Documentation sur le sujet

Sonny Klotz - Jean-Didier Pailleux - Malek Zemni

UVSQ

23/02/2017

Plan général

Plan

DCbrain est un logiciel qui permet de visualiser ce qui ce passe sur les **réseaux physiques (de fluide)** afin de trouver/prédire les problèmes et optimiser ces réseaux.

DCbrain est un logiciel qui permet de visualiser ce qui ce passe sur les **réseaux physiques (de fluide)** afin de trouver/prédire les problèmes et optimiser ces réseaux.

Cette visualisation provient de données collectées à partir des réseaux physiques (à l'aide de mesures, de capteurs IOT, etc.) puis analysées grâce aux technologies du **Big Data**.

DCbrain est un logiciel qui permet de visualiser ce qui ce passe sur les **réseaux physiques (de fluide)** afin de trouver/prédire les problèmes et optimiser ces réseaux.

Cette visualisation provient de données collectées à partir des réseaux physiques (à l'aide de mesures, de capteurs IOT, etc.) puis analysées grâce aux technologies du **Big Data**.

Réseaux physiques

Les réseaux qu'on va traiter dans le cadre de ce logiciel sont des réseaux industriels physiques (des réseaux de fluide, de distribution). Il s'agit des réseaux industriels tels les réseaux de distribution pétrolière, gazière, électrique, etc.

Big Data

Décrit des ensembles de très gros volumes de données, à la fois structurées, semi-structurées ou non structurées, qui peuvent être traitées et exploitées dans le but d'en tirer des informations intelligibles et pertinentes.

Big Data

Décrit des ensembles de très gros volumes de données, à la fois structurées, semi-structurées ou non structurées, qui peuvent être traitées et exploitées dans le but d'en tirer des informations intelligibles et pertinentes.

Exemple de sources :

Big Data

Décrit des ensembles de très gros volumes de données, à la fois structurées, semi-structurées ou non structurées, qui peuvent être traitées et exploitées dans le but d'en tirer des informations intelligibles et pertinentes.

Exemple de sources :

- Capteurs utilisés pour collecter les informations climatiques, de trafic, consommation (Smart cities, Internet des Objets).

Big Data

Décrit des ensembles de très gros volumes de données, à la fois structurées, semi-structurées ou non structurées, qui peuvent être traitées et exploitées dans le but d'en tirer des informations intelligibles et pertinentes.

Exemple de sources :

- Capteurs utilisés pour collecter les informations climatiques, de trafic, consommation (Smart cities, Internet des Objets).
- Messages sur les réseaux sociaux

Big Data

Décrit des ensembles de très gros volumes de données, à la fois structurées, semi-structurées ou non structurées, qui peuvent être traitées et exploitées dans le but d'en tirer des informations intelligibles et pertinentes.

Exemple de sources :

- Capteurs utilisés pour collecter les informations climatiques, de trafic, consommation (Smart cities, Internet des Objets).
- Messages sur les réseaux sociaux
- Enregistrements transactionnels d'achat en ligne

Big Data

Décrit des ensembles de très gros volumes de données, à la fois structurées, semi-structurées ou non structurées, qui peuvent être traitées et exploitées dans le but d'en tirer des informations intelligibles et pertinentes.

Exemple de sources :

- Capteurs utilisés pour collecter les informations climatiques, de trafic, consommation (Smart cities, Internet des Objets).
- Messages sur les réseaux sociaux
- Enregistrements transactionnels d'achat en ligne
- Signaux GPS de téléphones mobile

Problème

Comment utiliser et donner du sens à ces masses de données enregistrées sur ces réseaux ?

Plan

Définition

Ensemble de techniques de statistique descriptive.

Définition

Ensemble de techniques de statistique descriptive.

- **Objectifs** : une description succincte, regrouper les données.

Définition

Ensemble de techniques de statistique descriptive.

- **Objectifs** : une description succincte, regrouper les données.
- **Les données** : tableaux de données quantitatives et qualitatives.

Définition

Ensemble de techniques de statistique descriptive.

- **Objectifs** : une description succincte, regrouper les données.
- **Les données** : tableaux de données quantitatives et qualitatives.
- **Avantages** : traitement en masse, représentations graphiques.

Plan

Big Data

Le Big Data fait référence à la masse de données collectée. On considère du Big Data quand le traitement devient trop long pour une seule machine.

Big Data

Le Big Data fait référence à la masse de données collectée. On considère du Big Data quand le traitement devient trop long pour une seule machine.

Le Big Data est caractérisé par les 3V :

Big Data

Le Big Data fait référence à la masse de données collectée. On considère du Big Data quand le traitement devient trop long pour une seule machine.

Le Big Data est caractérisé par les 3V :

- le Volume de données considérable à traiter.

Big Data

Le Big Data fait référence à la masse de données collectée. On considère du Big Data quand le traitement devient trop long pour une seule machine.

Le Big Data est caractérisé par les 3V :

- le Volume de données considérable à traiter.
- la Variété de ces données qui peuvent être brutes, non structurées ou semi-structurées

Big Data

Le Big Data fait référence à la masse de données collectée. On considère du Big Data quand le traitement devient trop long pour une seule machine.

Le Big Data est caractérisé par les 3V :

- le Volume de données considérable à traiter.
- la Variété de ces données qui peuvent être brutes, non structurées ou semi-structurées
- la Vélocité qui désigne le fait que ces données sont produites, récoltées et analysées en temps réel.

Big Data

Le Big Data fait référence à la masse de données collectée. On considère du Big Data quand le traitement devient trop long pour une seule machine.

Le Big Data est caractérisé par les 3V :

- le Volume de données considérable à traiter.
- la Variété de ces données qui peuvent être brutes, non structurées ou semi-structurées
- la Vélocité qui désigne le fait que ces données sont produites, récoltées et analysées en temps réel.

Les traitements de cette quantité importante de données est massivement "parallélisé" avec MapReduce/Hadoop.

Le Machine Learning est :

Le Machine Learning est :

- Une discipline scientifique centrée sur le développement, l'analyse et l'implémentation de méthodes automatisables, offrant la possibilité à une machine d'évoluer grâce à un processus d'apprentissage à partir des données et à effectuer des tâches de façon performante.

Le Machine Learning est :

- Une discipline scientifique centrée sur le développement, l'analyse et l'implémentation de méthodes automatisables, offrant la possibilité à une machine d'évoluer grâce à un processus d'apprentissage à partir des données et à effectuer des tâches de façon performante.
- Un traitement statistique de masses de données réunissant à la fois mathématiques appliquées et informatique.

Le Machine Learning est :

- Une discipline scientifique centrée sur le développement, l'analyse et l'implémentation de méthodes automatisables, offrant la possibilité à une machine d'évoluer grâce à un processus d'apprentissage à partir des données et à effectuer des tâches de façon performante.
- Un traitement statistique de masses de données réunissant à la fois mathématiques appliquées et informatique.
- Utilisé lorsque le Big Data rend inopérant les méthodes statistiques traditionnelles.

Le Machine Learning est :

- Une discipline scientifique centrée sur le développement, l'analyse et l'implémentation de méthodes automatisables, offrant la possibilité à une machine d'évoluer grâce à un processus d'apprentissage à partir des données et à effectuer des tâches de façon performante.
- Un traitement statistique de masses de données réunissant à la fois mathématiques appliquées et informatique.
- Utilisé lorsque le Big Data rend inopérant les méthodes statistiques traditionnelles.

Le Machine Learning est composé de plusieurs types d'algorithmes d'apprentissage (supervisé, non supervisé, semi-supervisé, par renforcement).

Plan

Les graphes sont un outil très efficace pour la représentation de structures complexes comme les réseaux physiques dans notre cas.

Les graphes sont un outil très efficace pour la représentation de structures complexes comme les réseaux physiques dans notre cas.

DCbrain emploie une approche basée sur une représentation de ces réseaux physiques en **graphes de flux**.

Les graphes sont un outil très efficace pour la représentation de structures complexes comme les réseaux physiques dans notre cas.

DCbrain emploie une approche basée sur une représentation de ces réseaux physiques en **graphes de flux**.

Cette représentation en graphe permet de prendre en compte les spécifités des flux du réseau, c'est-à-dire :

Les graphes sont un outil très efficace pour la représentation de structures complexes comme les réseaux physiques dans notre cas.

DCbrain emploie une approche basée sur une représentation de ces réseaux physiques en **graphes de flux**.

Cette représentation en graphe permet de prendre en compte les spécifités des flux du réseau, c'est-à-dire :

- retranscrire les données du réseau sous forme de flux

Les graphes sont un outil très efficace pour la représentation de structures complexes comme les réseaux physiques dans notre cas.

DCbrain emploie une approche basée sur une représentation de ces réseaux physiques en **graphes de flux**.

Cette représentation en graphe permet de prendre en compte les spécifités des flux du réseau, c'est-à-dire :

- retranscrire les données du réseau sous forme de flux
- analyser (calculer) et représenter les données liées au flux du réseau

La représentation du réseau en graphe de flux a pour avantage de :

La représentation du réseau en graphe de flux a pour avantage de :

- repérer beaucoup plus facilement des anomalies dans le réseau

La représentation du réseau en graphe de flux a pour avantage de :

- repérer beaucoup plus facilement des anomalies dans le réseau
- simuler des évolutions du réseau

La représentation du réseau en graphe de flux a pour avantage de :

- repérer beaucoup plus facilement des anomalies dans le réseau
- simuler des évolutions du réseau

Ce genre de graphe peut être utilisé pour tout réseau physique de fluide, par exemple les réseaux électriques :

La représentation du réseau en graphe de flux a pour avantage de :

- repérer beaucoup plus facilement des anomalies dans le réseau
- simuler des évolutions du réseau

Ce genre de graphe peut être utilisé pour tout réseau physique de fluide, par exemple les réseaux électriques :

Graphe de flux : d'un réseau électrique :

- Nœuds : des connections
- Arcs : canaux pour acheminer l'électricité (câbles)

Les méthodes d'analyse de données classiques sont limitées au **données structurées**.

Les méthodes d'analyse de données classiques sont limitées au **données structurées**.

Données structurées

Données organisées et représentées dans un format prédéfini, selon une structure permettant de les traiter selon diverses combinaisons, afin de mieux exploiter les informations. Exemple : les bases de données relationnelles.

Les méthodes d'analyse de données classiques sont limitées au **données structurées**.

Données structurées

Données organisées et représentées dans un format prédéfini, selon une structure permettant de les traiter selon diverses combinaisons, afin de mieux exploiter les informations. Exemple : les bases de données relationnelles.

Données non-structurées

Données brutes non-organisées selon un format prédéfini qui permet d'y accéder et de les traiter plus facilement. Exemple : du texte brut.

Il est nécessaire d'employer une méthode d'analyse qui convient à la structure des données représentées par les graphes.

Il est nécessaire d'employer une méthode d'analyse qui convient à la structure des données représentées par les graphes.

L'analyse des données à partir d'un graphe (de flux dans notre cas) est réalisée grâce à la méthode de **graph mining**.

Il est nécessaire d'employer une méthode d'analyse qui convient à la structure des données représentées par les graphes.

L'analyse des données à partir d'un graphe (de flux dans notre cas) est réalisée grâce à la méthode de **graph mining**.

Le graph mining est donc une forme d'analyse de données semi-structurée dont le processus consiste à trouver et extraire des informations utiles à partir d'une masse de **données semi-structurées**.

Il est nécessaire d'employer une méthode d'analyse qui convient à la structure des données représentées par les graphes.

L'analyse des données à partir d'un graphe (de flux dans notre cas) est réalisée grâce à la méthode de **graph mining**.

Le graph mining est donc une forme d'analyse de données semi-structurée dont le processus consiste à trouver et extraire des informations utiles à partir d'une masse de **données semi-structurées**.

Données semi-structurées

Forme intermédiaire entre données structurées et non-structurées, elles ne sont pas organisées selon une format prédéfini, mais comportent néanmoins des informations associées (telles que des balises de métadonnées) qui les rendent plus faciles à traiter (en permettent l'adressage des éléments qu'elles renferment). Ce genre de données peut être représentée par un **graphe**. Exemple : des données XML ou HTML.

Il s'agit d'extraire des sous-graphes (miner des sous graphes fréquent, répétitifs dans le graphe en entrée) qui décrivent l'information recherchée du graphe. Ces informations peuvent ensuite être utilisées pour :

Il s'agit d'extraire des sous-graphes (miner des sous graphes fréquent, répétitifs dans le graphe en entrée) qui décrivent l'information recherchée du graphe. Ces informations peuvent ensuite être utilisées pour :

- classification et catégorisation du graphe (par l'analyse de la fréquence des sous graphes)

Il s'agit d'extraire des sous-graphes (miner des sous graphes fréquent, répétitifs dans le graphe en entrée) qui décrivent l'information recherchée du graphe. Ces informations peuvent ensuite être utilisées pour :

- classification et catégorisation du graphe (par l'analyse de la fréquence des sous graphes)
- effectuer des regroupement : trouver des relations entre différents éléments du graphe (par exemple trouver un groupe d'amis inter-connectés dans un graphe de réseau social)

Il s'agit d'extraire des sous-graphes (miner des sous graphes fréquent, répétitifs dans le graphe en entrée) qui décrivent l'information recherchée du graphe. Ces informations peuvent ensuite être utilisées pour :

- classification et catégorisation du graphe (par l'analyse de la fréquence des sous graphes)
- effectuer des regroupement : trouver des relations entre différents éléments du graphe (par exemple trouver un groupe d'amis inter-connectés dans un graphe de réseau social)
- prédire le comportement des éléments d'un graphe (prédire les préférence des utilisateurs dans un graphe de réseaux sociaux, prédire les évolutions dans un graphe de réseau physique)

Domaines d'utilisation :

Domaines d'utilisation :

- Social network analysis (analyse des groupes d'amis dans les réseaux sociaux)

Domaines d'utilisation :

- Social network analysis (analyse des groupes d'amis dans les réseaux sociaux)
- Applications chimiques et biologiques (développement de médicaments par l'analyse du graphe moléculaire et de ses évolutions)

Domaines d'utilisation :

- Social network analysis (analyse des groupes d'amis dans les réseaux sociaux)
- Applications chimiques et biologiques (développement de médicaments par l'analyse du graphe moléculaire et de ses évolutions)
- Réseaux physiques (prédire les évolutions)

Domaines d'utilisation :

- Social network analysis (analyse des groupes d'amis dans les réseaux sociaux)
- Applications chimiques et biologiques (développement de médicaments par l'analyse du graphe moléculaire et de ses évolutions)
- Réseaux physiques (prédire les évolutions)

Principaux algorithmes :

Domaines d'utilisation :

- Social network analysis (analyse des groupes d'amis dans les réseaux sociaux)
- Applications chimiques et biologiques (développement de médicaments par l'analyse du graphe moléculaire et de ses évolutions)
- Réseaux physiques (prédire les évolutions)

Principaux algorithmes :

- Apriori-based Approach

Domaines d'utilisation :

- Social network analysis (analyse des groupes d'amis dans les réseaux sociaux)
- Applications chimiques et biologiques (développement de médicaments par l'analyse du graphe moléculaire et de ses évolutions)
- Réseaux physiques (prédire les évolutions)

Principaux algorithmes :

- Apriori-based Approach
- Pattern-Growth Approach

Plan

Conclusion

Et notre application ...