



WMMFB40: Traitement des données temps réel et données hétérogènes

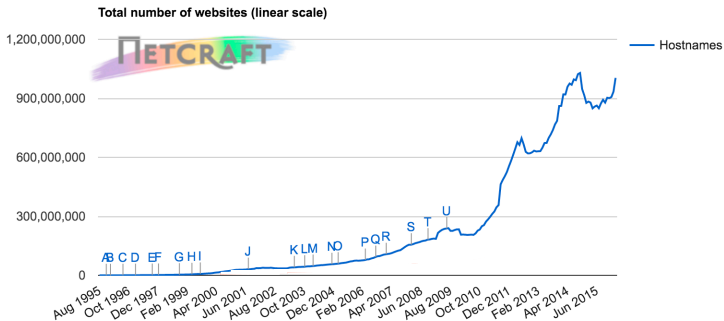
Massih-Reza Amini

Université Grenoble Alpes
Laboratoire d'Informatique de Grenoble
Massih-Reza.Amini@imag.fr

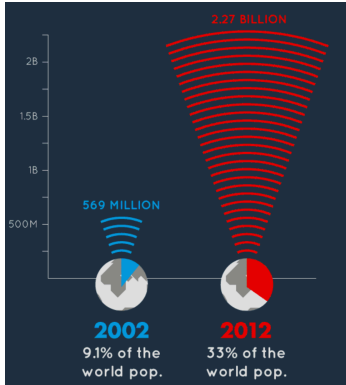


Ère de BigData

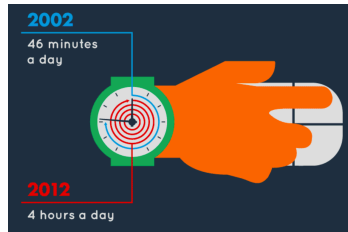
- In great part due to the rapid development of the Web this last 20 years,



Ère de BigData: Nouvelles pratiques et habitudes

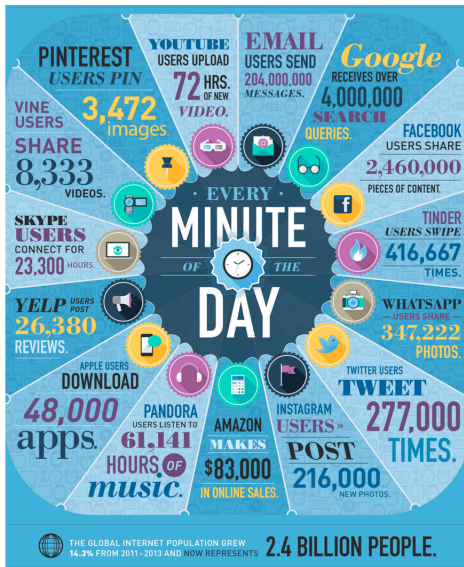


Nombre d'internautes



Temps de connexion

Ère de BigData: Génération accrue des données



Ère de BigData: Valeur des données

- ❑ According to the EMC project¹, in 2020 there will be 40 zetta bytes (40×10^{21} bytes) of unstructured data on the web.
- ❑ These data are considered as the oil of the *XXI* century.²
- ❑ Need to develop new automatic tools for information access.

¹ <http://www.emc.com/leadership/digital-universe/index.htm>

² [http:](http://www.lepoint.fr/technologie/les-data-petrole-du-xxie-siecle-14-03-2012-1441346_58.php)

[//www.lepoint.fr/technologie/les-data-petrole-du-xxie-siecle-14-03-2012-1441346_58.php](http://www.lepoint.fr/technologie/les-data-petrole-du-xxie-siecle-14-03-2012-1441346_58.php)

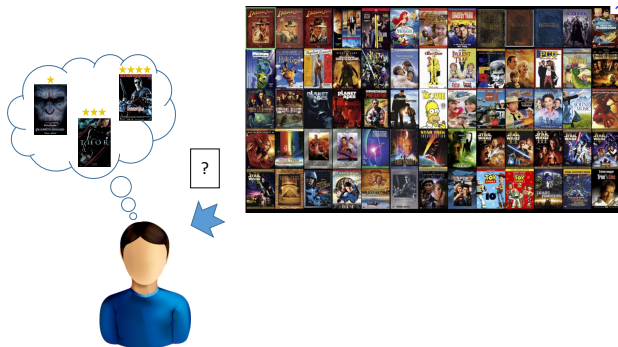
Ère de BigData: Ère de découverte

- ❑ Chris Anderson, l'éditeur en chef de *Wired Magazine* a publié dans son article Long Tail (Longue Traîne)
 - ❑ *"We are leaving the Information Age and entering the Recommendation Age."*

- ❑ Dans CNN Money, "The race to create a smart Google"
 - ❑ The Web, they say, is leaving the era of search and entering one of discovery. What's the difference? Search is what you do when you're looking for something. Discovery is when something wonderful that you didn't know existed, or didn't know how to ask for, finds you.

Systèmes de Recommandation

- ❑ Sur la base des comportements (passés), un SR prédit l'appétence d'un utilisateur pour un produit.
- ❑ Le but étant de recommander de nouvelles choses qui ne sont pas vues auparavant par les gens.



Systèmes de Recommandation

- ❑ Sur la base des comportements (passés), un SR prédit l'appétence d'un utilisateur pour un produit.
- ❑ Le but étant de recommander de nouvelles choses qui ne sont pas vues auparavant par les gens.

Plus de ...

- ❑ 60% des films vus sur Netflix sont des films recommandés,
- ❑ 35% of sales on Amazon are through recommendation,
- ❑ 38% of clicks on Google are generated over recommended products.



Définition d'un système de recommandation

- Un système de recommandation peut être défini comme la recherche d'une fonction qui prédit automatiquement l'appétence d'un utilisateur pour un produit.

- Ce calcul est souvent basé sur
 - Le comportement de l'utilisateur
 - Ses relations avec d'autres utilisateurs
 - Le contexte,
 - etc.

Différents types de recommandation

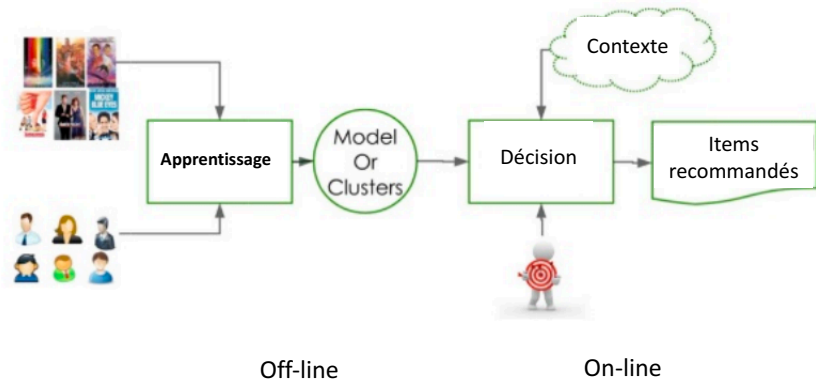
- ❑ Filtrage collaboratif basé sur les utilisateurs
 - ❑ Faire des recommandations en fonction de la combinaison de l'expérience de chaque utilisateur avec celles des autres
 - ❑ À partir de la matrice de utilisateur (en ligne) vs. produits (en colonne); calculer des scores de similarité entre les utilisateurs
 - ❑ Trouver les utilisateurs similaires et recommander à un utilisateur des produits choisis par les autres de son groupe, qu'il n'a pas encore choisi.
- ❑ Filtrage collaboratif basé sur les produits
 - ❑ À partir de la matrice de utilisateur (en ligne) vs. produits (en colonne); calculer des scores de similarité entre les produits
 - ❑ Recommander à un utilisateur des produits dont les similarités avec les produits qu'il a choisis sont élevés.

Définition plus formelle

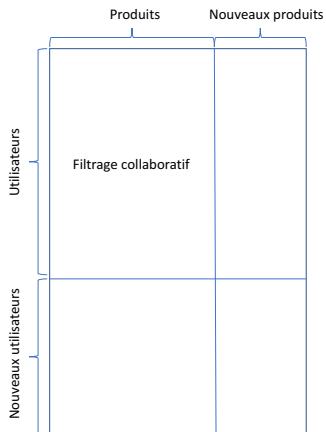
- Soit \mathcal{U} l'ensemble de tous les utilisateurs, et \mathcal{I} l'ensemble de tous les items possibles pour la recommandation
- soit f la fonction d'utilité qui mesure l'utilité de l'item $i \in \mathcal{I}$ pour l'utilisateur $u \in \mathcal{U}$, i.e. $f : \mathcal{U} \times \mathcal{I} \rightarrow \mathbb{R}_+$
- Le but est ainsi de choisir pour chaque utilisateur $u \in \mathcal{U}$, l'item $i \in \mathcal{I}$ qui maximise la fonction f :

$$u \in \mathcal{U}, i^* = \operatorname{argmax}_{i \in \mathcal{I}} f(u, i)$$

Processus en deux étapes



Différentes situations



Apprentissage d'un espace latent commun

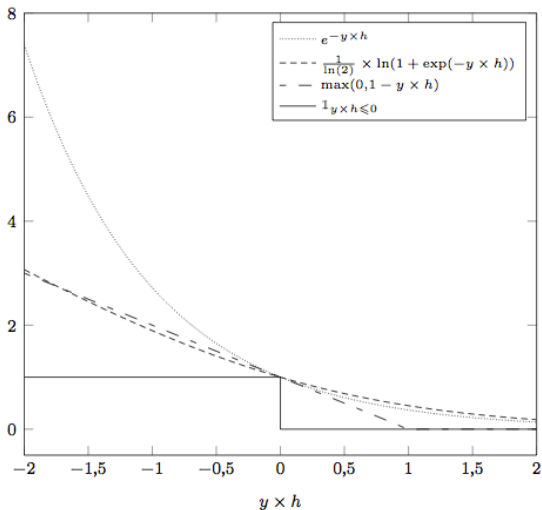
- ❑ Prix de 2M\$ de NetFlix: arriver à améliorer de 10% les performances dans le sens de la distance au moindre carré
- ❑ Système gagnant à base de l'apprentissage et de la factorisation matricielle $A = \mathbf{U}\mathbf{V}^\top$ En résolvant le problème

$$\operatorname{argmin} R_m(A, \mathbf{U}, \mathbf{V}) = \sum_{r_{ui}} \mathbb{1}_{r_{ui} \neq \mathbf{U}_u \mathbf{V}_i^\top} + \lambda(\|\mathbf{U}_u\|^2 + \|\mathbf{V}_i\|^2)$$

Or la minimisation de l'erreur de classification empirique est impossible ...

- ❑ La fonction empirique R_m n'est ni continue, ni dérivable.
- ❑ La minimisation d'une borne supérieure convexe $\hat{\mathcal{L}}(\mathbf{w})$ de R_m prenant la valeur 1 en 0 conduira à la même minimiseur que R_m .

Bornes supérieures convexes classiques de l'erreur



Optimisation convexe non-contrainte

- ❑ Le problème d'apprentissage se définit comme un problème d'optimisation non-contraintes plus facile.
- ❑ Considérer l'expansion de Taylor de la fonction objectif autour de son minimiseur

$$\hat{\mathcal{L}}(\mathbf{w}) = \hat{\mathcal{L}}(\mathbf{w}^*) + (\mathbf{w} - \mathbf{w}^*)^\top \underbrace{\nabla \hat{\mathcal{L}}(\mathbf{w}^*)}_{=0} + \frac{1}{2} (\mathbf{w} - \mathbf{w}^*)^\top \mathbf{H} (\mathbf{w} - \mathbf{w}^*) + o(\|\mathbf{w} - \mathbf{w}^*\|^2)$$

- ❑ La matrice Hessienne est symétrique et d'après le théorème de Schwarz ses vecteurs propres $(\mathbf{v}_i)_{i=1}^d$ forment une base orthonormée.

$$\forall (i, j) \in \{1, \dots, d\}^2, \mathbf{H}\mathbf{v}_i = \lambda_i \mathbf{v}_i, \text{ et } \mathbf{v}_i^\top \mathbf{v}_j = \begin{cases} +1 & : \text{si } i = j, \\ 0 & : \text{otherwise.} \end{cases}$$

Optimisation convexe non-contrainte (2)

- Chaque vecteur de poids $\mathbf{w} - \mathbf{w}^*$ peut être décomposée de façon unique sur cette base

$$\mathbf{w} - \mathbf{w}^* = \sum_{i=1}^d q_i \mathbf{v}_i$$

- C'est à dire

$$\hat{\mathcal{L}}(\mathbf{w}) = \hat{\mathcal{L}}(\mathbf{w}^*) + \frac{1}{2} \sum_{i=1}^d \lambda_i q_i^2$$

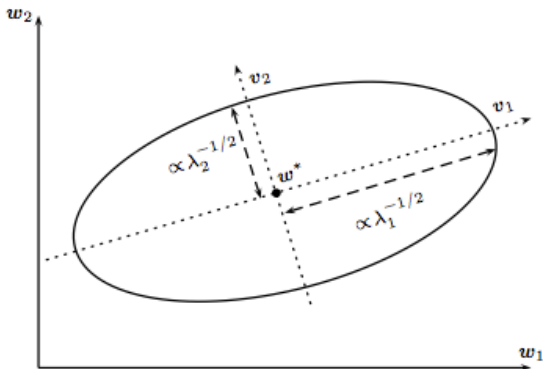
- De plus, comme la fonction objectif admet un minimum global, la matrice Hessienne est définie positive

$$(\mathbf{w} - \mathbf{w}^*)^\top \mathbf{H}(\mathbf{w} - \mathbf{w}^*) = \sum_{i=1}^d \lambda_i q_i^2 = 2(\hat{\mathcal{L}}(\mathbf{w}) - \hat{\mathcal{L}}(\mathbf{w}^*)) \geq 0$$

Toutes les valeurs propres de \mathbf{H} sont positives.

Optimisation convexe non-contrainte (3)

- Ceci implique que les lignes de niveau de $\hat{\mathcal{L}}$, définies par des lignes à valeur constante, sont des ellipses

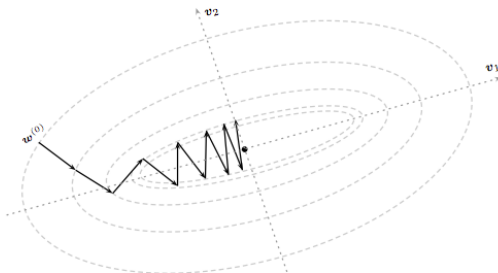


Algorithme de descente de Gradient

- L'algorithme de descente de gradient est un procédé itératif de la mise à jour des poids :

$$\forall t \in \mathbb{N}, \mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} - \eta \nabla \hat{\mathcal{L}}(\mathbf{w}^{(t)})$$

Où $\eta > 0$ est le pas d'apprentissage



References



A.P. Dempster, N.M. Laird, D.B. Rubin
Discriminant Analysis and Statistical Pattern Recognition.
Journal of the Royal Statistical Society, Series B, 39(1):1–38,
1977.



G. Celeux, G. Govaert.
A classification EM algorithm for clustering and two stochastic versions.
Research Report-1364, INRIA, 1991.



G.J. McLachlan
Discriminant Analysis and Statistical Pattern Recognition.
Wiley Interscience,
1992.



J. B. MacQueen
Some Methods for classification and Analysis of Multivariate Observations,
Proceedings of 5th Berkeley Symposium on Mathematical Statistics
and Probability, (1): 281–297,
1967