

SEANCE DU 08 décembre 2015.

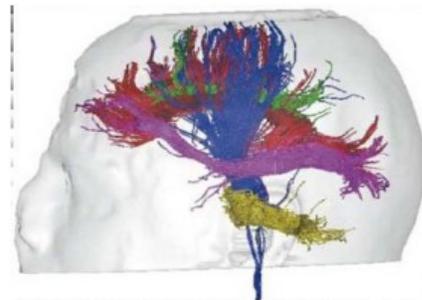
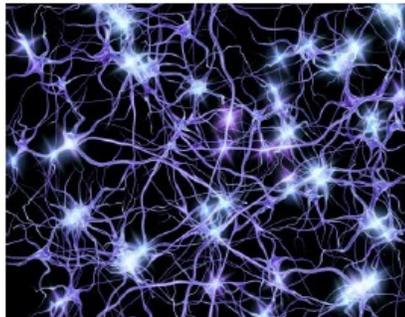
Restitution de l'intervention de :
Georges Linarès et Didier Josselin

Par l'équipe d'auditeurs : Barbara, Camille, Joëlle, Michèle, André, Gilles et Roland

TITRE : Informatique, mémoire et oubli.
Deuxième partie

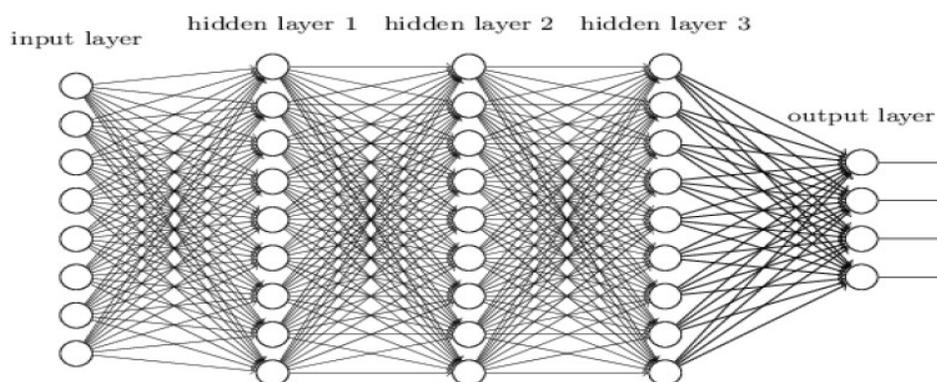
Georges Linarès : On va passer par des choses qui sont bien plus loin des aspects technologiques, très liés en tous cas aux bases technologiques de la mémoire en informatique. Ce que l'on a vu jusqu'à présent c'est que c'était quand même des principes de mémorisation basés sur l'idée de la commode à tiroirs dont je vous avais parlé tout à l'heure. Mais ce que l'on a vu aussi, c'est qu'un cerveau biologique, ce n'est pas seulement vrai pour les humains, ça ne fonctionne pas du tout pareil. D'abord la mémoire a une fonction globale, c'est-à-dire que l'on a un accès par le contenu. Les informations qu'on mémorise sont dépendantes les unes des autres. C'est ce qui permet aussi de faire de la reconstruction.

La mémoire comme fonction globale : les réseaux de neurones



Les gens qui font de l'informatique se sont fait cette remarque, ils ont considéré quel était le support de la mémoire chez l'homme, dans un cerveau biologique. Et dans un cerveau biologique on a des mécanismes qui sont dits partiels qui sont nécessaires, on a une quantité de mémoire limitée et le besoin de mémoriser tout le temps des informations et des processus de reconstruction et il y a toute la plasticité et la capacité d'adaptation du cerveau. La capacité d'adaptation du cerveau repose sur l'apprentissage, la mémorisation de nouvelles informations, de nouvelles règles. Et forcément la disparition d'autres informations. Se disant ça, les informaticiens ont cherché à construire des cerveaux artificiels.

La mémoire comme fonction globale : les réseaux de neurones



Alors il faut faire attention à ça, ce côté très anthropomorphique, c'est-à-dire essayer de faire construire des machines qui fonctionneraient exactement comme un cerveau humain c'est un peu compliqué. Notamment parce que l'on ne sait pas très bien comment le cerveau fonctionne. Donc essayer de reproduire quelque chose que l'on ne connaît pas ou que l'on connaît mal, forcément ça a des limites. Mais on s'est quand même inspiré de la façon dont les cerveaux sont structurés et organisés globalement, sont construits au niveau du neurone et des liens entre les neurones que sont les synapses, et évidemment les mécanismes d'apprentissage.

Comment un cerveau apprend ? Apprendre c'est mémoriser de l'information, pas seulement mémoriser sous la forme directe, c'est aussi avoir des mécanismes d'abstraction. Essayer d'avoir des représentations qui doivent être au plus haut niveau dans le sens qu'elles doivent être plus loin que l'observation et qui sont plutôt sur les règles, les mécanismes que l'on va trouver au-dessus de ce qui est observé.

Le réseau de neurones codé dans l'informatique, en gros c'est des ensembles de neurones, on les organise par couche. Tous les modèles ne sont pas comme ça, mais tous les modèles classiques sont comme ça. On code dans la mémoire de notre ordinateur des structures qui sont composées de neurones et des liens entre ces neurones. En général ces réseaux ce sont des réseaux qui ont des capacités d'apprentissage. Par exemple on présente un stimulus, une donnée observée à l'entrée du réseau et on s'attend à ce que le réseau produise une réponse en sortie qui correspond à ce que l'on attend. Le lien entre l'entrée et la sortie on le fait en apprenant, en donnant au réseau plein d'exemples. Je crois que Didier va dérouler un exemple en géographie, mais c'est à appliquer à pleins de domaines. Si vous avez un *smartphone* ; dans le système de reconnaissance de la parole par exemple, on a des systèmes d'identification du contenu d'une image, derrière on a des réseaux de neurones.

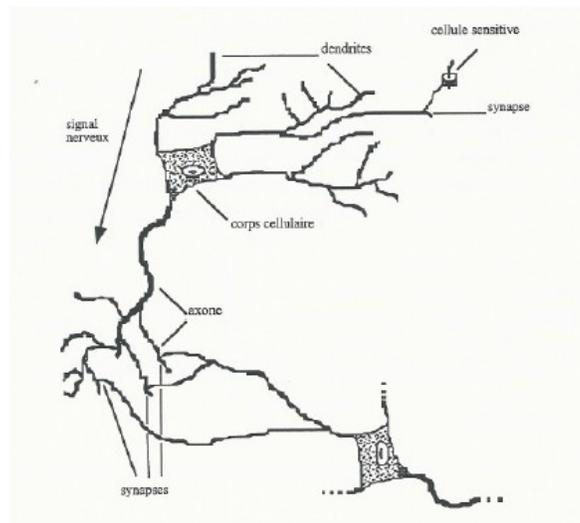
Je passe la parole à Didier.

Didier Josselin : Si je lance successivement 2 morceaux de craie à Jean-Robert, le premier il ne le rattrape pas parce qu'il est surpris, normalement le second il doit s'en saisir. Alors pourquoi ? Le réseau de neurones en fait est excité et avec ce genre d'approche, on essaie de reproduire le fait que la deuxième fois le neurone de Jean-Robert aura été excité et du coup il y a un poids qui s'est mis sur certaines fonctions à l'intérieur de ses neurones. Quand on regarde la structure que Georges a présentée tout à l'heure, c'est une structure avec des entrées et des sorties, des tubes, mais on sait que la gaine de myéline transporte l'information, vous avez des axones, vous avez des neurotransmetteurs, qui accélèrent ou au contraire ralentissent le flux nerveux. Tout ça n'est pas si simple.

Donc en informatique on est quand même loin d'une analogie, mais il faut savoir que l'on a une capacité des outils en informatique à modéliser des systèmes de décision en mimant, en quelque sorte, les réseaux de neurones.

Moi je me suis heurté il y a un certain nombre d'années à une question. En fonction de l'évolution du sol sur des parcelles agricoles dans le Vercors, sur 8 communes, environ 1000 parcelles, j'avais tout un tas de descripteurs géographiques à la parcelle, à l'exploitation agricole, aussi des éléments géophysiques, il fallait que j'essaie de comprendre, de prévoir éventuellement, dans 15 ans comment serait le paysage en fonction de l'évolution passée des 25 dernières années.

La mémoire des réseaux de neurones pour estimer la déprise agricole



Ce que j'ai fait, j'ai constitué un réseau de neurones et je lui ai demandé d'apprendre. D'apprendre ce qu'il s'est passé dans ces données, cet espèce d'échantillon que j'avais et qui, sur une durée de 25 ans, ce qui représentait une série d'évolutions des parcelles. Donc j'avais un certain nombre de variables, par exemple, le mode d'utilisation de la parcelle ou aussi l'évolution entre les deux périodes de temps que j'avais, voir si les parcelles avaient été abandonnées ou au contraire si elles s'étaient renforcées dans leur usage, en fonction par exemple de jeunes qui s'installaient, ou des choses comme ça.

J'avais un échantillon spatial, tout dans le Vercors, et j'avais différentes communes qui avaient été choisies par les gens avec qui je travaillais, pour être un peu représentatives de l'agriculture de montagne du secteur. J'avais d'autres variables qui allaient expliquer d'autres entrées, des variables d'entrée comme on avait tout à l'heure dans le réseau de neurones, qui étaient par exemple : si l'agriculture avait des aides, quelle était la structure des exploitations, cela est-il structurant dans l'exploitation agricole ? Ensuite est-ce qu'on a les circuits de fonctionnement des ventes qui fonctionnent bien ?

Variables explicatives de l'exploitation agricole

- Thème "Exploitation-structure" :
- * ATE : atelier(s) de production de l'exploitation,
 - * SAU : SAU de l'exploitation,
 - * BOV : proportions des bovins dans le cheptel,
 - * VLA : proportion des vaches laitières dans le cheptel,
 - * QUO : position de l'agriculteur vis-à-vis de ses quotas laitiers,
 - * COM : mode de commercialisation dominant des produits;
- Thème "Exploitant-aides" :
- * CGI : aide du Conseil Général de l'Isère,
 - * AGE : âge de l'exploitant,
 - * SUC : succession présumée,
 - * AID : nombre d'aides reçues par l'agriculteur,
 - * MOE : main d'oeuvre permanente disponible sur l'exploitation,
 - * REG : régime d'activité de l'exploitant.

Fig. III.3 - Les thèmes concernant l'exploitation

- Thème "Valeur-aménagement" :
- * BIO : "valeur écologique" (ZNIEFF, réserve de chasse, etc.),
 - * PAY : "valeur paysagère" (d'après jurys communaux),
 - * AME : aménagements réalisés depuis 10 ans sur la parcelle (accès, eau, clôture, etc.),
 - * PRO : procédures locales (remembrement, APF, OGAFF, etc.),
 - * POS : catégorie au Plan d'Occupation des Soils,
 - * REV : valeur locative annuelle théorique (revenu cadastral).
- Thème "Contraintes-risques" :
- * FEN : nuisances moyennes sur la parcelle,
 - * SUP : superficie de la parcelle,
 - * ACC : type d'accès à la parcelle,
 - * RIS : risques sur la parcelle,
 - * DIS : distance séparant la parcelle du siège d'exploitation;
- Thème "Entretiens-perception" :
- * ENT : le cas échéant, niveau d'entretien de l'herbe,
 - * TRL : taux de recouvrement ligneux,
 - * VOI : environnement dominant de la parcelle (en terme d'occupation du sol),
 - * LIS : état global de la lisière bordant la parcelle,
 - * JUS : justification principale de l'entretien de la parcelle,
 - * PLA : place de la parcelle dans le système,
 - * FCT : fonction principale de la parcelle;
- Thème "Foncier-propriétaires" :
- * LEP : lien de parenté entre l'ancien exploitant et l'actuel,
 - * DEV : devenir de l'ancien exploitant,
 - * FVL : type de fais-valoir sur la parcelle,
 - * PRP : nature de la propriété,
 - * PRF : profession du propriétaire,
 - * RES : lieu de résidence du propriétaire;
- Thème "Physique-biologie" :
- * ALT : altitude moyenne,
 - * ORI : orientation,
 - * SOL : nature du sol,
 - * PFF : "pierrosité" et profondeur du sol,
 - * VEG : durée de la végétation,
 - * HYD : hygrométrie de la station;
- Thème "Production-pâturages" :
- * CHE : cheptel directement utilisateur de la parcelle (si pâturage),
 - * CHO : estimation de la charge des animaux sur la parcelle,
 - * EQP : matériel utilisé pour l'exploitation de la parcelle,
 - * VOL : estimation du volume de production des parcelles fauchées,
 - * RGA : force du regain de l'herbe sur la parcelle,
 - * FER : type de fertilisation sur la parcelle.

Fig. III.4 - Les thèmes attachés à la parcelles

Variables explicatives à la parcelle

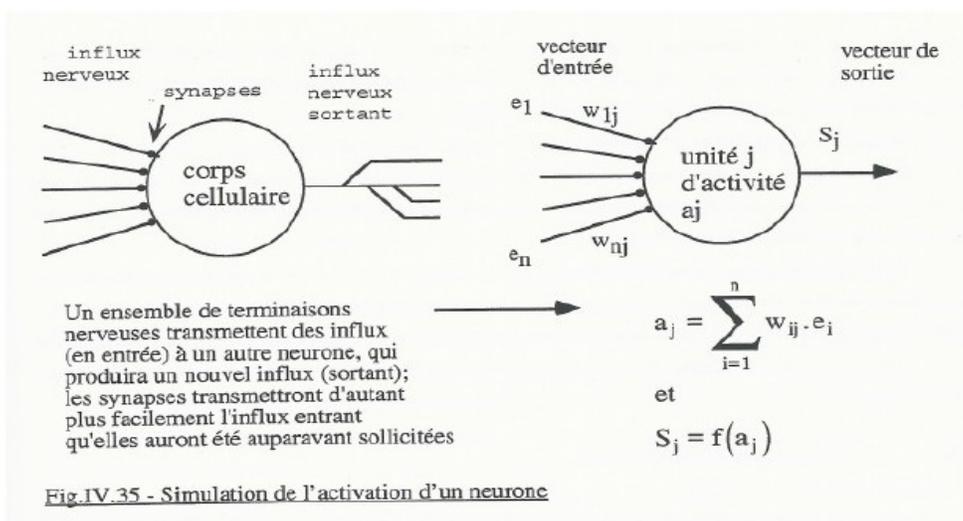
Et puis on avait tout un tas d'explications qui étaient liées à la parcelle agricole. Donc là, on change de résolution, on avait des données sur l'exploitation et puis on désagrège, on devient plus précis. On a par exemple des éléments sur l'aménagement. Est-ce que l'on est dans des zones écologiques ? Quels sont les risques, est-ce qu'il y a des risques sur la parcelle ? La distance de la parcelle à la zone d'exploitation, donc vous voyez tout ce qui est dans la structure parcellaire qui compose cette exploitation agricole.

On a aussi des éléments sur la perception de l'entretien. C'est-à-dire, finalement, cette parcelle, qu'elle est sa place dans l'exploitation, si il y a des rotations de cultures, des choses comme ça. Au niveau des éléments sur le foncier : quelle est la nature de la propriété ? Parce qu'il y a plus de risque d'abandon lorsque l'utilisateur, l'agriculteur n'est pas propriétaire de la parcelle, il y a tout un tas de critères qui font que c'est plus ou moins valable sur le long terme d'avoir une exploitation selon qu'il possède les parcelles ou pas. Après il y a des contraintes physiques, de biologie, "la pierrosité", la difficulté d'exploiter. Également des pratiques, partage des outils agricoles, des choses comme ça.

Et donc on a modélisé ça dans un réseau de neurones. Ça peut vous paraître bizarre mais en fait c'était ma thèse il y a un certain nombre d'années, et j'ai été un des premiers en géographie, à faire ça, à utiliser en quelque sorte un bio-mimétisme, en modélisant ce problème qui est un problème vraiment humain, social, avec un réseau de neurones.

Le principe est simple, c'est le principe de la craie de tout à l'heure, c'est le poids qui va se mettre à l'intérieur du réseau de neurones. C'est à travers ce poids qu'ensuite en quelque sorte, vous allez rediriger le flux dans le tube, et vous allez prendre une décision qui va être séparée entre des tubes de sortie, je simplifie, mais l'idée c'est qu'au niveau mathématique, vous avez ce que l'on appelle une rétro-propagation. C'est-à-dire qu'en gros, vous avez une sorte de calcul de ressemblance entre les différents critères qui rentrent dans le tube. A la sortie vous avez une classification et au fur à mesure ça devient discriminant, vous avez un système de rétro-propagation qui va jouer sur ces poids et qui va renforcer la sélectivité du réseau de neurones.

Analogie biologie et informatique



L'idée c'est ça, c'est une excitabilité interne de réseaux de neurones qui sont interconnectés. Et ce qui est extraordinaire là dedans, c'est que c'est super puissant mais que par contre on ne sait pas comment ça marche à l'intérieur. Il y a des publications de statisticiens qui montrent qu'il y a des analogies entre les "perceptrons", c'est-à-dire ce que Georges a évoqué tout à l'heure, avec une seule couche, un "perceptron" monocouche, c'est donc de la régression normale de base. Après vous avez la régression multiple qui correspond au multicouches, et vous avez par exemple les cartes de Kohonen qui sont d'autres réseaux de neurones qui correspondent à des nœuds dynamiques, à algorithmes basés sur des cas mixtes donc il y a des travaux qui existent, qui montrent que la statistique a déjà travaillé là dedans, donc on n'est pas loin, quelque part, d'avoir réinventé "la roue" d'après les statisticiens et quand on discute avec les informaticiens, ils disent que non, on est allé beaucoup plus loin. Et c'est vrai, quand on regarde dans la littérature on sent qu'il y a un renouveau de ces méthodes qui vraiment explosent au niveau des algorithmes et des possibilités.

Donc un réseau monocouche, ce que j'appelle un "perceptron" avec une entrée et une sortie, c'est un réseau un peu simple, alors qu'avec un réseau multicouches vous avez plusieurs possibilités et à la fin, quand vous creusez à l'intérieur, quand vous regardez la façon dont sont répartis les poids, c'est très difficile de trouver un déterminisme qui fait que vous allez avoir une bonne classification. Pourtant vous allez voir qu'en géographie je l'ai utilisé avec de bons résultats.

Types de réseaux de neurones Perceptron multi-couche

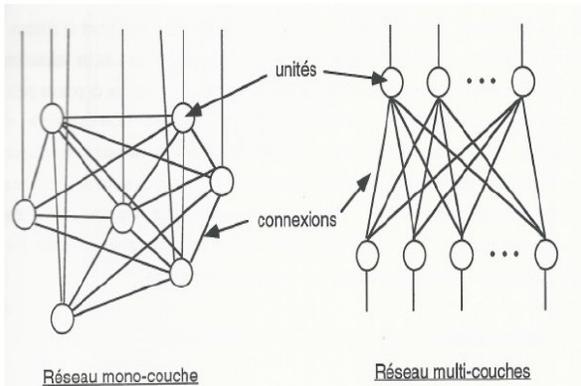


Fig. IV.37 - Unités et couches dans les réseaux de neurones

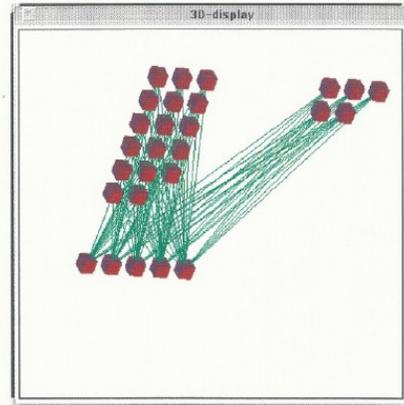


Fig. IV.38 - Structure du réseau de neurones utilisé
En haut à gauche, on trouve l'ensemble des unités d'entrées (correspondant aux 17 modalités des 4 variables "explicatives"), les unités de la couche cachée (5 en tout) et les unités de sortie (correspondant aux 5 modalités de la variable "décision").

Pour mes parcelles agricoles, j'avais 5 possibilités que nous allons détailler maintenant. La première possibilité c'était la forte déprise agricole. C'est-à-dire que les parcelles par rapport aux profils dominants de la combinaison des critères, ces parcelles-là auraient plutôt tendance à être en forte déprise. La probabilité qu'elles soient abandonnées est très forte. Sur la commune de Malleval, il y a beaucoup de forêts et il y a plus de nouvelles parcelles parce qu'en probabilité elles sont moins orientées vers la déprise. Tout cela n'est pas un tube qui prend des décisions très nettes, tout à l'heure quand j'ai jeté le morceau de craie à Jean-Robert, il a fait un mouvement esthétique, il n'a pas fait un mouvement de robot. C'est ça qui est extraordinaire avec les réseaux de neurones, c'est que vous avez une espèce de panoplie des possibles. Dans les littératures il y a la théorie des probabilités mais vous avez aussi la théorie des possibilités. Là on est dans les possibilités, on retrouve une parcelle qui se retrouve dans plusieurs catégories, avec une certaine probabilité d'arriver. Donc c'est très riche et très souple, c'est ça l'intérêt et c'est là que l'on voit une différence fondamentale avec d'autres méthodes qui seraient plus cassantes et qui feraient des partitions de solution, et là on a des superpositions des sorties.

Question de la salle : Si j'ai bien compris les zones rouges sont pratiquement destinées à être abandonnées ?

Réponse D.J : Ce que j'ai oublié de vous dire c'est que j'ai appris au réseau de neurones un peu comme tout à l'heure. Là il a été excité par l'ensemble des cinq cents à mille parcelles que j'avais sélectionné. Après je lui ai donné une commune, qui est Malleval, qu'il ne connaissait pas ou alors qui était dans la liste mais je le lui donne seul. On regarde comment cette commune va être reclassée par le système qui a appris. C'est-à-dire que j'ai stocké la mémoire en quelques sortes. Dans mon échantillon qui était représentatif, je me suis arrangé pour qu'il soit représentatif d'une variété spatiale des agricultures de montagne du Vercors. Là j'obtiens une espèce de mémorisation de la relation entre des variables explicatives et la probabilité de dépris ou pas. C'est cet outil là qui est une espèce de boîte dans laquelle je rentre n'importe quelle commune. Je fais l'hypothèse que son profil est cohérent avec celui de mon système qui va faire la classification et à la sortie elle va se mettre dans telle ou telle classe. C'est ce que j'ai fait là au niveau de la parcelle. J'ai injecté l'ensemble des critères. Ce sont les mêmes critères qu'on a vu, sur lesquels j'ai appris, j'ai des zéros pris pour la commune de Malleval, qui est une commune top pour faire du ski de fond. C'est aussi la commune de la secte apocalyptique qui s'est fait sauter la tête il y a plusieurs années vous vous souvenez. Donc à certains moments il ne faut pas y aller. Mais ceci dit c'est une commune qui est connue pour différentes raisons. Le temple du soleil c'était là à côté des pistes de ski....

Dans le principe, j'ai illustré en fait l'idée de mémoire. Ce système qui vous paraît peut être artificiel mais c'est de l'intelligence artificielle. On peut l'utiliser pour faire de la classification en science sociale. Donc vous voyez comme il y a une perméabilité entre les disciplines ce qui fait d'ailleurs que je suis actuellement dans le laboratoire de Georges Linarès. J'apprends beaucoup avec les méthodes des sciences exactes. On essaie d'apporter, nous sciences sociales, un regard un peu plus appliqué sur des problématiques environnementales. Par exemple, là nous voyons le cas d'emprise où dans la plus part des parcelles il y a stabilité c'est-à-dire que les profils qui sont repérés à Malleval font que la majorité des parcelles n'ont pas bougé dans le modèle. Si on fait l'hypothèse que ce modèle de prévision, entre guillemet, de vingt cinq ans si on le reproduit pour le futur, ce qui entre parenthèses serait faux puisqu'on ne sait pas de quoi sera fait l'avenir, on pourrait imaginer que peut être la pérennité des parcelles sera plutôt correcte ici. Il y a quand même quelques parcelles de reprise qui sont liées de mémoire à de l'alpage parce qu'en fait à Malleval en bas vous avez des communes de fond de vallée où les gens mettent des animaux. C'est très complémentaire en alpage pour faire du fromage ou avoir du lait. Du coup ces parcelles là sont particulièrement intéressantes car elles sont facilement accessibles car assez plate. Du coup elles sont beaucoup utilisées pour l'alpage.

Question de la salle : Est-ce que cela peut influencer l'avenir ?

Réponse D.J : Pour répondre à votre question, à la fin de ma thèse j'ai conclu avec une phrase du genre : « *Toute ressemblance avec des paysages futurs n'est que pure fiction.* » Cela vous donne une idée de la croyance que j'ai dans l'avenir de ce genre de méthode en termes de prévision. En revanche, je peux vous dire qu'en tant que "*classifieur*", ou classificateur, je ne sais pas comment on peut dire, pour arriver parmi un ensemble de communes candidates que je pourrais faire en zone de montagne, si je fais l'hypothèse que ces communes peuvent être classées par ce classificateur et s'il est représentatif de tout cela et bien je peux parvenir à une bonne classification. Pour nous c'est intéressant en science d'avoir des classifications car cela nous permet d'arriver à mieux cerner les problèmes évolution de l'agriculture. Peut-être là-dedans il y a des gens qui ont trouvé des systèmes qui sont pérennes que nous n'avions pas détecté. Grâce à cela ils vont créer une catégorie en sortie et qui sera une catégorie intéressante à développer avec une nouvelle agriculture. Typiquement, l'agriculture biologique, au début, personne ne l'avait vu venir. Avant c'était 1%- 2%. Maintenant, c'est devenu un modèle, pas dominant, mais qui est fort dans la production agricole. Donc ce qui est extraordinaire avec ces outils là c'est qu'on peut travailler à la marge. On peut regarder où est l'innovation. On trouve les grandes tendances qui valident le modèle. On a ces petites catégories, un peu parallèles où l'on peut réfléchir pour voir parce que c'est souvent là que se trouve l'innovation. L'innovation, elle n'est pas dans la masse de l'évolution à mon sens. Elle est plutôt sur la marge donc on a la mémoire de ce qui se fait et à la fin on a sur le bord de la mémoire des petits mouvements, qu'il faut aller regarder nous scientifiques, sur lesquels il y a l'innovation.

Là c'est la commune de Malleval. Les communes dont je parlais se trouvent ici. Vous retrouvez la forme ici et ici c'est les communes d'alpage. Ici ce sont les parcelles près fauchés, des parcelles bien entretenues qui restaient dans la partie la plus stable de l'évolution.

Question de la salle : Au niveau du monde agricole, quelle est la part de disparition du monde agricole dans ces régions ?

Réponse D.J : En fait c'est beaucoup moins grave que ce qu'on pensait. C'est-à-dire qu'il y a des regroupements d'exploitations. Souvent les agriculteurs sont âgés et il y a une diminution du nombre d'agriculteurs en nombre mais par contre l'emprise reste assez importante. Pour le Vercors, j'y ai travaillé deux ans en pratiquant divers métiers, cela c'était pendant une partie de mon travail de thèse, ce qui est bien avec le Vercors c'est que c'est pas mal entouré. Il y a

Grenoble qui n'est pas loin donc il y a des effets de tourisme. Par contre, la déprise se fait souvent en bordure dans les zones difficiles à exploiter. C'est pour cela que les critères de pentes, par exemple, alliés aux critères sociaux font que là où vous voyez apparaître des systèmes de friches sur les bords, vous voyez la forêt regagner du terrain. Ce n'est pas gravissime puisqu'on a besoin d'oxygène mais ceci dit en termes d'exploitation et de production il y a une légère baisse. Mais elle se fait à la marge. Elle ne se fait pas dans la masse des parcelles qui de toute façon sont associées à des exploitations qui sont solides et qui en plus ont été regroupées parce que il y a eu des personnes âgées qui n'ont pas eu de descendants. Du coup, il y a eu souvent des gros agriculteurs qui ont récupéré ces terres et cela tourne pas mal.

Il y a également les aides de la PAC. A l'époque, j'avais travaillé sur l'article 19 qui était la première directive européenne pour aider l'agriculture de montagne. Je ne sais pas si vous voyez dans le Vercors le Mont Aiguille ? C'est joli. Il y a une échelle pour monter derrière, s'il y en a qui font de l'escalade sachant qu'il y en a qui en font mais bon... Le mont Aiguille c'est un peu une icône du paysage du Vercors. Sous ce mont, vous avez un paysage agricole qu'il faut préserver parce que c'est l'association entre le paysage agricole et cette espèce de "saillance" montagnaise qui vraiment fait le cachet sur les cartes postales et tout ça. Vous pouvez constater que l'agriculture c'est plus que de la simple production, c'est également du paysage. Alors ils ne veulent pas qu'on les appelle comme ça mais à terme c'est clair il va falloir qu'on trouve des subventions. C'est important car la fonction de l'agriculture c'est beaucoup plus que la production agricole surtout en montagne.

Illustrations de cette étude

Prévision du réseau de neurone sur la commune de Malleval (Vercors) : forte déprise

Prévision du réseau de neurone sur la commune de Malleval (Vercors) : déprise moyenne

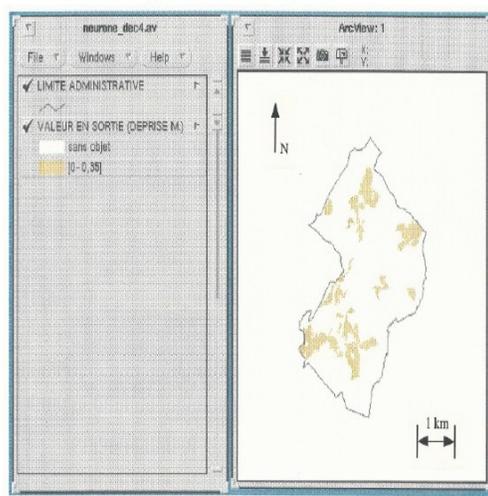
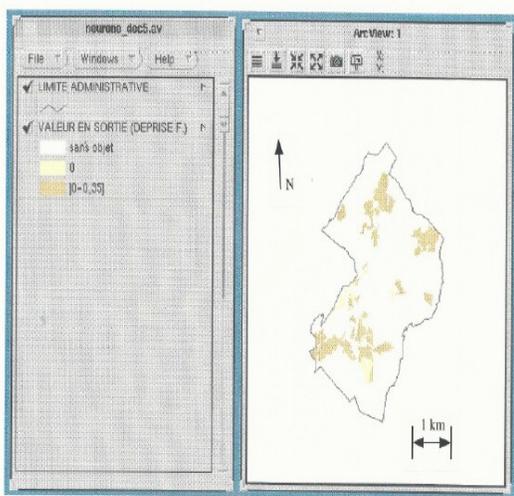


Fig. IV.43 - Valeurs en sortie du réseau à rétropropagation de gradient pour la décision de forte déprise (commune de Malleval)

Fig. IV.42 - Valeurs en sortie du réseau à rétropropagation de gradient pour la décision de déprise moyenne (commune de Malleval)

Prévision du réseau de neurone sur la commune de Mallevial (Vercors) :
déprise faible

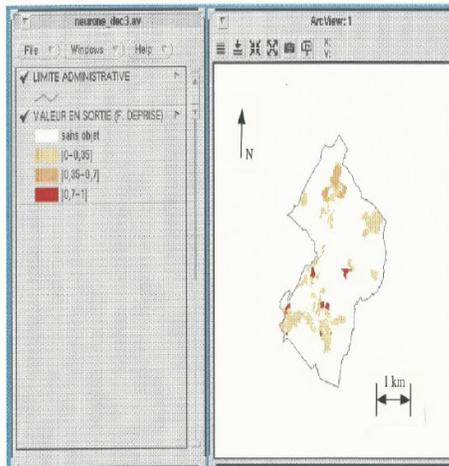


Fig. IV.41 - Valeurs en sortie du réseau à rétropropagation de gradient pour la décision de faible déprise (commune de Mallevial)

Prévision du réseau de neurone sur la commune de Mallevial (Vercors) :
emprise

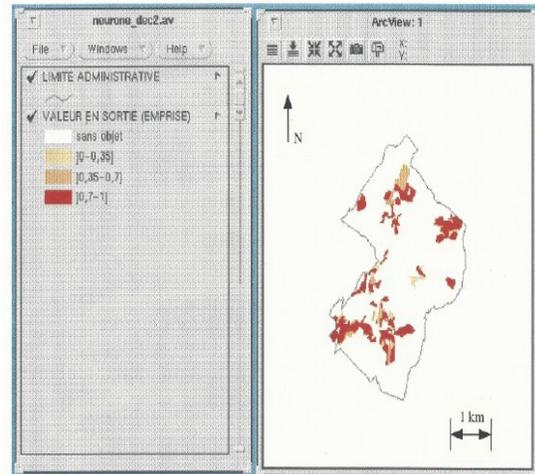


Fig. IV.40 - Valeurs en sortie du réseau à rétropropagation de gradient pour la décision d'emprise (commune de Mallevial)

Prévision du réseau de neurone sur la commune de Mallevial (Vercors) :
reprise

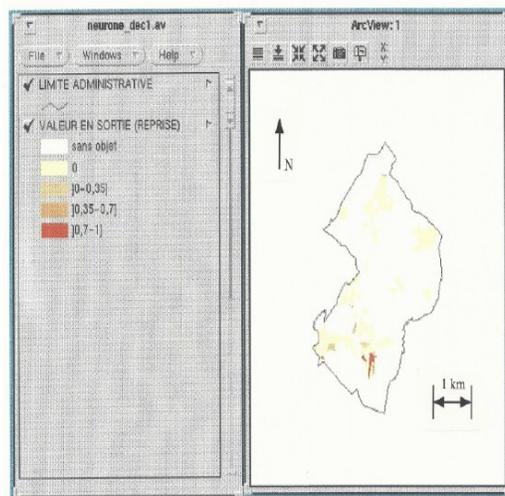
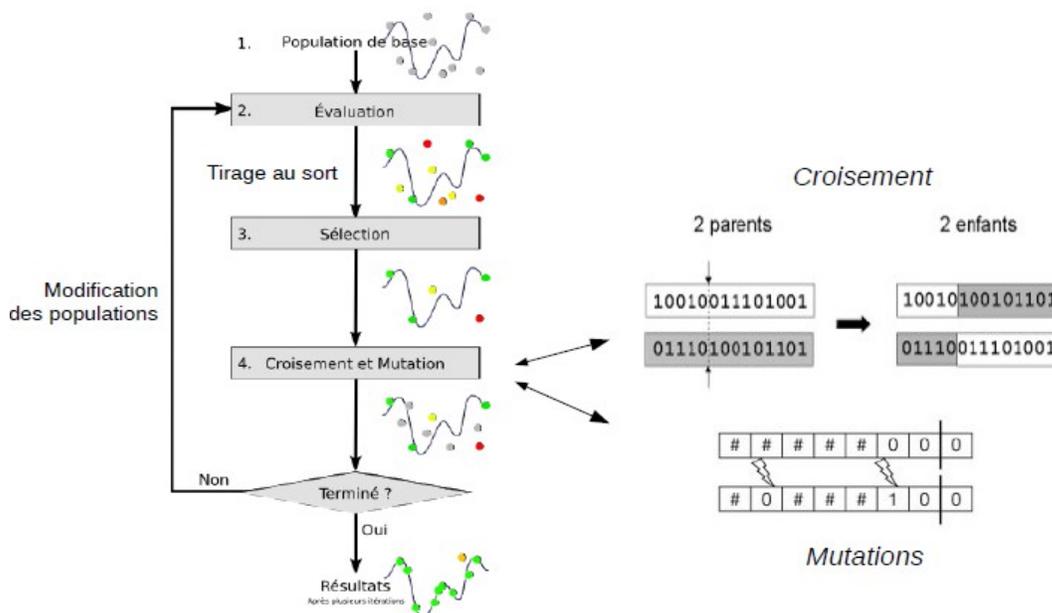


Fig. IV.39 - Valeurs en sortie du réseau à rétropropagation de gradient pour la décision de reprise (commune de Mallevial)

Là, je vous ai présenté un exemple de mémoire stockée dans une structure, une mémoire cache sans jeux de mot, une mémoire un peu immédiate. Il y a une autre mémoire que l'on peut utiliser qui est basée sur la théorie de Darwin et qui pour le coup est liée à la remarque de la dame de tout à l'heure sur le stockage de l'évolution, comme quoi l'ADN serait le prochain support de stockage de données dans le futur.

Cette mémoire c'est : Comment peut on à travers le temps améliorer un processus ? . Alors nous en géographie cela nous intéresse, par exemple dans la géographie des transports. On utilise des algorithmes qui nous trouvent des tracés optimums pour par exemple desservir des gens dans les villes, pour les amener au marché etc... On a utilisé ces méthodes pour justement arriver à améliorer cela. Comment cela fonctionne ? C'est simple. A la limite, si par exemple on décidait d'aller ensemble au parking, il y a plusieurs façons de s'y rendre si les portes sont ouvertes ou pas mais il y a une seule manière optimale c'est d'aller voir le gardien et de lui demander qu'il nous ouvre la porte. Donc on va commencer par créer une population de gens qui vont aller un peu partout, tester des solutions de transport ou plutôt d'accès à leur voiture qui seront plus ou moins bonnes. Puis à un moment on va voir que la meilleure cela va être celle qui sera la porte ouverte. Petit à petit, on va tous faire pareil parce qu'on va avoir une sélection en fait des meilleures populations. C'est-à-dire en fait que vous allez mourir dans la simulation et à chaque fois on va garder le meilleur. Celui qui va avoir été le plus optimal pour rejoindre sa voiture il va être conservé dans un tirage au sort. Ce tirage au sort va à chaque fois se focaliser sur ceux qui ont été les meilleurs. Cela est donc élitiste en quelque sorte.

Principe d'un algorithme génétique simulé



On mime la théorie de l'évolution. Comment cela marche ? On a donc une population de base, celle que je viens de créer à l'instant, et une évaluation celle d'un parcours dont on va calculer le temps et donc vous allez avoir une espèce de "scorum", de score, alors vous allez être bon ou vous n'allez pas être bon, mais c'est de là que l'on procède à la sélection. Parmi vous on en prend et on va croiser. Par exemple, il y en a un qui va partir à droite et puis qui va bifurquer à gauche, un autre va faire l'inverse. On va croiser leur trajectoire. On va faire un croisement génétique, exactement, s'il y a des profs de bio dans la salle, comme une mitose et cela va créer de la diversité génétique. On mime ce processus. On peut aussi injecter des mutations c'est-à-dire on peut, par exemple quelqu'un qui va aller rejoindre sa voiture, lui faire faire un petit écart aléatoire qui va peut-être améliorer finalement parce qu'il a fait une petite erreur, il est allé à

gauche parce qu'il a cru que c'était la porte mais en fait il est revenu. Si on lui enlève cet écart, il va être théoriquement meilleur. Donc du coup on a une population qui s'améliore, c'est-à-dire qu'on va créer avec ces mutations une nouvelle population. Cette population va commencer à exister. Ce qu'on va faire ensuite c'est le tirage de la bonne fortune. Effectivement, vous avez là la population... Alors imaginez-vous que l'on vous ait regroupé en neuf catégories et puis on va tirer au sort. La catégorie en fonction de votre score va grossir, du coup la probabilité que ce sont les meilleurs qui restent à la fin va augmenter. Je vous l'accorde c'est très élitiste voire même assez *bourrin* comme méthode mais c'est très efficace.

La roue de la bonne fortune...

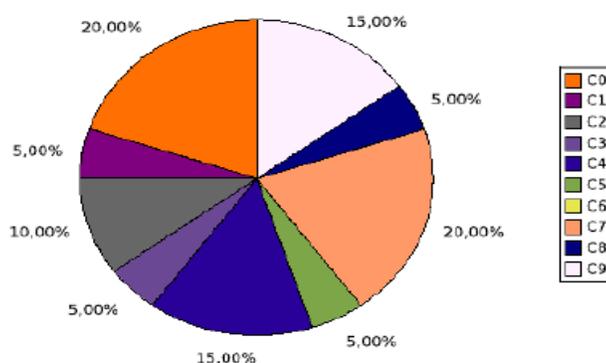


FIG. 5.5: Roulette des scores possible pour sélectionner les chromosomes à croiser

Je vais vous donner l'exemple de Besançon. Vous voyez ici c'est l'idée justement sur laquelle a travaillé un de mes thésards. Il a écrit des programmes afin de calculer les plus courts chemins. C'est ce qu'on appelle un transport à la demande. L'objectif c'est de desservir le centre de Besançon. Les gens partent d'un peu partout, aléatoirement. Au début, ils ont des trajets un peu chaotiques. Ce qu'on fait alors au fur et à mesure avec "*l'algogénétique*"; on va faire des mutations, on va brasser etc...on avance et on obtient des tracés qui globalement ont été améliorés et on a des nouvelles lignes qui convergent toutes vers le milieu. Ce n'est plus le halo un peu chaotique des premiers tracés. Ça c'est du à l'utilisation des algorithmes génétiques. Il faut savoir que j'ai travaillé sur Besançon pendant un certain nombre d'années et que le système « *Evolisgard* » c'était mon équipe qui l'avait conçu. On n'a pas utilisé ces algorithmes car ils étaient trop lents, on a utilisé d'autres algorithmes mais il faut savoir que tout cela participe aux méthodes qu'on utilise même en géographie et en aménagement qui sont basées sur la mémoire. En effet, là qu'est que l'on voit ? On voit que l'on mémorise les meilleures solutions. On arrive à la fin à avoir quelque chose qui n'est pas forcément optimal mais qui n'est déjà pas mal.

Là c'est un autre exemple sur Montbéliard avec un autre thésard qui lui a utilisé d'autres méthodes encore plus élitistes qui ont été assez efficaces aussi. Là on a travaillé directement sur l'environnement Google. On a des origines, des destinations aléatoires de clients. L'objectif c'est de les regrouper avec le même véhicule et de minimiser le nombre de véhicules et les distances parcourues. Tout cela on le met dans un "*algogénétique*" et celui qui a gagné, enfin la solution qui a gagné c'est celle où la somme des temps de parcours est le plus court, ou au contraire celle où

la qualité de service est la meilleure. On met ce que l'on veut. On a en fait une fonction d'objectif et le score qu'on a eu tout à l'heure, ce qui fait la qualité d'une solution c'est lié à cette fonction d'objection qui peut s'assimiler à une fonction à tête chercheuse. Par exemple, on voit des solutions où tous les clients ont été desservis avec des notions de boucles qui ont été faites comme une espèce de parcours optimum qui aurait relié ces points.

L'avantage c'est qu'on fait fi des considérations politiques dans la recherche, c'est cela qui est bien. Non non, je plaisante. On a bossé sur des systèmes opérationnels utilisés par année par des dizaines de milliers de clients sur différents sites comme Montbéliard, Besançon, d'autres communes de Franche Comté etc... malheureusement, la boîte qui éditait nos logiciels a coulé ce qui fait que j'ai passé un été à répondre au client qui me demandait ce que tout cela allait devenir. J'ai dû payer le serveur qui était dupliqué avec trois ordinateurs pour avoir de la redondance et éviter les risques. J'ai payé l'été un peu et puis après la boîte a coulé et cela nous a posé des problèmes. Là j'ai un ordinateur qui a mémorisé l'ensemble de ces applications parce que là c'est quand même neuf thèses. Là j'en suis à ma neuvième thèse sur les transports comme ça, soit trente-six ans de travail cumulés la dessus.

Un autre exemple, un troisième thésard, que j'ai épuisé il y a deux ans, a travaillé sur les algorithmes de fourmis. Quand vous voyez une fourmilière, on serait à même de penser que les fourmis connaissent les structures externes de leurs habitats. Apparemment, il n'en n'est rien. Je ne sais pas quels ont été les questionnaires effectués pour pouvoir mettre cela en évidence mais ceci dit il semblerait que ce ne soit pas le cas. Cela voudrait dire que la fourmi a une connaissance très partielle. La fourmi laisse des informations à ses copines du type : « *Viens, suis moi par là il y a de la bouffe et maintenant tu la ramènes à la fourmilière qui est là-bas.* ». Chacune ramène son petit grain de nourriture. Si vous regardez après ce que c'est qu'une fourmilière, vous constaterez que toutes les fourmilières se ressemblent. Il y a une émergence, c'est une des théories de la complexité. C'est-à-dire que si on distribue la connaissance, il y a peut-être une forme de mémoire, c'est très lié à ce que disait tout à l'heure Georges. Si on distribue ainsi la connaissance partielle, cette mémoire, on arrive à faire émerger quelque chose de structurant. C'est riche comme idée.

Alors nous on est parti là-dessus en géo. On s'est demandé est ce que les taxis ne peuvent-ils pas utiliser les pratiques de distribution de la connaissance comme les fourmis ? Ces artisans travaillent seuls comme ils sont souvent amenés à le faire, sachant qu'ils peuvent également travailler en équipe mais cela reste assez concurrentiel. Du coup, on a travaillé sur des algorithmes qui peuvent leur permettre de communiquer et d'échanger les informations. Comment ça fonctionne ? Les véhicules partent et au fur à mesure qu'ils prennent des clients, les taxis laissent des phéromones virtuels pour dire : « *Là il reste des clients. Vas y en priorité* ». Cela a l'air de rien mais à part le G7 qui fait ça, et encore on se demande comment se fait l'optimisation parce que à Paris des fois ou même à Lyon, vous ne trouvez pas de taxis alors qu'il y a le G7 et que donc ils devraient travailler dans des modes coopératifs comme celui-là. Donc je ne sais pas trop comment cela fonctionne. Nous, notre idée ce n'est pas de recréer le G7, c'est de recréer, de mettre à disposition une information qui serait partagée et qui permettrait simplement de savoir où sont les clients aux autres taxis. Une fois les clients embarqués, le taxi va regarder s'il reste d'autres clients et laisse sa phéromone. Sinon il continue à chercher au hasard, de manière aléatoire. On va voir ce que cela donne du coup.

Cela sont les résultats que l'on a fait sur une simulation sur Avignon et sur Dakar, parce qu'on a regardé ce qui se faisait sur ces différents sites. Comme ce sont des transports à la demande un peu innovants, on a regardé que les africains sont très forts : Taxis brousse, taxis clandestins, taxis allant au-delà de la capacité du véhicule... Cela vous donne une idée de l'efficacité. Je vois qu'il y en a qui ont voyagé en Afrique, c'est du vécu. On a fait plusieurs modèles. Je ne vais pas vous les décrire. C'est juste pour replacer le seul modèle qui fait de la mémoire : C'est celui que j'ai appelé coopération avec les phéromones parmi les autres.

Modèle de coopération indirecte inter-véhicules basé sur les phéromones

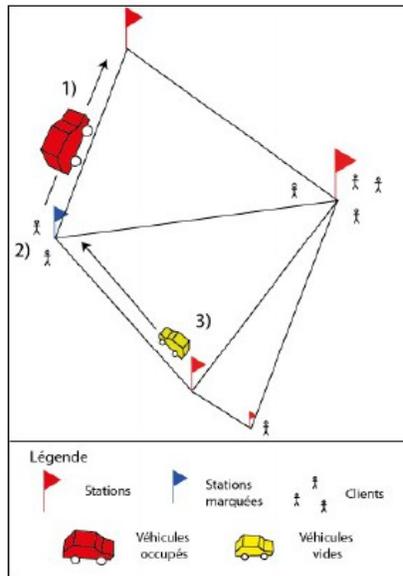


Illustration 54 : choix de déplacement basé sur la coopération inter-véhicule.

1. Les véhicules embarquent les clients aux stations et les déposent à leur destination
2. Une fois les clients embarqués, le véhicule va regarder s'il reste des clients en attente à la station. Si oui, il va marquer la station et informer les autres véhicules.
3. Quand il ne sont pas occupés, les véhicules se dirigent en priorité vers les stations marquées.

Dans les autres modèles il y a :

- Le modèle aléatoire : Tous les taxis sont indépendants ils ne communiquent pas, ils se promènent aléatoirement. C'est un petit peu ce qui se fait dans beaucoup d'endroits. Vous pouvez constater qu'en utilisant ce modèle ils réussissent à ne pas être si mauvais.
- Le modèle lié à la distance : Ils font en sorte que plus le client est proche et plus ils vont le chercher. Comme ils sont en compétition, c'est moins bon que l'aléatoire. C'est quand même incroyable. Cela veut dire qu'ils viennent en fait et le second vient juste derrière et après hop c'est terminé. Comme quoi il vaut mieux soit faire de l'aléatoire comme ça dans son coin sans rien dire à personne, faire son taf et pas écouter les bips radios ; soit faire du coopératif en communiquant.
- Le modèle potentiel lié à l'attractivité : je ne vais pas détailler mais on a des marchés entre guillemet où on a des clients qui veulent aller dans des zones attractives. On s'est basé la dessus pour faire le modèle. Dans certain cas, il a donné de bons résultats mais là en l'occurrence ce n'est pas terrible.
- Le modèle taxi clandestin africain : celui qui part avec des espèce de corridors, qui attend d'être plein pour partir. Ils égrainent leur clientèle en mode de cabotage avant de revenir au point de départ. C'est n'est pas très souple car il faut attendre que cela se remplisse etc... c'est pour cela que c'est un rendement assez faible.

En revanche, regardez le gain rien qu'avec la coopération par rapport au système aléatoire. Rien que le fait de laisser une petite phéromone, je ne parle pas de téléphoner à votre copain qui est le plus proche, juste laisser une phéromone accessible à l'ensemble des autres taxis. Regardez comme cette information-là ramène en efficacité. Cela vous donne le potentiel d'efficacité de cette méthode. Là cela fait émerger un système cohérent en quelque sorte qui est caché derrière. Et donc c'est une forme de mémoire puisque c'est une mémoire qui est très localisée puisque vous la posez quelque part et elle est distribuée et puis elle est prise ou pas prise par les véhicules.

Par exemple, prenez ce plan simplifié du réseau d'Avignon. On a fait la même chose sur Dakar. Il représente le centre ville et l'extérieur d'Avignon. On a tout un tas d'indicateurs qui nous permettent de regarder si on est bon ou pas. C'est à partir de cela qu'on va inventer les systèmes de demain en ce qui concerne les transports en appliquant la méthode de biomimétisme.

Il faut savoir qu'en biologie il y a beaucoup de mémoire. On voit les réseaux de neurones, les algorithmes génétiques. Ce sont des mémoires avec des temporalités différentes. On voit des phéromones qui ont une mémoire volatile, un peu la mémoire dont Georges parlait tout à l'heure.

Georges Linarès : Ce que Didier a montré là c'est lié à la fonction de mémoire. C'est d'autre façon d'enregistrer de la mémoire dans un ordinateur. Du coup on est assez loin de ce dont je vous parlais en début de présentation c'est-à-dire de codage dans la mémoire vive, par exemple, la métaphore de la commode. Ce qui est intéressant dans les réseaux de neurones notamment c'est que ce sont des systèmes qui capturent de l'information en représentant des exemples et ils apprennent au travers de ces exemples qu'on leur a représentés. Il y a donc un effet de mémorisation mais cela va un peu au-delà en fait c'est aussi un effet de généralisation, capacité d'inférence donc d'appliquer à des objets, à des cas nouveaux une logique qui a été capturée à partir des exemples. On va un peu au-delà en fait par simple mémoire. En tout il y a un truc intéressant, c'est que dans les réseaux de neurones on n'a pas du tout ce lien direct entre l'information mémorisée et un endroit dans lequel elle serait contenue, une sorte de support de la mémoire. On est dans une mémoire qui est répartie c'est-à-dire qu'un objet tel qu'il serait mémorisé dans le cadre général, une information telle qu'elle est mémorisée dans un réseau de neurones est en réalité mémorisée sur l'ensemble du réseau. Chaque neurone et les liens entre eux et le poids de ces liens, dont parlait Didier, c'est l'ensemble de tout cela qui va mémoriser l'ensemble des informations. C'est un point de vue qui est très global sur la mémoire.

Il y a un autre aspect qui est une distribution de la mémoire. Cela veut dire qu'on a une répartition de l'information sur une structure globale qui stocke la mémoire cela donne des capacités de robustesse, de plasticité. Ça permet aussi de stocker des grandes quantités d'informations en faisant de façon un petit peu implicite de l'oubli. C'est une sorte de souvenir approximatif de ce qui s'est passé et dans le processus l'oubli est extrêmement important c'est-à-dire que dans la capacité de généralisation d'un réseau de neurone, dans sa capacité à appliquer une règle, une logique qui a été observée et/ou apprise à partir d'exemples, cette capacité de l'appliquer à d'autres exemples c'est basé sur la sélection de ce qui est important et de ce qui ne l'est pas. On ne voit pas cela, c'est dans la logique du réseau.

En fait, l'intelligence vient de ce que l'on se souvient et de ce que le réseau a su oublier. En fait ces deux aspects : mémoire et oubli, c'est l'ensemble des deux qui font l'intelligence de la machine. Donc cette idée de distribution on la retrouve à d'autres niveaux et notamment sur le Web. Le Web est un ensemble énorme de machines interconnectées. Je vous ai dit tout à l'heure la quantité d'informations qui était stockée globalement sur de l'informatique et aussi sur le web. Là ce sont des statistiques que j'ai trouvées sur le nombre de noms de machines et sur les sites actifs. Alors cela va être un petit peu différent car vous pouvez avoir des machines qui hébergent de l'information, qui stockent de l'information qui n'est pas active et d'autres qui seraient actives. Là ce qui est intéressant c'est qu'on avait un niveau très marginal entre 1995 à 2000. Mais ces quelques dernières années, il y a eu une véritable explosion de tout cela avec des rapports de un à cent million. Actuellement on a plus de cent millions de noms d'hôtes référencés. Donc en gros, indépendamment de l'activité on a plus de cent millions de machines qui sont supposées accessibles par le net.

Le Web comme mémoire distribuée

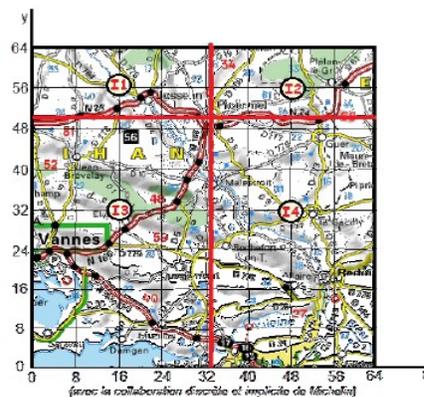
- ensemble énorme de machines interconnectées
 - ces machines partagent de l'information
 - la structure matérielle est masquée



Didier Josselin : Quand vous allez sur Internet et que vous tapez sur Google une adresse cela prend combien de temps ? Pour répondre, cela est quasiment instantané. Est-ce que vous imaginez, tout à l'heure Georges nous a dit qu'une image prenait une capacité d'un giga-octet ; est-ce que vous imaginez les quantités d'images de Google sur l'ensemble de la terre à quoi cela correspond ? Quand vous mettez tous les réseaux, quand vous zoomez, vous avez vu, vous avez toutes les images, cela fait des milliers et des milliers, je ne sais pas même à quoi cela correspond mais c'est énorme. Comment vous pouvez expliquer que vous pouvez zoomer en un quart de seconde sur une adresse sur cet ensemble d'information ? Et bien c'est ce que je vous propose maintenant de comprendre.

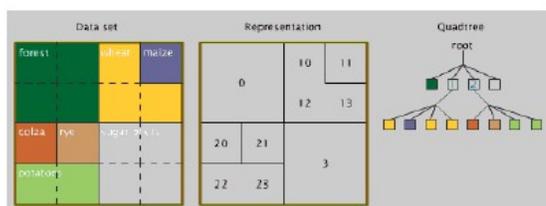
Déjà cette idée qu'on va voir d'indexation spatiale elle n'est pas très nouvelle au sens que sur les cartes IGN vous avez cette sorte de quadrillage. Cela donne une idée d'un repérage à la base on va dire relatif parce que c'est cela l'idée. Vous avez un carré à côté de l'autre etc... mais on peut aller beaucoup plus loin et c'est ce qu'on va essayer de voir ensemble.

L'indexation spatiale : c'est pas nouveau ;)

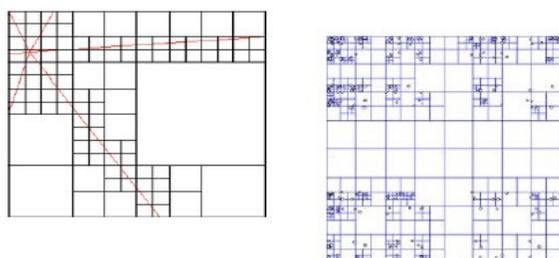
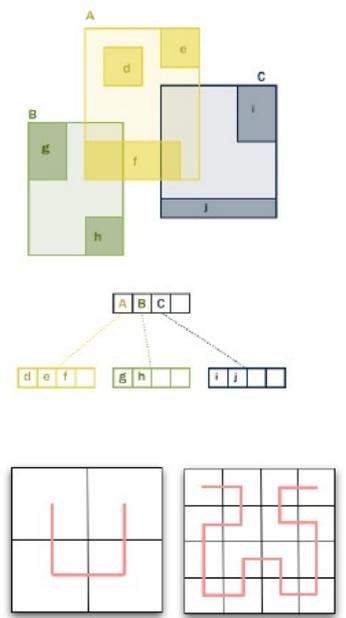


Typiquement, les informaticiens ont inventé des outils extraordinaires qui sont les arbres en général : les tris. Ici par exemple, on a en haut à gauche un exemple de comment avec un arbre (au sens arborescence) on peut représenter un petit carré en fonction de comment il est agencé, inclus dans un carré de plus grande taille. Si on prend, comme il l'appelle : « la racine » en haut, vous avez le carré vert qui correspond à la forêt qui est là. Ensuite à droite vous avez un mélange de couleurs et vous descendez dans l'arbre. A chaque fois que vous descendez dans l'arbre regardez ce qui se passe. Vous éliminez tout le reste puisque vous avez cette indexation. Le premier niveau quand vous êtes sur le vert vous avez déjà enlevé d'un coup 75 % de l'information à aller chercher pour faire un zoom. Si vous savez votre « adresse », quand vous donnez votre « adresse » sur Google, elle est à l'intérieur de ce carré vert, vous avez éliminé 75% et vous descendez dans votre arbre. Vous continuez à un niveau au-dessus et vous savez que vous être à l'intérieur du carré suivant. Mais si vous faites un calcul de localisation à l'intérieur de ces quatre carrés cela vous prend un quart de seconde et paf vous éliminez à nouveau 75% de l'information. En d'autres termes, au troisième niveau vous avez déjà éliminé 75 % x 75% x 75 % de l'information. Vous faites cela avec un niveau d'échelle croissant et là vous avez à travers l'arbre une méthode extrêmement rapide pour aller zoomer immédiatement sur une image. Vous ne le voyez pas, mais peut être avez-vous constaté quand cela ne fonctionne pas très bien, les images vous parviennent par carrés. Donc derrière il y a cette idée d'indexation qui n'est pas forcément physique mais qui est liée de manière logique à cela.

Principe



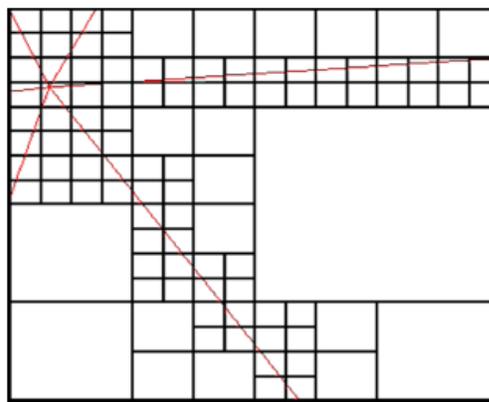
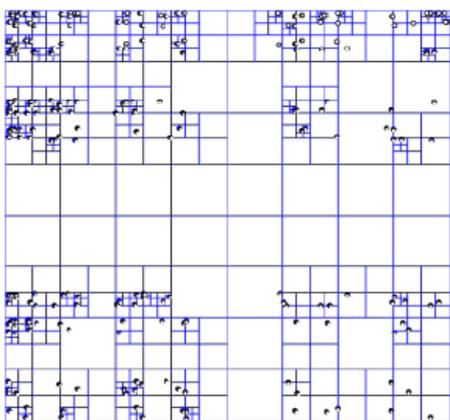
R-tree Hierarchy



Alors après il y a plein de méthodes permettant d'économiser cette information. Par exemple, quand on n'a pas de réseau qu'il n'y a que des prés ou alors on ne s'intéresse pas à l'information qui est là et bien on ne va pas mettre d'indexation. On va gagner encore, on peut jouer sur l'arbre. On peut élaguer l'arbre, on peut faire un arbre qui ne soit pas complet, qui soit plus simplifié pour permettre de descendre plus vite dedans. Ici on voit une approche fractale puisque vous allez zoomer à chaque fois et vous retrouverez des petits points qui sont des échantillons de quelque chose. Donc vous voyez il y a différentes méthodes. En haut à droite, vous avez une autre méthode qui n'est pas forcément une agrégation, c'est un R-tri. Bon vous en

avez plein après... De toute façon tous ces arbres sont meilleurs que si vous aviez à faire, vous imaginez, le reste de savoir vos coordonnées, savoir où vous devriez aller, faire un calcul de distances, rechercher les éléments les plus proches et zoomer dessus... cela prendrait des années. Donc vous voyez, on est dans des processus où grâce à une formalisation de ce type-là, qui finalement est une forme de simplification pour moi de la mémoire puisqu'on a une image qui est en mémoire et là on a une structure qui permet d'aller très très vite dans cette information.

Découpages fractales en échelles



Voilà c'est juste pour que vous identifiez la notion. Je sais que les fractales c'est l'un de vos dadas (référence sur le cours de J-P Cohen-Addad l'année traitant des frontières). Les structures, on les retrouve. Les fractales sont utilisées notamment pour faire de la compression d'image parce que là cela va vous permettre de faire de la compression. C'est exactement ce que disait Georges tout à l'heure, partout où il y a une homogénéité, où il y a une texture qui est équivalente, on va considérer que ce n'est qu'un seul groupe de pixel et ainsi économiser de l'information.

Georges Linarès : Alors juste pour revenir sur l'accès aux informations sur le web, il y a un aspect que l'on n'a pas trop développé et qui est absolument critique, c'est le moteur de recherche. Vous savez ce que c'est un moteur de recherche. Quand vous faites une requête sur Google ou un autre moteur, je vais vous parler en utilisant ce que je faisais en recherche, le travail du moteur de recherche c'est de trouver ce que vous voulez ou plutôt du moins ce qu'il comprend que vous voulez. Vous avez formulé sous forme de requête les mots clés que vous voulez. Évidemment, et cela Didier vient d'en parler, pour accéder à l'ensemble des contenus pertinents il faut analyser le contenu et également comment il peut être pertinent et correspondre au mieux à cette requête. Il faut aussi être capable de le faire à grande vitesse parce que quand vous avez vu la quantité de données stockées sur Internet il faut aussi que ce traitement aille très vite. En fait dans cette façon d'accéder par mots clés au contenu, il y a aussi toute une problématique liée à l'indexation et à la façon dont on va coder l'accès à cette information qui est contenue et ce que Didier vous a présenté. Les arbres qu'il vous a montrés c'est une logique qui permet d'accéder à ce que l'on veut de façon très rapide sans tout parcourir ce qui serait évidemment impossible.

Il y a un autre aspect qui est intéressant et qui pose des questions à plein de niveaux différents c'est que du coup la mémoire sur le web est-ce que c'est encore la capacité de mémoriser l'information ou est-ce que c'est la capacité d'y accéder ? En fait dans le moteur de recherche, l'interface entre la donnée et l'utilisateur fait l'analyse de la requête. Il fait ressortir les documents supposés correspondre à votre requête. Il y a peut-être des documents qui ne sortent jamais de là. Qu'est-ce que serait un document qui ne serait jamais envoyé par un moteur de recherche ? Est-ce qu'on a mémorisé un document, est-ce qu'on peut considérer qu'on a mémorisé l'information si elle se trouve quelque part sur le web mais qu'on n'arrive jamais à y accéder ? Donc en fait cette question de l'information stockée et de la technique d'accès ou du moyen d'accès à l'information, finalement cela oblige à reconsidérer ce qui est de la mémorisation et ce qui est de l'oubli. Finalement c'est le moteur de recherche, c'est lui, en vous rendant accessible ce qui est mémorisé ou au contraire en vous le masquant, en ne vous rendant plus accessible quelque chose qui est stocké quelque part c'est finalement lui qui fait la fonction mémoire et oubli, puisque c'est lui qui rend l'information vivante.

Alors c'est un point qui est intéressant et qui pose un tas de problème car dans les moteurs de recherche il y a des logiques qu'on ne maîtrise pas très bien. Ce que je veux dire, c'est comment Google fait le lien entre votre requête et l'algorithme qu'utilise Google. Le lien entre votre requête et les documents qui vous sont présentés . Il y a des choses que l'on sait que l'on connaît sur lesquelles il y a des sujets de recherche, des technologies qui évoluent. On sait comment faire le lien entre un document texte par exemple et quelques mots clés qui représenteraient la requête. Mais Google ne gagne pas ainsi son argent. Quand je dis Google, c'est le moteur de recherche en général dont je parle. Je n'ai pas décidé de donner des actions à l'un plutôt qu'à l'autre. Les moteurs de recherche ont en général des stratégies qui sont commerciales. Ils ont des modèles économiques qui font qu'à un moment il faut qu'ils gagnent leur vie d'une façon ou d'une autre. Ils peuvent gagner leur vie en faisant de la publicité. Ce qu'ils font ou alors en regardant ce que vous faites, les requêtes que vous effectuez, ce que vous recherchez, ce que vous avez acheté des choses sur le net et vous proposent des similaires. Ils utilisent également des référencement pour gagner leur vie.

Question sur le mode économique du moteur de recherche libre : Donc en faisant en sorte que lorsque vous tapez votre requête ceux qui sont intéressés à apparaître d'abord, bien sûr, à condition qu'ils mettent la main au portefeuille. Ce n'est absolument pas un truc secret dont je vous fais part, il s'agit juste du modèle économique des moteurs de recherche. Cela pose un problème car finalement ce que l'on considérerait comme un grand espace de stockage, neutre, partagé, accessible par tous, gratuit et libre ; en réalité c'est une idée purement utopique ou du moins en tout cas trompeuse. Les données peu importe qu'elles soient stockées quelque part si on n'y a pas accès ou si on nous en donne une vision qui est quelque peu déformée par tout un mécanisme économique virtuel, on n'a pas exactement le retour qu'on cherchait. Cela on peut le voir dans les deux sens. A la fois on peut ne pas avoir accès à une info qui nous serait utile finalement mais qui n'est pas indexée par le moteur de recherche ou qui est "*secondarisé*" par ce dernier puisque cela va apparaître très loin dans la liste de recherche. Et puis il peut y avoir l'effet inverse. Il peut y avoir des données qu'on voudrait voir disparaître parce que cela nous est personnel. Par exemple, un film que vous aviez fait à quinze ans et que vous voudriez effacer des réseaux. La question est de savoir si on est encore propriétaire de ses propres informations mise en ligne. Est-ce que l'information est mémorisée ou pas et pour combien de temps ? En tout cas cela pose la question de nos droits sur les informations que l'on dépose sur le net.

Il existe un décalage temporel entre l'instant où quelqu'un met de l'information en ligne et le moment où cette information est disponible beaucoup plus tard avec une persistance de l'information dans le temps. Cette information qui a été reproduite

L'ordinateur est aussi de la mémoire avec un décalage temporel mais c'est cela la mémoire. Un livre c'est une forme de mémoire. Cela stocke de l'information durant un temps indéterminé. C'est cela de la mémoire.

Est-ce que quand on dépose de l'information sur Internet on en est encore propriétaire ? Le problème c'est que quand je l'ai déposée je ne la consulte plus, pour qu'il y ait problème il faut que ce soit connu. C'est plutôt cela. Ce qui agit dessus aussi c'est la quantité d'informations émises. Un adolescent qui fait usage de l'information sur le net est ce qu'il a bien conscience qu'elle y est peut-être à vie et surtout accessible par tout un chacun. C'est aussi la façon d'utiliser une mémoire de cette immense capacité qui est mise en cause. C'est une mémoire dans un format inhabituel. Quand on dépose de l'information qui est susceptible d'être distribué de cette façon-là, on n'en maîtrise pas la persistance.

Question de la salle sur le mode économique du moteur de recherche libre

Réponse de G L : Le problème des moteurs de recherche libres c'est que l'infrastructure technique est très lourde. En fait le modèle du moteur libre c'est que c'est un peu compliqué d'avoir autant de serveurs que Google. Ils offrent un service, certains sans pub, mais plus lent. Ceci étant dit, je ne veux pas faire de publicité, mais il existe une société française qui propose un moteur de recherche libre : Qwant. Par contre ce qu'il disent c'est qu'il n'y a aucune trace personnelle. Autrement dit si vous cherchez sur le net une paire de chaussures via leur moteur de recherche vous ne trouverez pas sur d'autres pages consultées à la suite de la recherche, des publicités d'autres sociétés qui vous proposeront leurs chaussures.

On a besoin d'éthique et de simplification mais en même temps on a envie d'utiliser les moteurs de recherche. Alors maintenant reste à savoir quel type de moteur de recherche souhaite t on consulter ?

Question de la salle : le problème de la mémoire ce n'est pas seulement d'accéder ou de ne plus accéder à la mémoire, c'est la qualité de l'information à laquelle on a accès. Je m'exprime mal. Y a-t-il des travers ou des déformations dans l'information que l'on retrouve. Par exemple dans le cas des associations, comme la mémoire est très associative comme pour les moteurs de recherche, on peut faire des associations plus ou moins bonnes. Du coup c'est encore autre chose ce n'est pas je ne demande pas à me faire oublier. Si des associations sont faites sur Internet à mon propos est ce que ses associations vont être effacées.

Réponse de G L : C'est encore plus compliqué à maîtriser car c'est le regroupement et toute la logique derrière l'analyse des requêtes, d'évaluation de la similarité des documents entre eux. Il y a toute cette mécanique-là à développer mais considérer que deux documents répondent bien à une requête c'est parce que ces deux documents se ressemblent. C'est de cette association dont vous parlez. Pourquoi deux documents se ressemblent ? On peut imaginer qu'ils ont une similarité que l'on ne voudrait pas voir apparaître ou le contraire. Il y a un autre aspect qui est l'effet de masse. Quand vous tapez une requête sur Google, vous obtenez une somme de réponses qui se ressemblent beaucoup car le système est validé pour chaque document de la requête mais il n'y a pas de réponse globale. Il y a donc plein de réponses qui se ressemblent. La première est la mieux "scoré", la mieux évaluée par le moteur par rapport à la requête. La seconde est le second meilleur document par rapport à la requête. Du coup on se retrouve avec une somme volumineuse de réponses qui ne sont pas forcément très informatives.

En réalité il y a un travail la dessous en laboratoire, c'est d'essayer de considérer, de faire des résumés, de traduire la diversité des contenus, de chercher la diversité des contenus dans l'espace réduit qui est l'écran pour éviter ce phénomène d'écrasement par la mémoire. S'il y a beaucoup de documents ils vont peut-être masquer toute la diversité qu'il y a derrière. Alors que quand vous demandez une requête vous avez peut-être envie de voir cette diversité y compris des documents qui seraient mis à la marge. Il y a des questions derrière donc on va passer à la neutralité *net* c'est cette façon d'équilibrer c'est ce que l'on décide de montrer à un utilisateur qui fait une requête

Question de la salle : Quand on fait une requête la plupart des réponses amènent vers le même site. Comment cela se fait-il qu'il n'y a pas une analyse de ces réponses-là ?

Réponse de G L : Vers le même site oui mais par contre il faut qu'il soit très similaire et l'analyse n'est pas faite. Entre deux documents pourquoi on en choisirait un mais pas l'autre et qu'on déciderait de prendre un autre document qui est lié mais qui est plus éloigné de la requête parce qu'on pense qu'il apporte quelque chose d'original et qu'on pourra peut-être y trouver son compte. Alors je ne dis pas qu'il n'y a pas de solution, mais je dis que c'est compliqué parce que en fait c'est aussi la façon d'évaluer ce que veut vraiment l'utilisateur. Parce qu'il y a ça aussi : comment formule t on la requête ?

Remarque auditeur / Je ne sais pas si vous connaissez l'existence du magazine *Clé* qui a fait un numéro spécial en septembre sur la mémoire justement et dedans il y avait un article qui disait que la taille du cerveau avait arrêté de croître à partir du moment où les gens avaient commencé à écrire et à stocker l'information à l'extérieur de leur cerveau. Donc c'est marrant le rapport. Alors peut être le *Cloud* va peut-être nous amener à avoir des petites têtes, je sais pas... ou peut être que toutes ces connexions vont nous amener à avoir la grosse tête.

Question de la salle : est qu'au niveau des neurotransmetteurs et de ces substances chimiques inter-neurones vient toute la richesse et la complexité des sections neuronales. Est-ce que du coup en informatique il y a une équivalence des neurotransmetteurs ?

Réponse de G L : C'est un peu ce que décrivait Didier tout à l'heure, c'est qu'en fait entre deux neurones informatiques, pour le coup, il y a un lien et que ce lien est validé. C'est le « W » qui le montre... c'est une valeur, cette valeur qui dit à quel point il y a une dépendance entre les deux neurones. Donc quand le neurone est activé, il va plus ou moins, son activation va plus ou moins provoquer l'activation du neurone connecté et le niveau de dépendance entre les deux ça dépend de ce poids. Je ne sais pas si l'on peut considérer que ce soit du même ordre que les neurotransmetteurs

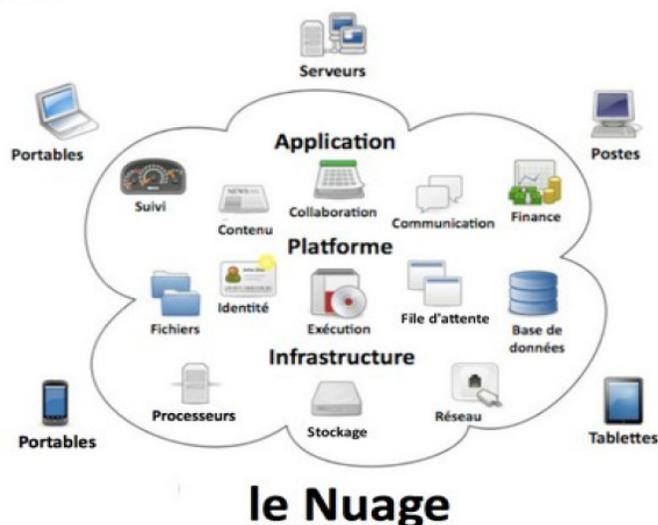
Réponse de D.J : Les neurotransmetteurs se sont des substances chimiques qui sont beaucoup moins rapides que les neurones, il faut le savoir. Cela fait la régularité de l'information. La dopamine et tous ces trucs font des connexions des plus rapides. Il faut garder en temps que ce n'est pas parce que vous allez complexifier un système qu'il va être plus performant. Typiquement, nous en géographie quand on construit des réseaux on met des difficultés à se déplacer super précises, millimétriques. On n'a pas de meilleur résultat dans l'optimisation que si l'on fait quelque chose qui est fait à peu près correctement. On s'est aperçu que, toute l'énergie que l'on mettait pour mettre les stops, les feux et les machins... et bien l'un dans l'autre avec tous les flux, tout cela se « moyennise » et il suffit, et d'ailleurs on a déjà vu cela, il y a un truc qui s'appelle Kiss : *Kepp It Simple and Stupid*. Cela veut dire qu'il y a des gens en ce moment qui travaillent dans la complexité, qui cherchent des modèles ultrasimples en fait et qui soient un peu génériques et capables d'absorber tout un tas de ...

Georges Linares : Et par ailleurs c'est un peu le principe d'un réseau de neurones, la structure d'un réseau de neurones. C'est très simple un neurone, c'est simple, c'est connecté. Après si l'on regarde à une échelle microscopique un neurone, il est extrêmement sophistiqué. Ce qui fait la complexité d'un réseau de neurones c'est sa complexité globale, c'est ce qui fait que nous sommes plein de structures très simples et qui sont connectées entre elles. Donc structurellement c'est très simple, l'intelligence vient de la masse on remplace finalement si vous voulez...en fait c'est stupide et à force cela devient pénible.

Un mot sur le *Cloud*. C'est un principe récent qui existe en réalité depuis assez longtemps et qui se démocratise depuis assez peu de temps. L'idée c'est qu'on a un stockage. On stocke des informations sur Internet mais en fait ce n'est pas une sorte de contrôle, ce n'est pas une sorte de stockage individuel. Ce que je veux dire c'est que si vous voulez mettre des données quelque part à priori vous n'allez pas les mettre sur Internet. Par contre le *Cloud* c'est justement ça, cette sorte de mélange de tout cela. En tout cas, c'est une structure sur Internet qui est dédiée au stockage individuel des informations par les utilisateurs. Donc le *Cloud* c'est un ensemble d'ordinateurs qui sont connectés entre eux. On ne voit pas du tout la complexité de ce groupe d'ordinateurs mais cela offre juste un service à l'utilisateur qui s'occupe de stocker l'information. Vous avez un fichier, je ne sais pas, vous avez des films en vidéo par exemple plutôt que de les stocker sur votre machine, d'acheter des disques durs, enfin l'approche traditionnelle, et bien l'idée là c'est de stocker sur un espace qui est distant, qui est accessible par Internet, qui est géré par un prestataire de service un opérateur qui a des gros paquets de machines et qui va offrir un service payant qui est de stocker vos données. Donc l'intérêt c'est qu'en fait on est un peu moins dépendant de l'espace. Par exemple vous avez un disque dur qui fait cinq cents giga et vous vous rendez compte qu'en fait il en fallait sept cent ou vous achetez un disque dur qui fait deux téras et vous n'en occupez que cinq cent giga. Voilà donc le fait de « *méthodiser* » l'espace de stockage a un effet de parage et d'optimisation. Du coup c'est avec cette marge là que des gens gagnent de l'argent.

Le Cloud

- *Cloud computing* : ensemble de ressources partagées



le Nuage

Après les *Cloud* ce n'est pas seulement au niveau de la mémoire, il y a une sorte de généralisation de cette idée de nuage avec ce qu'on appelle le *Cloud computing* c'est une ressource distante accessible à peu près n'importe où et n'importe quand à partir du moment où l'on est connecté et qui s'adapte facilement au besoin par exemple pour faire des calculs qui nécessitent beaucoup de puissance de calcul, en général ce n'est pas les gens individuellement mais cela le pourrait ou alors pour décompresser des films par exemple ou faire du traitement vidéo qui est un peu lourd. Le nuage peut alors mettre en votre possession un ensemble de processeur dont vous avez un usage très ponctuel et donc aucune raison de l'acheter pour vous. Alors c'est un peu généralisé, vous avez les stockages des données, les processeurs, les applications... Vous avez besoin ponctuellement d'une application alors plutôt que de l'avoir à la maison pour en fait la sous utiliser pourquoi ne pas utiliser l'application qui est partagée ou mutualisée au travers du Web et quand vous êtes tous connectés la vitesse de transfert de données sur le réseau est assez rapide donc ce n'est pas pénalisant.

Question de la salle : n'y a-t-il pas un danger de perte d'informations? Il y a des fournisseurs d'accès qui offrent des espaces de stockage...

Réponse de G L : *Ouaie* mais ce qu'il y a c'est que ces fournisseurs d'accès ils peuvent financer une infrastructure parce qu'elle est à une échelle qui n'est pas du tout celle d'un foyer. Ces prestataires peuvent mettre en place des structures qui sont très robustes parce que c'est partagé parce que vous n'allez pas chez vous avoir un modulateur qui aura la capacité de cinquante mille machines. Vous pouvez là avoir des planchers qui sont suspendus parce que les vibrations ça peut provoquer des problèmes sur les machines vous pouvez avoir des protections. Vous pouvez mettre en œuvre des dispositifs qui ont un sens quand on a une grande quantité de machines et qui n'en n'ont plus quand on a une ou deux machines.

Question de la salle : Problème de l'accès aux données quand on n'a pas accès à Internet ?

Réponse de G L : Là, il y a un autre problème. C'est le principe des *caches* de dissocier la synchronisation de nuage local des données que l'on peut utiliser et du stockage.

Question de la salle : Sur la propriété des données

Réponse de G L : Qui est propriétaire...Ne plus avoir de données chez soi, c'est sûr qu'on peut se faire piquer les données. Déjà qu'on peut se les faire piquer quand elles sont chez soi.. Ceci dit c'est un vrai problème d'avoir des données qui sont hébergées par des sociétés commerciales qui ne sont pas trop encadrées. C'est l'aspect juridique qui reste dans le flou.

Réponse de D.J : On peut juste dire que par rapport au CNRS, notre réseau scientifique de connexion, une de nos politiques c'est de ne plus travailler avec Google. On a des systèmes qui nous permettent maintenant de mettre des informations. On pense que Google est parasité entre guillemet, enfin certains le pensent, parasité par le gouvernement américain. Il y a des tas d'informations qui sont analysées et donc c'est du vol d'idées. Quand vous mettez des informations sur Google, c'est écrit qu'elles ne vous appartiennent plus, c'est quand même la réalité. Il faut le savoir. Il y a une réaction en ce moment qui est assez nette, les discours sont importants au CNRS pour dire : « *ne mettez plus vos données sur Google on vous donne des outils pour avoir notre propre intranet* » en espérant qu'on ne vienne pas nous piquer nos idées chez nous. On a quand même des spécialistes en informatique qui sont très forts et qui nous évitent ce genre de problème.

Juste pour vous montrer une dernière image sans laquelle on n'aurait jamais pu trouver certains sites archéologiques. Ceci pur vous faire méditer sur l'idée que le numérique ça a du bon. Pourquoi ? Parce qu'on peut changer d'échelle, parce qu'on peut le manipuler à travers les arbres dont je vous ai parlé tout à l'heure. Et là si vous êtes en tracteur pour le coup vous allez voir des traces vertes mais elles ne vont pas vous indiquer que c'est un site archéologique. En revanche, dès que vous changez d'échelle tac... vous pouvez avoir un changement d'objet vous le repérez et vous pouvez trouver des sites extraordinaires. Donc le numérique, le stockage de l'information, c'est important.

Photographie aérienne du site protohistorique de Grézac (Charente-Maritime)

Grande nécropole gauloise avec diverses structures funéraires. Circulaires ou carrés, de toutes tailles, ces "monuments funéraires" abritaient des sépultures à incinération.



Après je vous ai mis des informations sur informatique et liberté qu'on utilise beaucoup en géographie.

Lois et textes fondamentaux



Le 23 mai 2014, la Cour de Justice de l'Union Européenne rendait sa décision sur le « droit à l'oubli numérique », instauré ainsi le **droit à l'oubli** du même nom auprès de Google et à disposition de chaque internaute judiciaire européen.

Comment est venu au tel droit ? Quelles ont été les avancées réalisées ou ce sera avant cette décision ?

Quelques rappels sur l'évolution de la législation française dans ce domaine.

La loi Informatique et Libertés (Février 1978)

Cette loi reconnaît quatre droits essentiels : à savoir :

- Le droit d'information
- Le droit d'accès
- Le droit de rectification

Informatique et Liberté (1978)

Directive européenne 95/46/CE : la CNIL avant-gardiste

- Texte de référence à dimension européenne en matière de protection des données à caractère personnel. Cette directive entend harmoniser les normes en matière de protection des données personnelles dans l'UE, en vue de faciliter leur libre circulation. Elle stipule également que ces données ne doivent pas être soumises à un traitement automatisé, sauf si celui-ci remplit les exigences posées par trois principes : proportionnalité, transparence et finalité légitime.

- Par ailleurs, l'article 28 demande à chaque État-membre d'instituer une autorité de protection des données personnelles, sur le modèle général de la Commission nationale Informatique et libertés (CNIL) établie en France.

Conclusion de Georges Linares : Ce qu'on a essayé de vous montrer c'est que la mémoire c'est une question qui est très centrale en informatique. Ce n'est pas seulement la barrette de mémoire en support technique de la mémoire mais c'est aussi des principes dont on représente l'information la façon dont on l'organise globalement. Il y a beaucoup de variété dans tout cela. En réalité, la mémoire ce n'est pas une technologie qui est mure, il y a des évolutions au niveau technologique mais aussi sur les méthodes et parmi aussi leur représentation. Ceci étant dit le matériel évolue très rapidement c'est ce que je vous ai montré au départ. Globalement, les structures qui supportent la mémoire évoluent radicalement c'est ce qu'on vous a montré en vous parlant du web et du *cloud*. Ce n'est pas seulement une évolution quantitative du support technique mais c'est aussi une évolution dans la façon dont tout cela se développe et sur les usages.

Conclusion de Didier Josselin : En géographie vous avez vu que la mémoire est très géographique. Elle est distribuée, du coup il y a ces notions de distances, de nœuds, de connexion. C'est très lié la géographie et la mémoire. Je voudrais dire que par rapport aux réseaux sociaux j'ai un certain optimisme par rapport à certaines notions de mémoire car le fait de la distribuer justement va peut être faire qu'on va sauver des choses. Donc je me dis que peut être les jeunes avec leur nouvelle façon d'échanger, de partager c'est peut être une solution. Vous avez vu tout à l'heure le système de coopération, même avec des phéromones on arrive à améliorer des systèmes parce qu'on partage l'information. Moi je pense que les solutions de la mémoire et de son usage sont dans le partage à travers les réseaux sociaux. J'espère que les jeunes feront mieux que nous de ce côté-là. Je sais, je suis « *has been* » de ce côté-là !