

Analyse Ontologique de scénario dans un contexte Big Data

Marwan Batrouni*, Aurélie Bertaux*
Christophe Nicolle*

*Laboratoire électronique, Informatique et image (Le2i) - FRE CNRS 2005
Bât. I3M - Rue Sully - 21000 Dijon - France
prenom.nom@u-bourgogne.fr,
<http://checksem.fr>

Notre analyse du domaine prédictif a établi deux constats : il n'existe pas de définition universelle de scénario et l'analyse de scénarios ne prend pas en compte suffisamment de contraintes pour établir des prédictions de qualité prouvée. Nous proposons donc une définition universelle des scénarios pour répondre au premier point. Pour répondre au second nous nous sommes intéressés au fait que les prédictions sont plus précises si le volume et l'origine des données étaient conséquent et hétérogène. Ce constat nous a porté dans le domaine du Big Data auxquels doivent être ajoutés la sémantique, la Valeur et la Véracité des données pour une prédiction fiable et sensée. Nous montrons que les ontologies permettent de répondre à l'ensemble de ces critères. Par ailleurs, nous avons identifié que les machines d'états répondent aux besoins (en terme d'outil et de formalisation) de l'analyse de scénario. Nous proposons donc à partir de la définition uniformisée des scénarios, de les analyser par le biais des machines d'états basées sur des ontologies. Cet article ébauche les prérequis de formalisation par la combinaison de 3 domaines qui apportent chacun les réponses aux différents éléments de la définition : l'analyse de scénario, les ontologies et le Big Data.

1 Analyse de scénarios dans un environnement Big Data

Dans le contexte du Big Data, l'analyse de scénarios doit exploiter ses outils pour améliorer son efficacité. Pour cela deux technologies (parmi les 3 ci-dessous) dans la sphère Big Data sont particulièrement intéressantes, à savoir l'analyse prédictive et l'analyse prescriptive.

- **Analyse descriptive : ce qui s'est passé.** ETL, MapReduce, les lacs de données et les algorithmes de clustering...
- **Analyse prédictive : ce qui se passerait.** L'exploration de texte notamment via CRISP-DM.
- **Analyse prescriptive : ce qui devrait arriver.** L'apprentissage machine, la recherche opérationnelle, l'analyse statistique, et la modélisation et la simulation.

Le Big Data fournit des outils capables d'améliorer l'efficacité de l'analyse de scénarios. Cependant les prédictions devant être crédibles et pertinentes il faut prendre en compte la *Valeur* et la *Véracité* des données. Ces deux aspects sont traités par l'utilisation d'ontologies.

2 Les ontologies et le Big Data

Les ontologies sont définies par "une spécification formelle et explicite d'une conceptualisation partagée". Outre leur aspect formel, les ontologies présentent 4 facettes pertinentes : la sémantique par l'inférence pour répondre à la Valeur et pour répondre à la Véracité : la consistance (par la cohérence garantie par les axiomes de la logique de description), la décidabilité (par $SHOIN(\mathcal{D})$, $SROIQ(\mathcal{D})$ et $SHIF(\mathcal{D})$) et la complétude (par l'Open World Assumption, qui permet d'intégrer des connaissances nouvelles et hétérogènes).

3 Ontologies et analyse de scénario : formalisation

En première étape, nous avons établi des définitions et une terminologie concernant le domaine du discours (\mathcal{D}) et l'espace-état qui sont des blocs de construction essentiels pour construire le système dynamique à la base du concept *World*.

Un espace-état X est l'ensemble des n -tuples (d_0, d_1, \dots, d_n) résultant du produit cartésien de tous les éléments de \mathcal{D} .

$$X = \prod_{i \in I} D_i = \left\{ f : \left(f : I \rightarrow \bigcup_{i \in I} D_i \right) \wedge (\forall i \in I : (f(i) \in D_i) \wedge \phi(\{f(i)\})) \right\}$$

Où ϕ est le prédicat assurant l'intégrité de l'opération.

La description du *world*, i.e. une séquence d'états sélectionnés à partir de notre espace d'états ; on peut construire un *World* (i.e. l'ensemble des mondes possibles) à partir d'une liste d'exigences coïncidant avec les exigences pour une machine d'états.

Une machine d'état plus généralisée est de nature probabiliste appelée automate probabiliste, qui intègre des distributions de probabilités pour chaque transition possible d'un état à l'autre, ainsi que l'ajout à la définition d'un sixième tuple représentant un vecteur stochastique avec les probabilités de la machine d'état se trouvant dans un état initial donné.

Une façon de transformer une ontologie en un système dynamique est appelé *dérivation des relations ontologiques*, le processus implique une dérivation partielle des relations dans un graphe d'ontologie sur les événements dans un graphe de machine d'états. D'un point de vue conceptuel, il s'agit de ce qu'on appelle la *structure du premier ordre extensionnel* pour laquelle le concept déterministe de l'ontologie tient encore dans une ontologie dérivée. Dans cette interprétation, le domaine du discours et la machine d'états résultante mappent un *world* à une séquence possible d'états $w \equiv \{q_0, q_1, \dots, q_n\}$, dans le langage de l'automate. Par conséquence, la machine d'état est le producteur potentiel de tous les *World* possibles.

Définition 3.1. Scénario : Un scénario est une séquence conceptuelle et sémantiquement cohérente d'états, dans le but soit d'explorer l'avenir d'un système à l'étude, soit de recommander des décisions pour le rapprocher d'un état désiré.

Cette définition synthétise les points clés proposés dans diverses définitions précédentes de scénarios dans le domaine de l'analyse de scénario, à savoir : Séquence, Cohérence, Incertitude et Conceptuel. La définition ajoute également la dimension de cohérence sémantique et met l'accent sur les deux objectifs principaux des scénarios qui sont les objectifs exploratoires et normatifs.