Something that is decided; resolution:

Ex: She made a poor decision when she dropped out of school.

- A judgment, as one formally pronounced by a court:

Ex: It is the decision of this court that the appeal is granted.

- The quality of being decided; firmness:

Ex: He spoke with decision and calm authority.

Source: <http://www.dictionary.com/browse/decision>

Provisional definition of Decision [RONALD HOWARD]

“Decision-making is what you do when you do not know what to do”



A definition of Decision in our context

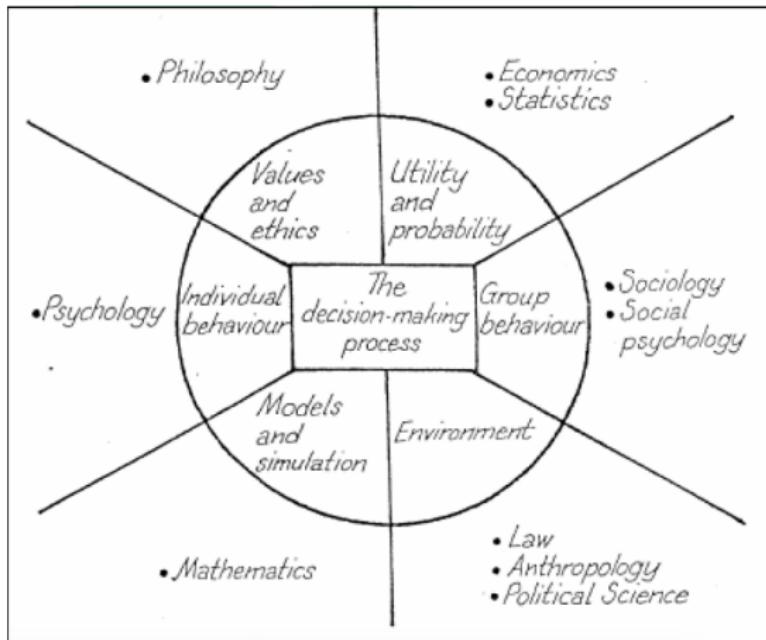
- A choice that you make about something after thinking about several possibilities
 - Ex: We need to take a lot of factors into account in our decision-making.
 - Ex: She has had to make some very difficult decisions.
 - Ex: The company will reach/come to/make a decision shortly.

Source:

<https://dictionary.cambridge.org/dictionary/english/decision>

Decision in many domains

Philosophy, Economics, Mathematics, Operational Research, Psychology, Computer sciences, Political sciences, Biology? Theology?



What Decision Analysis is not !

- A general method for taking “good decisions”
 - Example 1: Choice of new job
 - Example 2: medical decision
 - Etc.
- What is a “good decision”?
 - Good for whom, according to what criteria, at which moment in time?
- Good decision processes vs. good decisions?
- A description on how “wise people” decide
 - Expert systems
 - Doctors / Politicians: Nuclear Industry vs Road safety; Prevention vs First Aid
- How do you recognize “wise people”?
 - Luck vs. Wisdom

Decision Analysis

- **Definition (B. Roy):** “consists in trying to provide *answers* to questions raised by actors involved in a *decision process* using a *model*”
- **Decision process:** strategy of intervention: aid, communication, justification, etc.
 - Many different ways to provide decision-aid
 - Difficulty to compare methods
 - What is a “good” Decision Analysis model ?
 - Different models may lead to different recommendations

Decision Analysis

- **Definition (B. Roy):** “consists in trying to provide *answers* to questions raised by actors involved in a *decision process* using a *model*”
- **Answers:** “Optimal solution” or “Good decision” is absent
- **Models:** formalized or not

Decision Making

- Intervention in a decision process:

- imagine compromises
- communicate
- coordinate
- control
- motivate
- conduct change

- Importance of “final choice” ?

Formal decision models

- A set of explicit and well-defined rules to collect, assess and process information in order to be able to make recommendations in decision and/or evaluation processes
- A perfect or not even a best formal decision model do not exist.
 - It is important to describe the decision model used (transparency?).
- Actually, defining a “perfect model” would be a difficult, if not impossible, task.

Examples of models

- **Astrology**

the astrologer “provide *answers* to questions raised by his/her client using a *model*”

- **Graphology**

- **Psycho-analysis**

- **Decision analysis makes use of explicit and formalized models**

Formalized models

- Drawbacks: complex, opaque
- Advantages:
 - Provide a clear language: communication tool
 - Capture the essence of a situation: structuration tool
 - Answers “what-if” questions (sensitivity, robustness): Exploration tool
- Example: choosing a bottle of wine

Possible objections

- I do not need such tools because I know how to decide
- Let's organize a high-level meeting to discuss it
- Intuition is often enough

Plan

- 1 Models
- 2 Decision theory and Decision analysis
- 3 Main steps of developing a decision model
- 4 Algorithms & Transparency
- 5 Our program

Developing a decision model (Step 1)

Formulation: Translate the problem scenario into a mathematical model

- Define the problem
- Develop a decision model
 - **Variables:** Measurable quantity that can be variable
 - **Parameters:** measurable quantity inherent to problem

Developing a decision model (Step 2)

Solution: Mathematical expressions from formulation are solved

- **Develop a Solution:** Manipulate model to arrive at the best solution. Ex:
 $\text{Time} = \text{Distance}/\text{Rate}$
- **Test Solution:** Does the solution make sense?

Developing a decision model (Step 3)

Interpretation: Implication of results

- Conduct sensitivity analysis:
 - what happens if parameters vary?
 - Testing outcomes under a variety of states
- Implement results: Enact solution & monitor it performs as expected

Possible problems

Possible problems

- **Defining the problem:** Conflicting viewpoints, impact on other stakeholders
- **Model development:** Fitting problem scenario to textbook model, understanding of others
- **Acquiring data:** Existence, validity
- **Developing a solution:** Limitations of one answer
- **Implementation:** Management and user support

Plan

- 1 Models
- 2 Decision theory and Decision analysis
- 3 Main steps of developing a decision model
- 4 Algorithms & Transparency
- 5 Our program

Un algorithme : c'est quoi ?

Définition (informelle)

Un algorithme est la composition d'un ensemble fini d'étapes, chaque étape étant formée d'un nombre fini d'opérations dont chacune est :

- définie de façon rigoureuse et non ambiguë;
- effective (i.e. pouvant être réalisée en un temps fini).

La notion d'algorithme est plus générale que celle de programme (indépendant du langage de programmation utilisé).

Un peu d'histoire

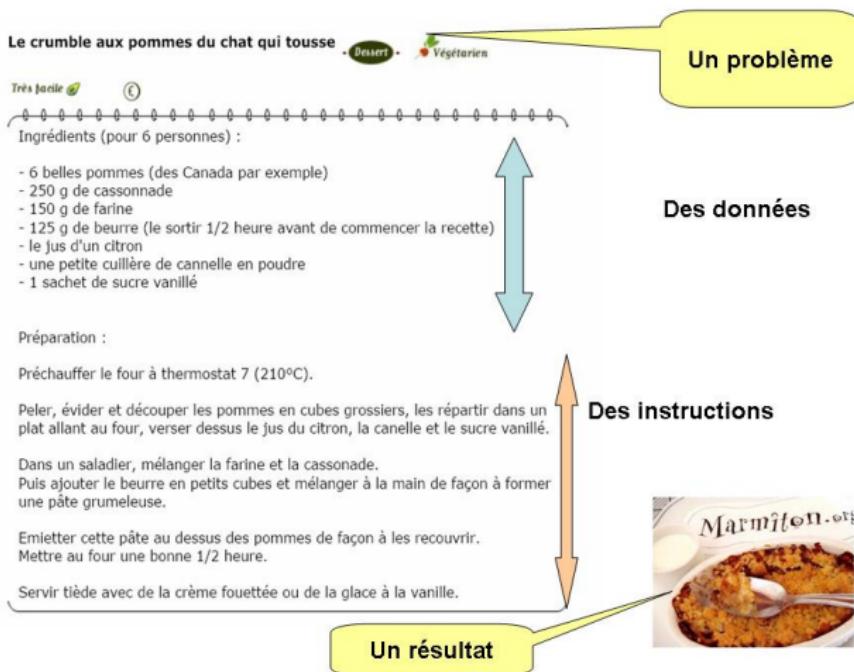
Le mot algorithme vient du nom du mathématicien Al Khwarizmi (820) : introduction de la numérotation décimale et des calculs s'y rapportant.

Des besoins contradictoires

Un algorithme doit :

- être simple à comprendre, à mettre en oeuvre et à mettre au point;
- mettre intelligemment à contribution les ressources de l'ordinateur, et plus précisément, il doit s'exécuter rapidement.

Un algorithme : c'est quoi ?



Classiquement assimilable à une recette de cuisine

Algorithme : définitions

Un algorithme =

- Description précise des opérations à faire pour résoudre un problème (suite d'instructions).
- Procédure de calcul bien définie qui prend en entrée une valeur, ou un ensemble de valeurs et qui donne en sortie une valeur, ou un ensemble de valeurs.

Un *bon* algorithme =

- Un algorithme **correct** : i.e. pour chaque instance en entrée, l'algorithme se termine en produisant la bonne sortie
⇒ Savoir prouver un algorithme
- Un algorithme **efficace** : mesure de la durée que met un algorithme pour produire un résultat
⇒ Savoir analyser la complexité d'un algorithme : i.e. détermination de l'espace mémoire et du temps d'exécution nécessaire à la résolution du problème.

Pour un problème donné, **plusieurs** algorithmes ou **aucun** sont possibles.

Un algorithme se termine en un **temps fini**.

Algorithme vs Programme

Deux termes souvent confondus

- Un algorithme est une entité mathématique décrivant une procédure;
- Le programme est une entité technique, qui implémente un algorithme dans un langage de programmation donné.

Transparence d'un algorithme

Tentative de définition

- Souvent comprise comme la publication du texte d'un algorithme ou du code source d'un logiciel
⇒ ces informations peuvent demeurer opaques pour le commun des citoyens, et même dans certains cas pour des experts.
- Le fonctionnement de certains types d'algorithmes, qui reposent sur des techniques d'I.A., ne peut être appréhendé indépendamment des jeux de données d'apprentissage.
⇒ ces données peuvent refléter des biais qui seront “apris” puis reproduits par l'algorithme.

Transparence d'un algorithme

Tentative de définition à travers cinq propriétés

- ① La loyauté
- ② L'équité
- ③ La Responsabilité ou Redevabilité (accountability):
- ④ L'explicabilité
- ⑤ L'intelligibilité

Les propriétés 1 et 2 sont des propriétés normatives extrinsèques et les deux dernières sont considérées comme des propriétés épistémiques intrinsèques.

Transparence d'un algorithme

La loyauté des algorithmes

- Un algorithme est **loyal** si la fonctionnalité affichée auprès de l'utilisateur est identique à la fonctionnalité connue du fournisseur (entreprise, concepteur, association, ...).
- Si le fournisseur dissimule une fonctionnalité dont il a une claire conscience qu'elle est remplie par l'algorithme, alors l'algorithme sera **déloyal**.

Exemples: tarification volatile, personnalisation de l'offre commerciale réalisée selon des critères différents des critères affichés, recommandation de trajets basés sur la présence de points d'intérêts commerciaux, réponse à des requêtes sur un moteur de recherche donnant la priorité à des liens sponsorisés plutôt qu'à la pertinence ...

Transparence d'un algorithme

La loyauté des algorithmes

- Un algorithme demeure **déloyal**, même s'il remplit les fonctionnalités affichées, s'il remplit en plus une autre fonctionnalité dont l'utilisateur n'a pas été clairement informé, comme par exemple la collecte de géolocalisation de l'utilisateur d'un smartphone à des fins commerciales.
- C'est la relation entre le fournisseur et l'utilisateur de l'algorithme, ou le fournisseur lui-même, qu'on peut en toute rigueur qualifier de loyal ou déloyal.

Transparence d'un algorithme

L'équité des algorithmes

- Un algorithme est **équitable** si les résultats qu'il produit n'induisent pas un effet discriminant ou **biais** à l'égard d'une catégorie particulière de la population
- La discrimination peut être introduite, de manière consciente ou inconsciente, prévisible ou imprévisible, au niveau de la conception de l'algorithme.
- **Exemple:** les développeurs d'une application de la ville de Boston pour signaler les dégâts sur la voie routière ont dû corriger un biais favorisant les quartiers les plus fortunés : la collecte des données était faussée par le fait que la probabilité de posséder un smartphone et de télécharger l'application était bien plus grande dans les populations les plus aisées.

Transparence d'un algorithme

L'équité des algorithmes

- L'inéquité doit être attribuée aux effets sociaux de l'algorithme.
- Ces effets sociaux dépendent de caractéristiques techniques de l'algorithme, mais ne s'y réduisent pas a priori : ils dépendent de l'interaction entre l'algorithme et son contexte social d'usage.

Transparence d'un algorithme

La responsabilité des algorithmes

- les concepteurs de systèmes algorithmiques ou ceux qui les utilisent doivent être en mesure de "**rendre des comptes**", c'est à dire de justifier les résultats de ces systèmes.

Loi numérique 2017: transparence des algorithmes décisionnels à caractère individuel, utilisés par la fonction publique

L'administration est tenue de fournir (sur demande), "sous une forme intelligible et sous réserve de ne pas porter atteinte à des secrets protégés par la loi", les informations suivantes :

- Le degré et le mode de contribution du traitement algorithmique à la prise de décision
- Les données traitées et leurs sources
- Les paramètres de traitement et, le cas échéant, leur pondération, appliqués à la situation de l'intéressé
- Les opérations effectuées par le traitement

Transparence d'un algorithme

L'explicabilité des algorithmes

- Un algorithme est **explicable** s'il est possible de donner à l'ensemble des utilisateurs, **quelque soit leur bagage éducatif** (sans expertise scientifique), une vision claire des procédures employées et des fonctionnalités remplies par l'algorithme, afin de permettre un usage informé
- L'explicabilité est donc un enjeu non seulement pour s'adresser à un public sans aucun bagage technique en informatique, mais aussi à des personnes n'ayant pas les capacités psychologiques, les compétences cognitives ni même l'expérience de la vie d'un adulte.

Transparence d'un algorithme

L'explicabilité des algorithmes: Expliquer à qui ?

L'exigence d'explication peut se traduire de différentes manières selon les publics visés.

Pour le citoyen lambda, il peut s'agir de comprendre la logique qui a conduit à un résultat particulier.

- Une recommandation,
- L'exposition à une publicité,
- Le classement d'une information,
- L'affectation à une université
- Un refus de demande de prêt,
- etc.

Il peut être utile de lui fournir des informations sur les critères pris en compte par le système, leurs pondérations, ou encore des contre-exemples (comme des profils voisins du sien conduisant à des résultats d'affectation universitaire différents).

Transparence d'un algorithme

L'explicabilité des algorithmes: Expliquer comment ?

Explicabilité, un Défi majeur pour deux raisons principales:

- Certains algorithmes reposent sur des techniques, comme l'apprentissage profond ou “deep learning” qui produisent des résultats intrinsèquement difficiles à apprêhender pour des humains.
- La quantité importante de facteurs (données d'entrée) pris en compte est parfois un obstacle à l'intelligibilité.

Transparence d'un algorithme

L'explicabilité des algorithmes: Expliquer comment ?

Certains travaux reposent sur

- L'hypothèse de la “boîte noire”, c'est à dire l'absence de connaissance du code du système, et fonctionnent par analyse de ses entrées et sorties.
- L'hypothèse de la “boîte blanche” et procèdent par analyse du code lui-même

L'avantage de la première démarche est d'être indépendante du type de technique mise en œuvre par le système algorithmique. La seconde, en revanche, permet de disposer d'informations plus riches sur le fonctionnement du système.

Transparence d'un algorithme

L'explicabilité des algorithmes: Expliquer comment ?

Plutôt que de tenter d'assurer une forme d'explicabilité a posteriori, quand le système est déjà déployé, l'idéal serait de prendre en compte cette exigence dès la phase de conception d'un système algorithmique, et de garantir ainsi "par construction" une meilleure compréhension de leurs résultats.

Transparence d'un algorithme

L'intelligibilité de l'algorithme

- Un algorithme est **intelligible** s'il est possible au concepteur de comprendre son fonctionnement et de vérifier s'il satisfait bien les propriétés désirées.
- L'intelligibilité est une forme d'explicabilité fondamentale, qui porte sur la capacité des concepteurs à s'expliquer l'algorithme qu'ils conçoivent.
- Ses limitations sont dues aux limites de l'état de l'art scientifique, voire à des limites scientifiques fondamentales.
- En anglais, explicabilité et intelligibilité sont désignés ensemble par le terme "**interpretability**".

Transparence d'un algorithme

Les liens entre les propriétés

- L'explicabilité dépend de l'intelligibilité : il est nécessaire de comprendre pour expliquer.
- Intelligibilité = explicabilité fondamentale.
- L'intelligibilité est fondamentale pour vérifier qu'un algorithme est loyal et équitable :
 - Il faut expliquer pour être loyal.
 - Le manque d'intelligibilité peut créer des effets discriminants inattendus.

Plan

- 1 Models
- 2 Decision theory and Decision analysis
- 3 Main steps of developing a decision model
- 4 Algorithms & Transparency
- 5 Our program

Our Program

- **Chapter 1:** Introduction à la transparence des algorithmes.

TP1: Rappels du langage Python

- **Chapter 2:** Le filtrage collaboratif

TP2: Systèmes de recommandation

- **Chapter 3:** La Transparence en choix social

TP3: Les méthodes de votes

- **Chapter 4:** La prise en compte de l'interaction entre critères de Décision

Projet: Nutriscore ou Hôtels de Luxe ou produits 60 millions de consommateurs, ... (présentation du projet à la 4e séance)

- **Chapter 5:** Une méthode de classification ordonnée ELECTRE TRI (MR-Sort)

Suite Projet

- **Evaluation:** $NF = 20\% (TP2 + TP3) + 80\% \text{ Projet}$ (soutenances à la dernière séance)