

Systemes de recommandations et Données environnementales



NEGRE Elsa

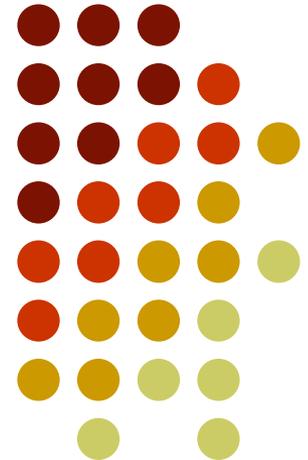
Université Paris-Dauphine

Elsa.Negre @lamsade.dauphine.fr

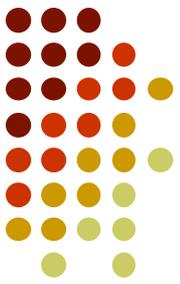
LAMSADE

UMR CNRS 7243

laboratoire d'analyse et modélisation de systèmes pour l'aide à la décision

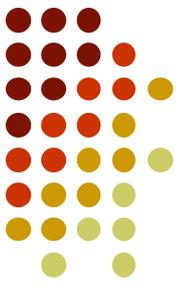


Plan



- Qu'est ce qu'un système de recommandations ?
- Les méthodes les plus utilisées
- Les améliorations possibles
- Un petit mot sur l'évaluation
- Applications
 - Gestion de crises - Systèmes d'alertes précoces
 - Smart Cities
 - Ailleurs...
- Conclusion et perspectives

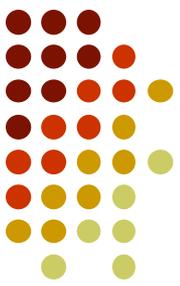
Un système de recommandation – Quèsaco?



- Un système de recommandation (RS) aide les utilisateurs qui n'ont pas suffisamment d'expérience ou la compétence nécessaire pour évaluer le nombre, potentiellement important, d'alternatives offertes par un site (web).
- Dans leur forme la plus simple les RS
 - recommandent à leurs utilisateurs des listes personnalisées et classées d'articles
 - fournissent aux consommateurs des renseignements pour les aider à décider quels articles acheter
- Les articles peuvent être :
 - musique, livres, films, ...
- De manière générale, on parle d'items

Un système de recommandation

– Quèsaco? (2)



$$r_{RS} : Users \times Items \rightarrow Ratings$$

- Problématique : estimation de scores pour des articles non encore évalués
- Définition [Hill+:CHI'95] :
 - Soit
 - C : ensemble de tous les utilisateurs
 - S : ensemble de tous les articles possibles
 - u : fonction d'utilité d'un article s à l'utilisateur c
 - Alors, choisir l'article s' non encore évalué de S qui maximise l'utilité de chaque utilisateur c de C

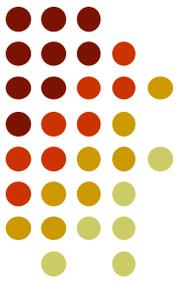


U(c,s)	Bridget Jones Baby	Jack Reacher 2	7 Mercenaries	Jack Reacher 1	Bridget Jones Diary	The Good, the Bad and the Ugly
Elsa				😊	😊	☹️
Zahra	😊				😊	😊
Dario			☹️			☹️



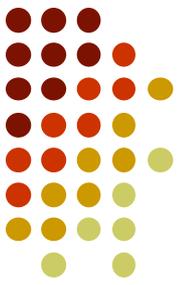
Elsa	😊	😊	☹️	😊	😊	☹️
------	---	---	----	---	---	----

Plan

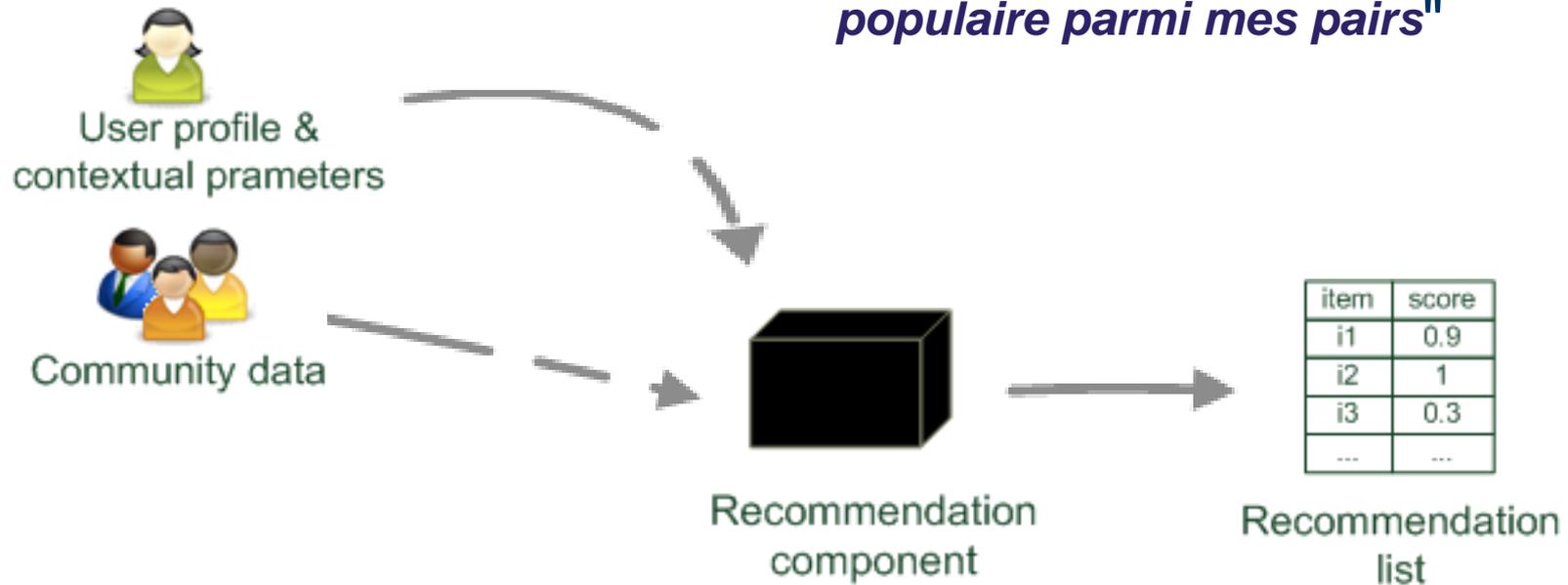


- Qu'est ce qu'un système de recommandations ?
- **Les méthodes les plus utilisées**
- Les améliorations possibles
- Un petit mot sur l'évaluation
- Applications
 - Gestion de crises - Systèmes d'alertes précoces
 - Smart Cities
 - Ailleurs...
- Conclusion et perspectives

Les méthodes les plus utilisées (1)

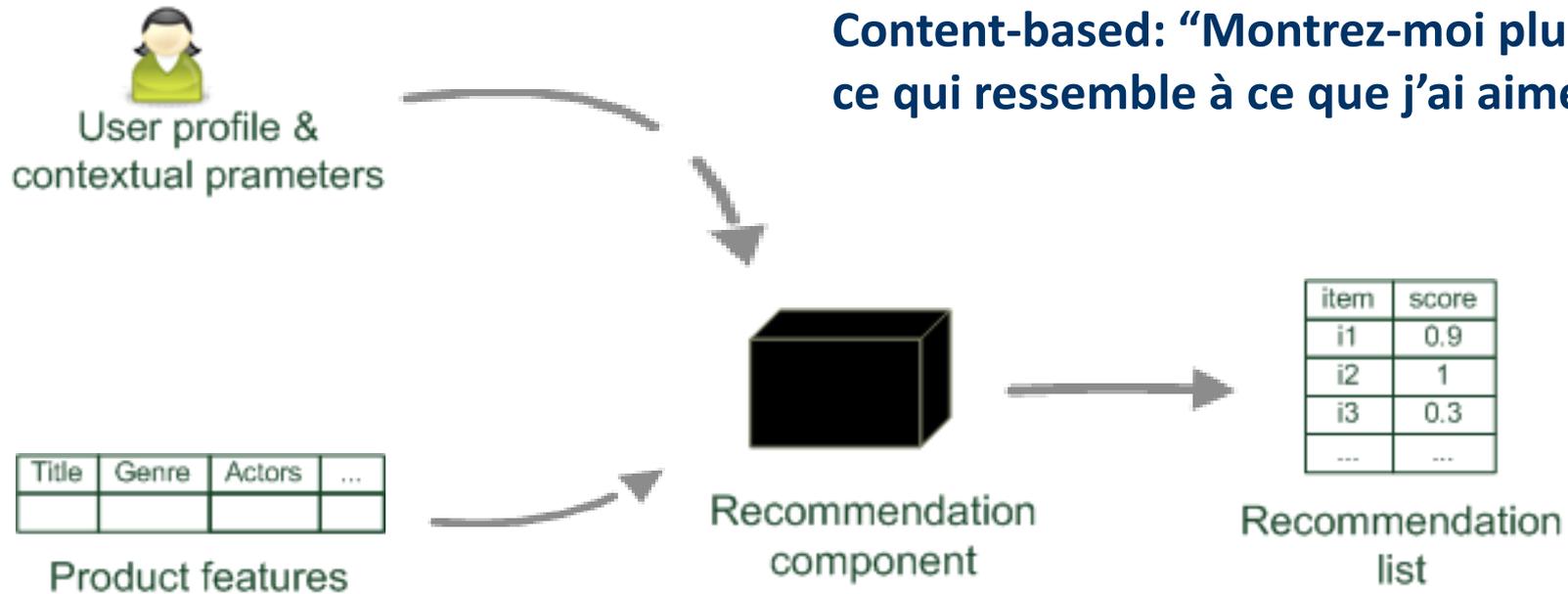
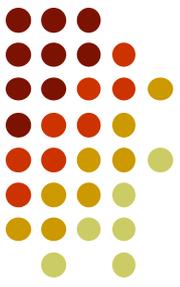


Collaborative: "Dites-moi ce qui est populaire parmi mes pairs"



Idée : Essayer de prédire l'opinion de l'utilisateur qu'il aura sur les différents items et être en mesure de recommander le « meilleur » item à chaque utilisateur en fonction des goûts/avis précédents de l'utilisateur et des avis d'autres utilisateurs qui lui sont semblables .

Les méthodes les plus utilisées (2)

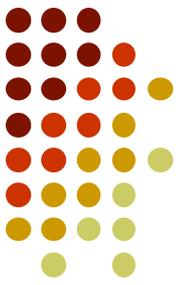


Idée : fait coïncider le profil de l’item avec le profil de l’utilisateur pour décider de sa pertinence pour l’utilisateur.

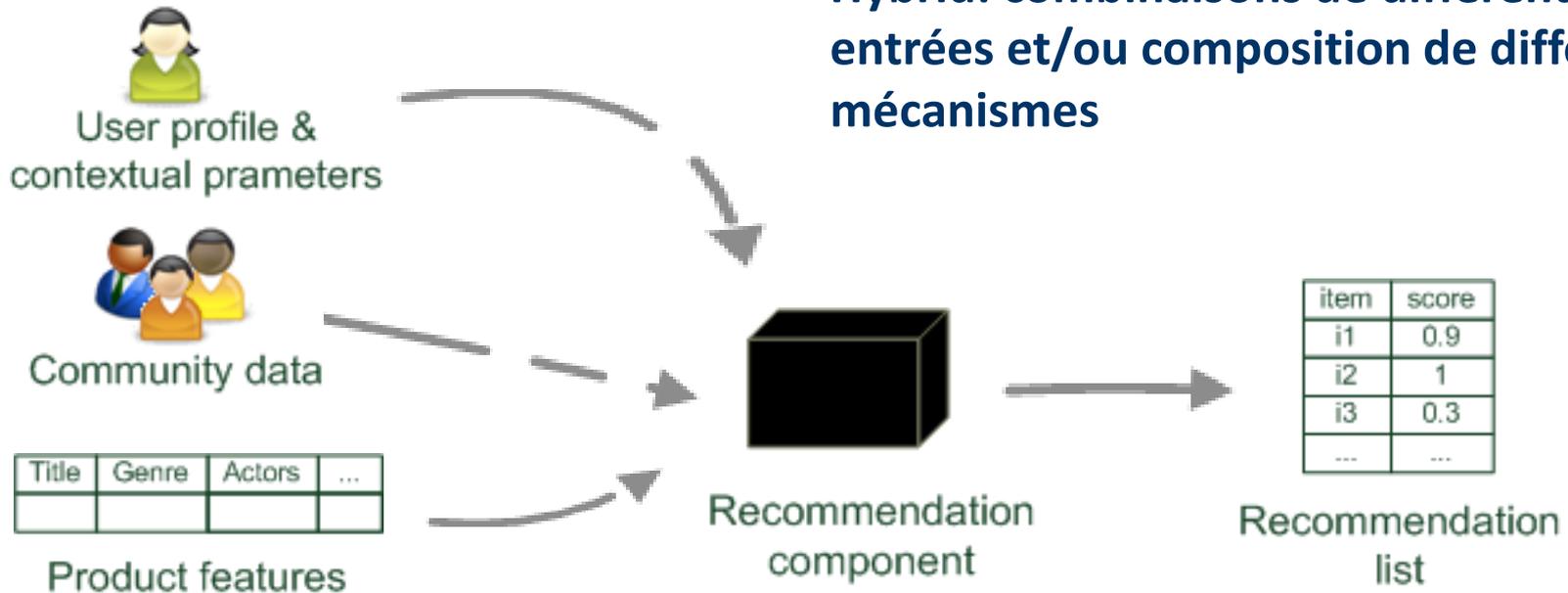
Le profil de l’utilisateur est basé sur les items que l’utilisateur a aimé dans le passé ou sur les intérêts qu’il a explicitement définis.

	Avantages	Inconvénients
Approches collaboratives	<ul style="list-style-type: none"> 👍 Utilise des recommandations d'autres utilisateurs (score) pour évaluer l'utilité des items 👍 Plus il y a d'utilisateurs, plus il y a de scores : meilleurs sont les résultats 	<ul style="list-style-type: none"> 👎 Trouver des utilisateurs ou groupes d'utilisateurs similaires est difficile 👎 <i>Cold start problem</i> <ul style="list-style-type: none"> 👎 Nouvel utilisateur = Pas de préférences 👎 Nouvel item = Pas de score 👎 Faible densité de la matrice utilisateur/score 👎 Portfolio Effect: Non Diversity Problem
Approches basées sur le contenu	<ul style="list-style-type: none"> 👍 Le profil des utilisateurs est la clé 👍 Le matching entre les préférences de l'utilisateur et les caractéristiques des items fonctionne aussi pour les données textuelles 👍 Pas besoin de données sur les autres utilisateurs. 👍 Pas de problème de faible densité 👍 Possibilité de faire des recommandations à des utilisateurs avec des goûts « uniques ». 👍 Possibilité de recommander de nouveaux items ou même des items qui ne sont pas populaires 	<ul style="list-style-type: none"> 👎 Tous les contenus ne peuvent pas être représentés avec des mots-clés 👎 Les utilisateurs avec des milliers d'achats/items sont un problème 👎 Nouvel utilisateur : pas d'historique ← Cold-start problem 👎 Les profils utilisateurs... 👎 Entièrement basé sur les scores d'articles et de sujets d'intérêt: moins il y a de scores, plus l'ensemble de recommandations possibles est limité. 👎 Impossible d'exploiter les jugements des autres utilisateurs

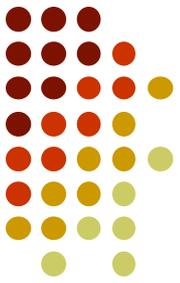
Les méthodes les plus utilisées (3)



Hybrid: combinaisons de différentes entrées et/ou composition de différents mécanismes

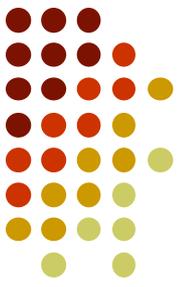


Plan



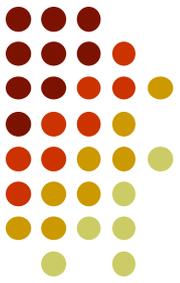
- Qu'est ce qu'un système de recommandations ?
- Les méthodes les plus utilisées
- **Les améliorations possibles**
- Un petit mot sur l'évaluation
- Applications
 - Gestion de crises - Systèmes d'alertes précoces
 - Smart Cities
 - Ailleurs...
- Conclusion et perspectives

Améliorations - Contextualisation

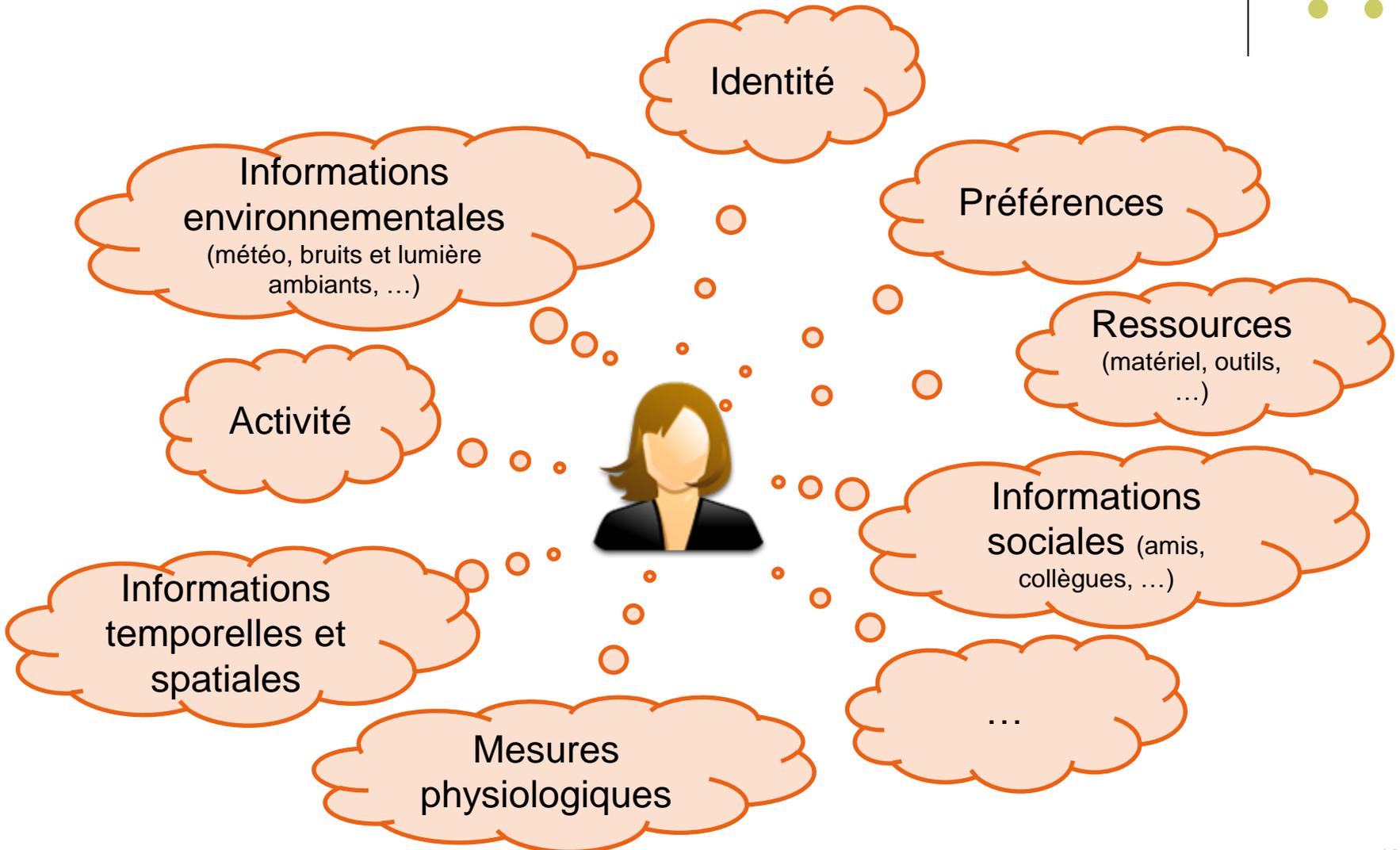


- Plus de 150 définitions différentes du contexte existent dans différents domaines...
- Définition de contexte la plus acceptée en *context-aware computing*:
 - “Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves” (Dey, 2001)
 - “A system is context-aware if it uses context to provide relevant information and/or services to the user, where relevancy depends on the user's task”. (Dey, 2001)

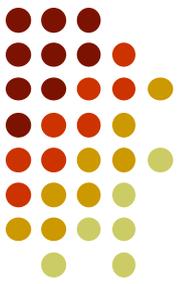
Obtenir le contexte?
Pertinence du contexte?



Contexte



CARS (1)

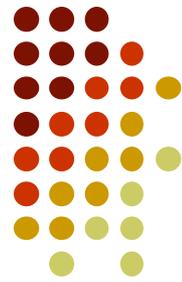


- Systèmes de recommandation = aider à réduire la surcharge d'information
- Approche complémentaire intéressante = intégrer les informations/données contextuelles

➔ Systèmes de recommandation contextuels (CARS) qui doivent générer des recommandations plus pertinentes en s'adaptant à un contexte spécifique de l'utilisateur.

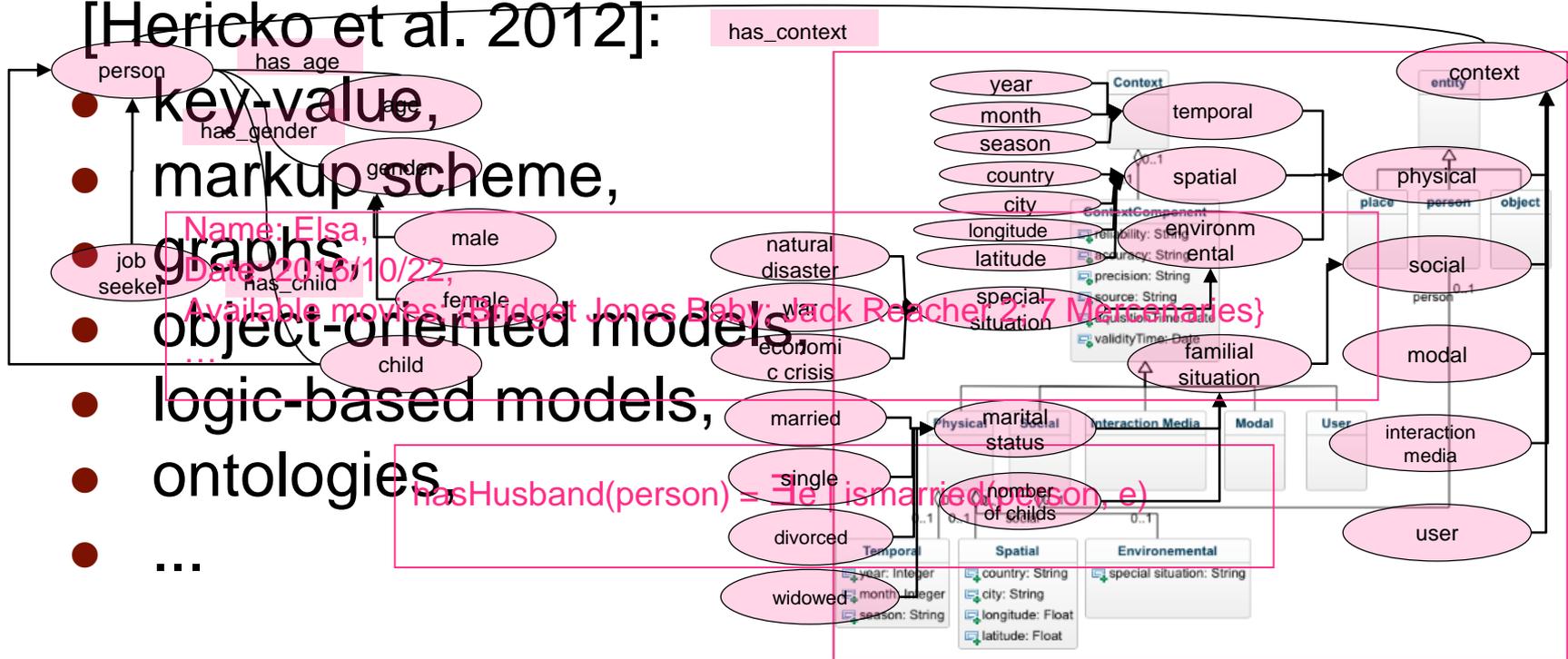
$r_{RS} : Users \times Items \rightarrow Ratings$

➔ $r_{CARS} : Users \times Items \times Contexts \rightarrow Ratings$



Modélisation du contexte

Différentes approches pour modéliser le contexte
[Hericko et al. 2012]:



CARS (2)

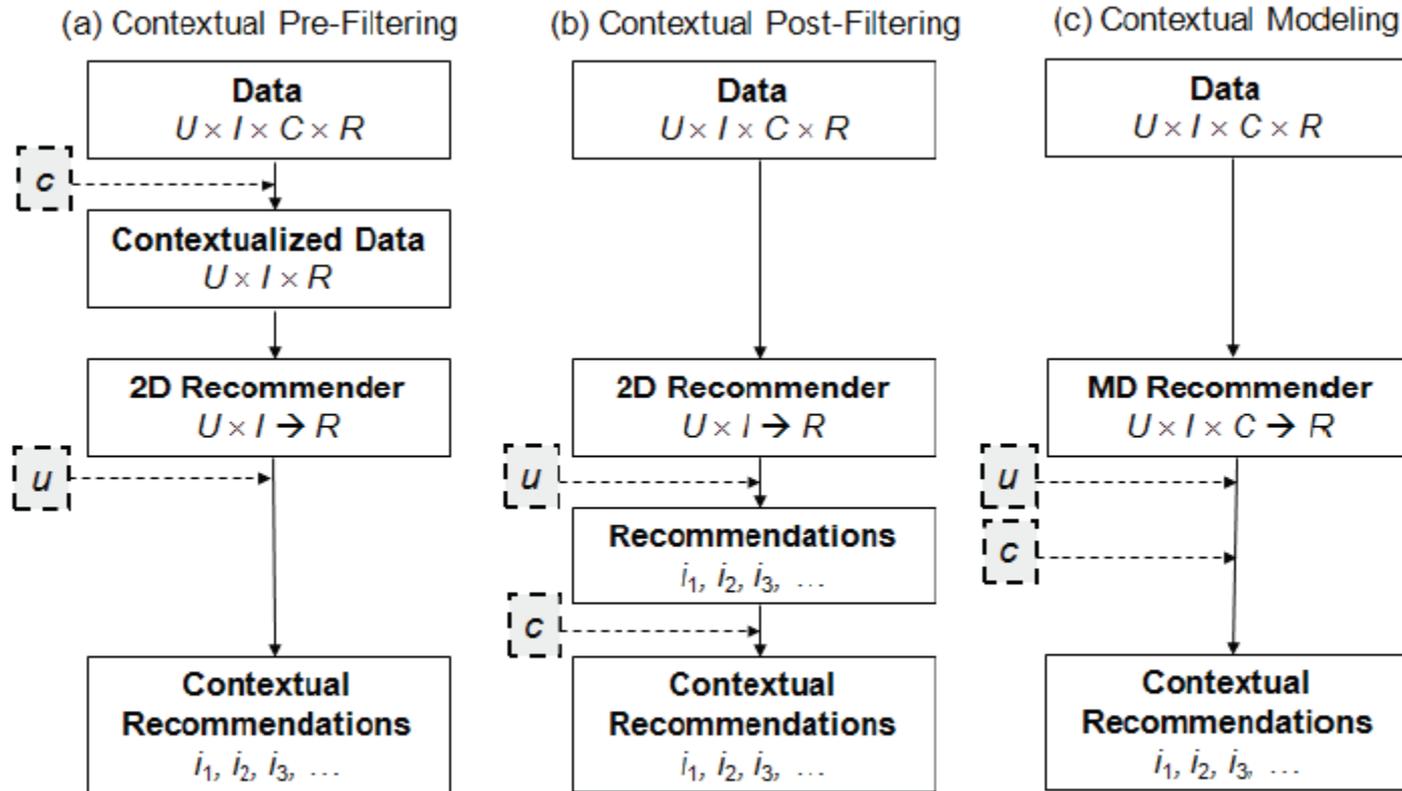
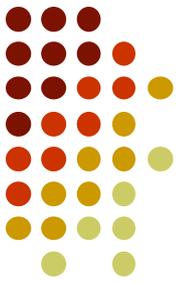
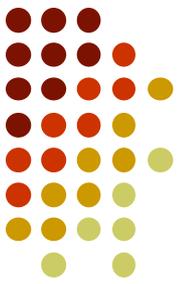


Fig. 4 Paradigms for incorporating context in recommender systems.

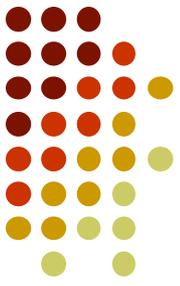
Source : Adomavicius & Tuzhilin, 2010

Plan



- Qu'est ce qu'un système de recommandations ?
- Les méthodes les plus utilisées
- Les améliorations possibles
- **Un petit mot sur l'évaluation**
- Applications
 - Gestion de crises - Systèmes d'alertes précoces
 - Smart Cities
 - Ailleurs...
- Conclusion et perspectives

Evaluation des systèmes de recommandation



- Mean Absolute Error (*MAE*) calcule l'écart entre les scores prédits et les scores réels.

$$MAE = \frac{\sum_{c \in C} \sum_{i \in I_{test_c}} |reco(c,i) - r_{c,i}|}{\sum_{c \in C} |I_{test_c}|}$$

- Root Mean Square Error (*RMSE*) , similaire à la MAE, mais met l'accent sur les plus grands écarts

$$RMSE = \sqrt{\frac{\sum_{c \in C} \sum_{i \in I_{test_c}} (reco(c,i) - r_{c,i})^2}{\sum_{c \in C} |I_{test_c}|}}$$

- Rappel et précision

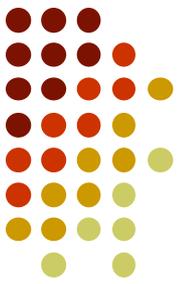
$$Précision_c = \frac{\text{Nombre de succès pour } c}{\text{Nombre total d'éléments recommandés à } c} \quad Rappel_c = \frac{\text{Nombre de succès pour } c}{\text{Nombre total d'éléments pertinents existants}}$$

- Le « rank score » est défini comme le rapport entre le score des items corrects et le meilleur score théorique réalisable pour l'utilisateur

$$RankScore = 100 \cdot \frac{\sum_c RankScore_c}{\sum_c RankScore_c^{max}}$$

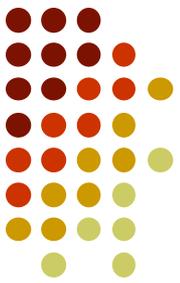
- ...

Plan



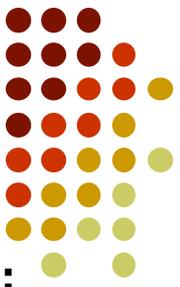
- Qu'est ce qu'un système de recommandations ?
- Les méthodes les plus utilisées
- Les améliorations possibles
- Un petit mot sur l'évaluation
- **Applications**
 - Gestion de crises - Systèmes d'alertes précoces
 - Smart Cities
 - Ailleurs...
- Conclusion et perspectives

Gestion de crises



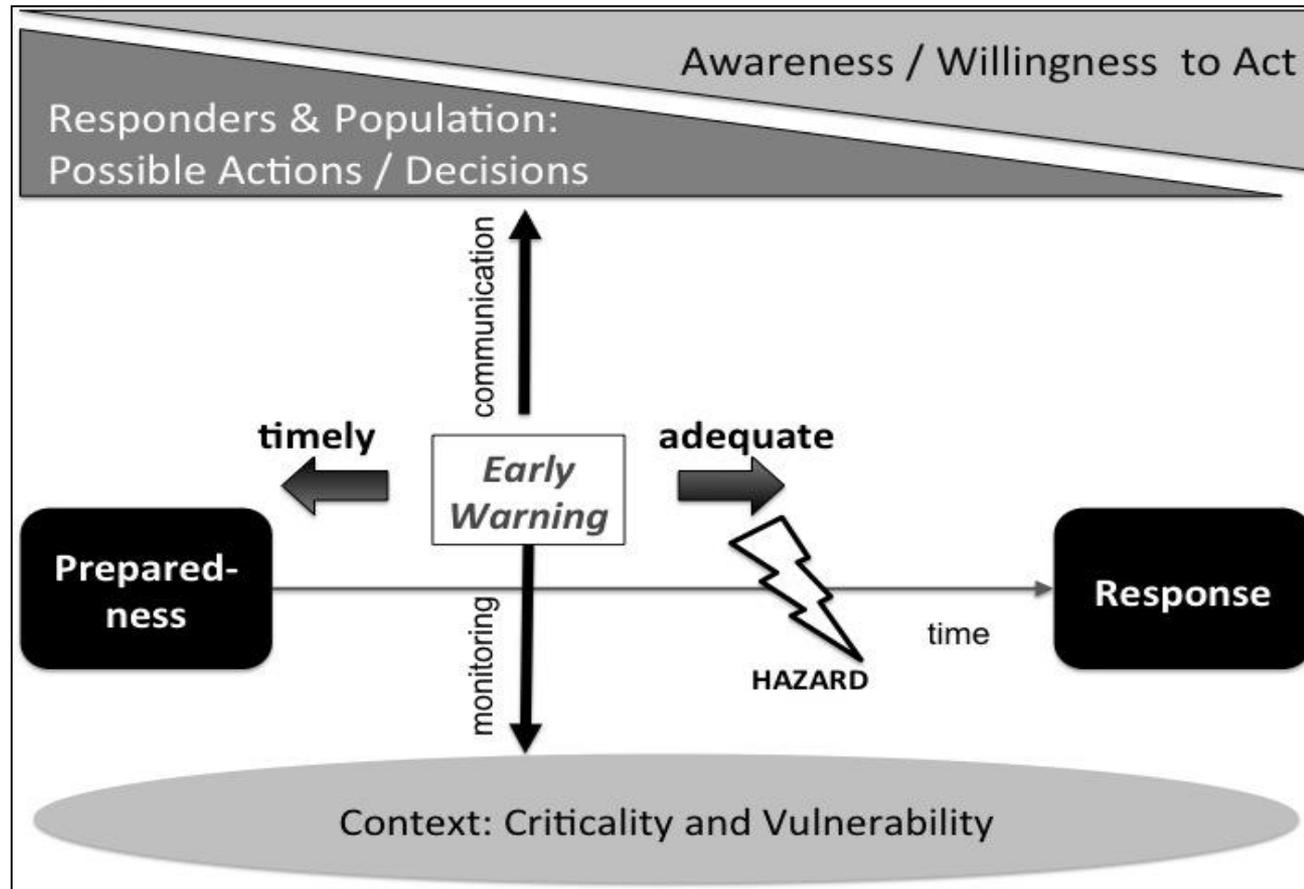
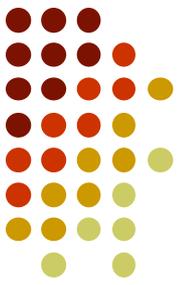
- La **gestion de crise** est l'ensemble des modes d'organisation, des techniques et des moyens qui permettent à une organisation de se préparer et de faire face à la survenance d'une crise puis de tirer les enseignements de l'évènement pour améliorer les procédures et les structures dans une vision prospective.
- Crise : risque pandémique, maladie émergente, crise sociale, crise économique, crise politique/géopolitique, catastrophe naturelle, conflits, ...

Systeme d'alerte precoce (EWS)



- Un **systeme d'alertes precoces** peut être defini comme une chaîne de systemes de communication d'information comprenant des capteurs, de la detection, de la decision et des sous-systemes transitoires, dans cet ordre, qui travaillent en collaboration, pour prevenir et signaler des perturbations affectant negativement la stabilite du monde physique ; et qui donnent suffisamment de temps au systeme de reponse pour preparer les ressources et les mesures d'intervention (actions) pour minimiser l'impact sur la stabilite du monde physique. (Waidyanatha, 2010)

Contexte décisionnel des EWS

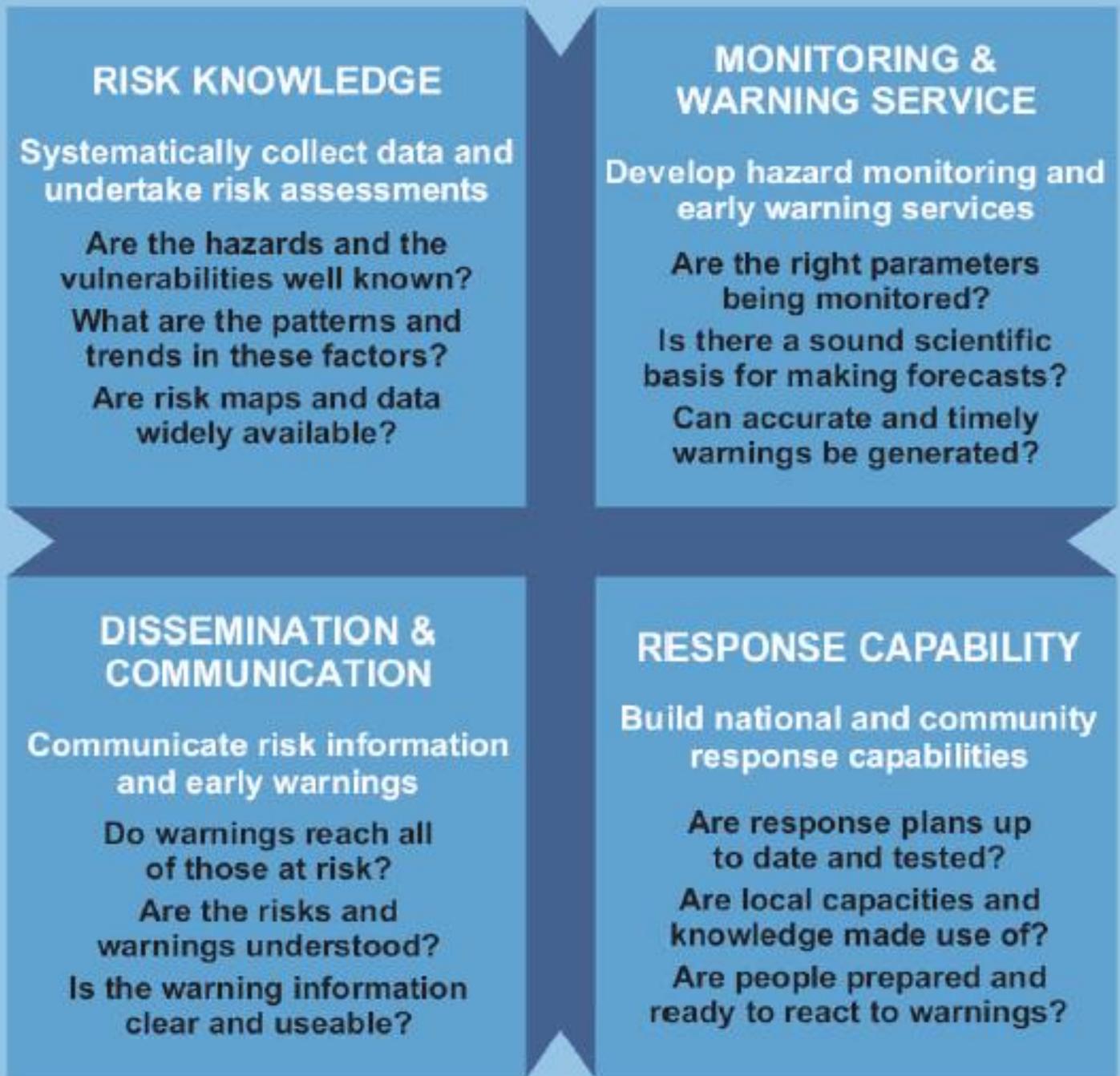


Plus l'alerte est déclenchée tôt, plus il y a de temps pour organiser et coordonner des actions préventives.

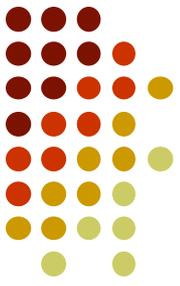
Malheureusement, les informations relatives au "danger" deviennent de plus en plus précises au fur et à mesure que le temps passe...

Un “bon” EWS

Les 4
éléments d'un
système
d'alertes
précoces



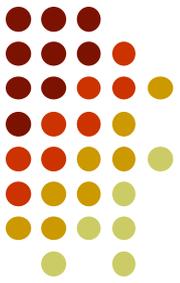
Les systèmes de recommandation en EWS



$u(w,a)$	Action 1	Action 2	Action 3	Action 4	Action 5
Alerte 1	8	7			
Alerte 2	9		3		
Alerte 3	3	5		5	5
Alerte 4	5	3		3	3

- Recommandations en EWS :
- Etant donné :
 - Un log d'alertes
 - Des indicateurs correspondants aux alertes
 - Une alerte déclenchée (courante)
- Soit
 - A : ensemble des actions possibles
 - W : ensemble d'alertes
 - u : fonction d'utilité d'une action a pour une alerte w
- Alors choisir l'action a' non encore évaluée de A qui maximise l'utilité pour chaque alerte w de W

Plan



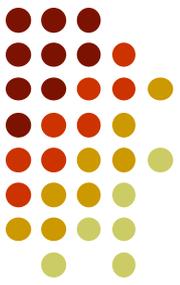
- Qu'est ce qu'un système de recommandations ?
- Les méthodes les plus utilisées
- Les améliorations possibles
- Un petit mot sur l'évaluation
- **Applications**
 - Gestion de crises - Systèmes d'alertes précoces
 - **Smart Cities**
 - Ailleurs...
- Conclusion et perspectives

Smart Cities



- De nombreuses villes à travers le monde se définissent comme « ville intelligente », mais cette définition est loin d'être bien positionnée et standardisée.
- Les villes se consacrent également aux TIC (Technologies de l'Information et de la Communication) pour soutenir les grandes collections de données et leur traitement, liés à la communication entre les citoyens et les institutions, les services privés et publics numériques, ...
- Intérêt: Améliorer les stratégies, les choix et les investissements de la ville

Smart Cities (2)

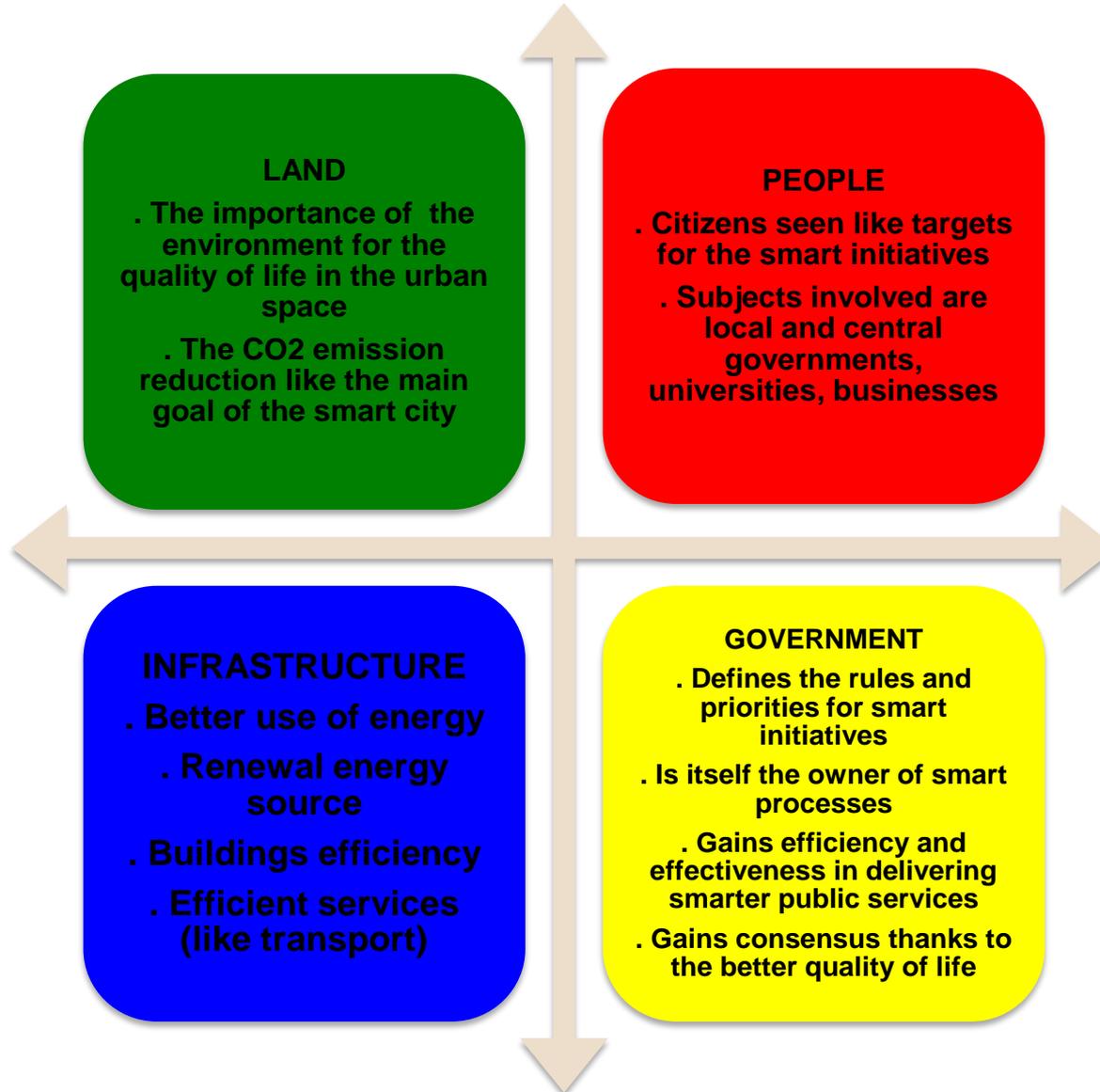
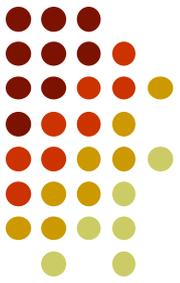


Cities are smart:

“when investments in human and social capital and traditional (transport) and modern (ICT) communication infrastructure fuel sustainable economic growth and a high quality of life, with a wise management of natural resources, through participatory governance”

(Caragliu et.al, 2011)

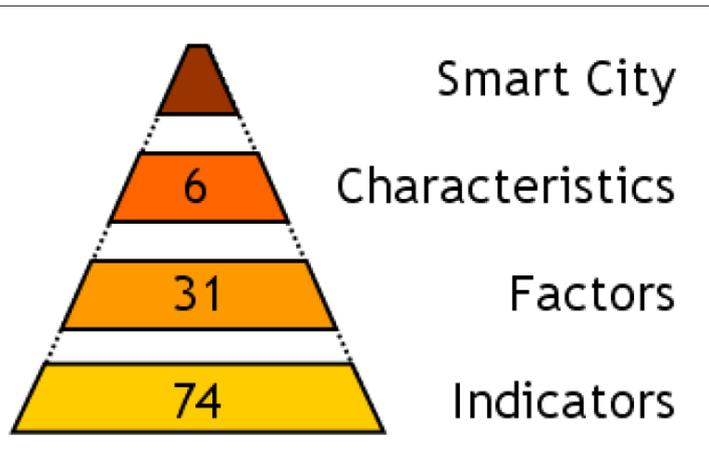
Smart Cities (3)



Dans la pratique... Classification



(Giffinger, 2007)



SMART ECONOMY (Competitiveness)

- Innovative spirit
- Entrepreneurship
- Economic image & trademarks
- Productivity
- Flexibility of labour market
- International embeddedness
- *Ability to transform*

SMART PEOPLE (Social and Human Capital)

- Level of qualification
- Affinity to life long learning
- Social and ethnic plurality
- Flexibility
- Creativity
- Cosmopolitanism/Open-mindedness
- Participation in public life

SMART GOVERNANCE (Participation)

- Participation in decision-making
- Public and social services
- Transparent governance
- *Political strategies & perspectives*

SMART MOBILITY (Transport and ICT)

- Local accessibility
- (Inter-)national accessibility
- Availability of ICT-infrastructure
- Sustainable, innovative and safe transport systems

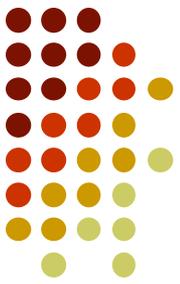
SMART ENVIRONMENT (Natural resources)

- Attractivity of natural conditions
- Pollution
- Environmental protection
- Sustainable resource management

SMART LIVING (Quality of life)

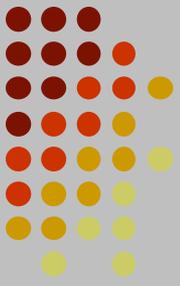
- Cultural facilities
- Health conditions
- Individual safety
- Housing quality
- Education facilities
- Touristic attractivity
- Social cohesion

Les systèmes de recommandation pour les Smart Cities



$u(v,a)$	Action 1	Action 2	Action 3	Action 4	Action 5
Ville 1	8	7			
Ville 2	9		3		
Ville 3	3	5		5	5
Ville 4	5	3		3	3

- Recommandations pour les Smart Cities:
- Etant donnés :
 - Un log de villes
 - Une ville qui veut améliorer son « intelligence » (courante)
- Soit
 - A : ensemble des actions possibles
 - V : ensemble de villes
 - u : fonction d'utilité d'une action a pour une ville v
- Alors choisir l'action a' non encore évaluée de A qui maximise l'utilité pour chaque ville v de V



Exemple - Hypothèses

Premièrement, nous restreignons notre espace de travail :

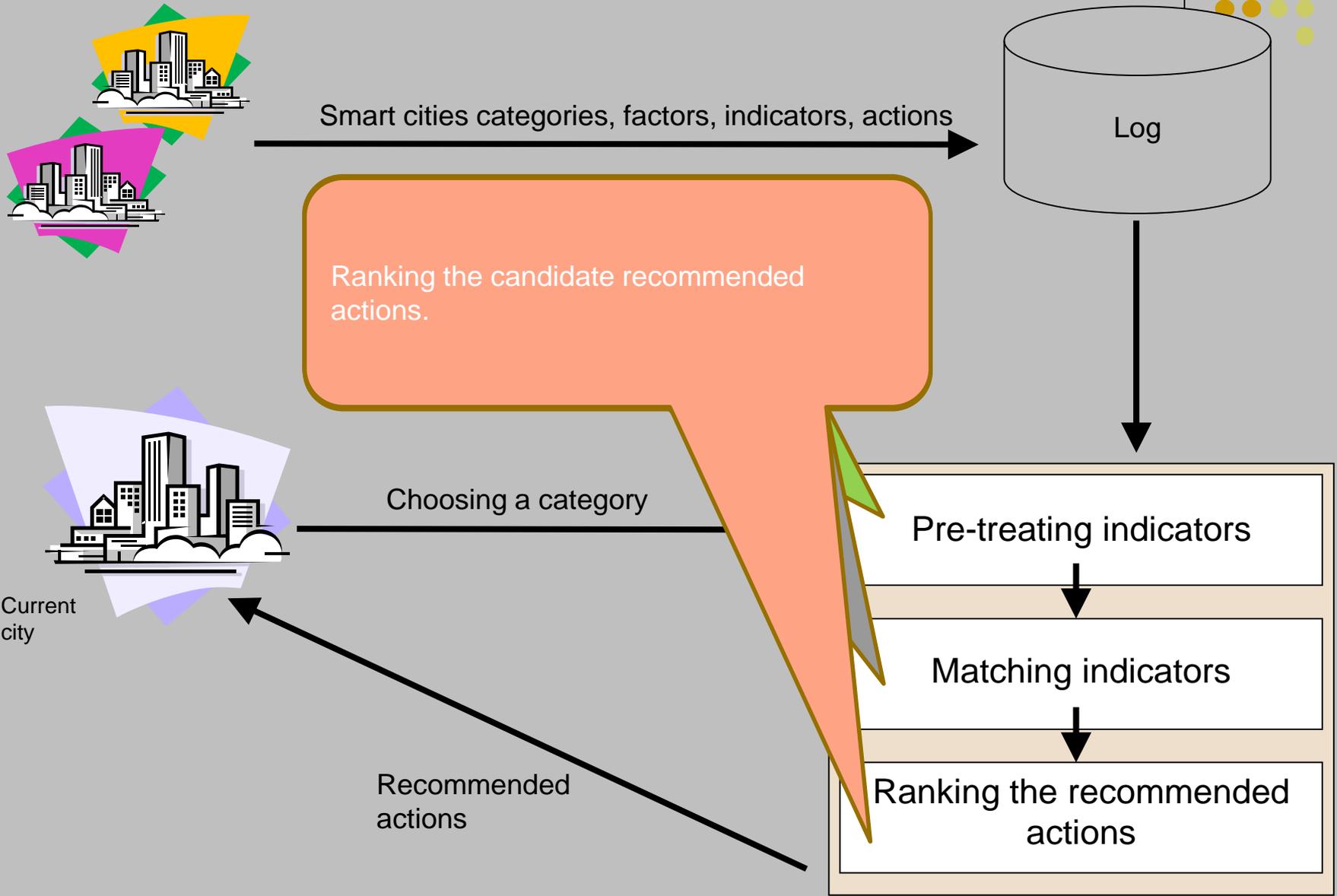
- Les notes / scores sont donnés par une personne autorisée à prendre la décision de mettre en œuvre des actions différentes et à travers ce score indique si l'action est (ou était) pertinente.
- Une ville peut être considérée comme un n-uplet contenant la description de la ville et les informations de la ville.
- Les informations de la ville sont un ensemble de catégories « intelligentes » où chaque catégorie est un ensemble de facteurs et chaque facteur est un 3-uplet spécifiant les indicateurs/valeurs correspondants et pour chaque indicateur, l'ensemble des actions mises en œuvre.
- Les données concernant les villes (catégories, facteurs, indicateurs, actions) sont enregistrées.
- Le log peut être une base de données ou une autre structure de données. Il sera alimenté par le système de recommandation au moment de l'utilisation du système par les villes.
- Pour les données initiales (problème de démarrage à froid), nous espérons utiliser des données officielles, publiques et libres, éventuellement enrichies avec la participation de volontaires.
- La ville, qui veut améliorer son « intelligence », choisit, au début, une seule catégorie de villes intelligentes (parmi les six) qu'elle souhaite rejoindre.
- Les valeurs des indicateurs sont numériques.
- Les indicateurs sont les mêmes pour une catégorie / un facteur donné pour chaque ville (seules les valeurs changent et peuvent être nulles).

So, we have, for each city $C_i, \forall i, j, k, n \in \mathbb{N}^{+*}$:

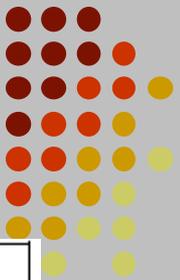
$$C_i = \langle Description_i, \{ Category_{ij}, \{ Factor_{ijk}, \{ \langle Indicator_{ijk1}, Value_{ijk1}, \{ Action_{ijk1}^1, \dots, Action_{ijk1}^{m_1} \} \rangle, \dots, \langle Indicator_{ijkn}, Value_{ijkn}, \{ Action_{ijkn}^1, \dots, Action_{ijkn}^{m_n} \} \} \} \} \rangle \rangle$$

where $\forall l \in \mathbb{N}^{+*}, m_l \in \mathbb{N}^+$.

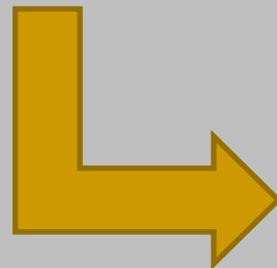
Example (2) - Overview



Exemple (3)

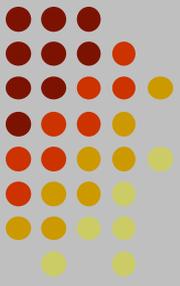


	smart environment					
	sustainable resource management				pollution	
	water consumption		electricity consumption		air quality	
	Value	Actions	Value	Actions	Value	Actions
Smallville	300 A1 A2	•after midnight shut down the public fountains, •do not water plants during summer	3 000 A4 A5	•after midnight shut down the public lights, •solar street lamps	$\frac{10}{10}$	\emptyset
Metropolis	100 000 A3	•after midnight shut down the public fountains, •do not wash cars during summer	200 000	•after midnight shut down the public lights	$\frac{8}{10}$ A6 A7	•restrict vehicles access into town center, •lower vehicles speed on big roads
Gotham	150 000	\emptyset	100 000	•after midnight shut down the public lights	$\frac{0}{10}$	\emptyset



u(c,a)	A1	A2	A3	A4	A5	A6	A7
Smallville	9	7		8	8		
Metropolis	8		6	7		5	5
Gotham				8			

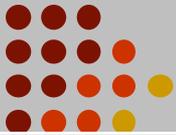
Exemple (4)



Step 1 : Pretreating indicators

- Input :
 - the log of smart cities (L)
 - the category (G) chosen by the current city.
- The Pretreat function is used to compute a set of intervals of indicator values. It searches among cities in L having factors and indicators corresponding to G, for each indicator, the minimal recorded indicator value (V_{min}) and the maximal recorded indicator value (V_{max}).
- Output : $[V_{min}, V_{max}]$

Exemple (5)



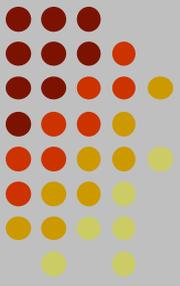
	smart environment					
	sustainable resource management				pollution	
	water consumption		electricity consumption		air quality	
	Value	Actions	Value	Actions	Value	Actions
Smallville	300	<ul style="list-style-type: none"> •after midnight shut down the public fountains, •do not water plants during summer 	3 000	<ul style="list-style-type: none"> •after midnight shut down the public lights, •solar street lamps 	$\frac{10}{10}$	∅
Metropolis	100 000	<ul style="list-style-type: none"> •after midnight shut down the public fountains, •do not wash cars during summer 	200 000	<ul style="list-style-type: none"> •after midnight shut down the public lights 	$\frac{8}{10}$	<ul style="list-style-type: none"> •restrict vehicles access into town center, •lower vehicles speed on big roads

- $L = \{\text{Smallville, Metropolis}\}$
- $G = \text{Smart environment}$



	smart environment		
	sustainable resource management		pollution
	water consumption	electricity consumption	air quality
Intervals	[300; 100 000]	[3 000; 200 000]	$[\frac{8}{10}; \frac{10}{10}]$

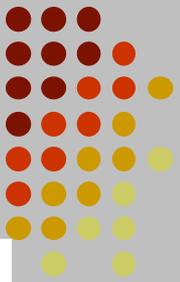
Exemple (6)



Step 2 : Matching indicators

- Input :
 - Intervals of indicator values (step 1)
 - Current city (C_c)
 - The log of cities (L)
- The Match function verifies if the indicator values of C_c belong to the intervals. If it does not belong, the set of corresponding actions in L is registered.
- Output : a set of unordered actions

Exemple (7)



smart environment			
sustainable resource management		pollution	
water consumption	electricity consumption	air quality	
Intervals	[300; 100 000]	[3 000; 200 000]	$[\frac{8}{10}; \frac{10}{10}]$
Gotham	150 000	100 000	$\frac{8}{10}$



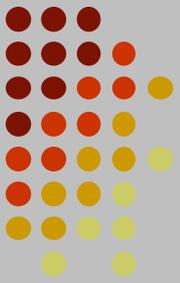
smart environment						
sustainable resource management				pollution		
water consumption		electricity consumption		air quality		
Value	Actions	Value	Actions	Value	Actions	
Smallville	300	3 000	<ul style="list-style-type: none"> •after midnight shut down the public fountains, •do not water plants during summer 	$\frac{10}{10}$	∅	
Metropolis	100 000	200 000	<ul style="list-style-type: none"> •after midnight shut down the public fountains, •do not wash cars during summer 	$\frac{8}{10}$	<ul style="list-style-type: none"> •restrict vehicles access into town center, •lower vehicles speed on big roads 	

A1

A2

A3

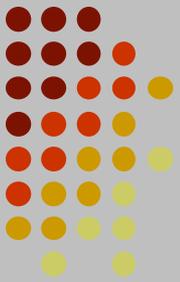
Exemple (8)



Step 3 : Ranking actions

- Input :
 - The unordered set of actions (step 2)
 - Satisfaction criterion
- The Rank function orders actions according to the satisfaction criterion (expressed by the city).
- Output : a set of ordered actions

Exemple (9)



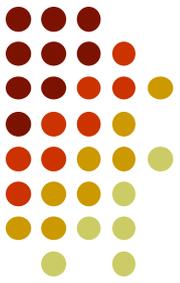
u(c,a)	A ₁	A ₂	A ₃	A ₄	A ₅	A ₆	A ₇
Smallville	9	7		8	8		
Metropolis	8		6	7		5	5
Gotham				8			



u(c,a)	A ₁	A ₂	A ₃	A ₄	A ₅	A ₆	A ₇
Smallville	9	7		8	8		
Metropolis	8		6	7		5	5
Gotham	9	7	6	8			

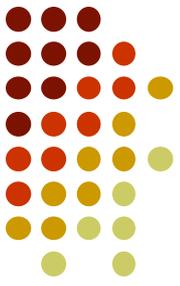
- Recommended actions are the ordered set {A₁, A₂, A₃}
- A₁ : *after midnight shut down the public fountains*
- A₂ : *do not water plants during summer*
- A₃ : *do not wash cars during summer*

Plan



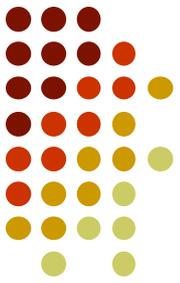
- Qu'est ce qu'un système de recommandations ?
- Les méthodes les plus utilisées
- Les améliorations possibles
- Un petit mot sur l'évaluation
- **Applications**
 - Gestion de crises - Systèmes d'alertes précoces
 - Smart Cities
 - Ailleurs...
- Conclusion et perspectives

Ailleurs...



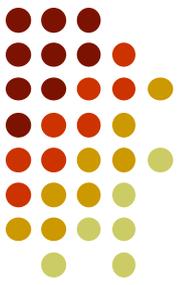
- Entrepôts de données / OLAP
- Recettes de cuisine
- Experts dans les Environnements collaboratifs de travail (CWE) à partir de traces d'interaction
- Supports d'apprentissage dans les MOOC (avec des agents)
- ...

Plan



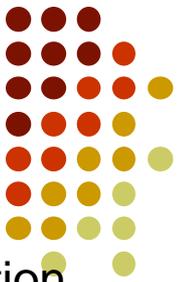
- Qu'est ce qu'un système de recommandations ?
- Les méthodes les plus utilisées
- Les améliorations possibles
- Un petit mot sur l'évaluation
- Applications
 - Gestion de crises - Systèmes d'alertes précoces
 - Smart Cities
 - Ailleurs...
- Conclusion et perspectives

Conclusion



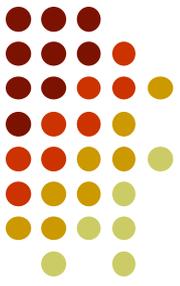
- Les systèmes de recommandation ont leurs racines dans de nombreux domaines de recherche :
 - Recherche d'information
 - Filtrage d'information
 - Classification de texte
 - ...
- Ils utilisent des techniques issues de différents domaines :
 - Machine learning
 - Data mining
 - ...
- abordent différents sujets :
 - Algorithmes de recommandation
 - Approches collaboratives, basées sur le contenu, hybrides, ...
 - Evaluation des RS
 - Contexte
 - ...

Perspectives



- Amélioration des techniques de filtrage collaboratif
 - Utiliser plus de sources de données (tagging data, demographic information, and time data)
 - Combiner des techniques
 - ...
- Multicriteria recommender systems
 - Exploiter les scores multicritères qui contiennent de l'information contextuelle, pour améliorer les recommandations
- Context awareness
 - Prendre en compte les aspects temporel, géographiques, ...
 - Contexte émotionnel ("Je suis tombée amoureuse d'un garçon. Je veux voir un film romantique.")
 - ...
- Group recommendations
- Les techniques de recommandation vont fusionner avec d'autres champs de recherche
 - User modeling
 - Personalized reasoning
 - ...
- Cold-start
- Context-Aware recommender systems → Context-Driven recommender systems

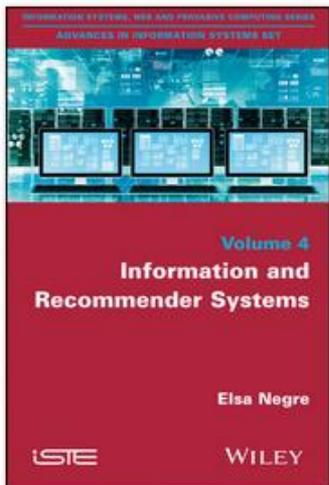
Implications non-négligeables pour les données environnementales



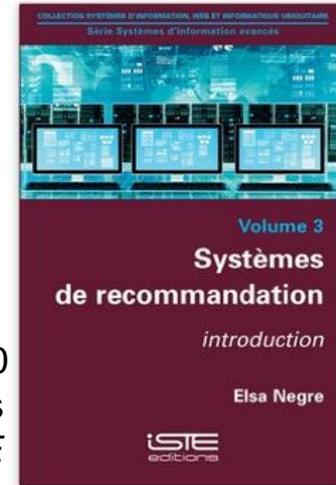
Questions ?

Information and Recommender Systems

Systèmes de recommandation



ISBN: 978-1-84821-754-6
92 pages
October 2015, Wiley-ISTE



ISBN : 978-1-78405-863-0
82 pages
Juin 2015, ISTE