

Réseaux de neurones formels : physique statistique et théorie de l'information

J.-P. NADAL

*Laboratoire de Physique Statistique**
Ecole Normale Supérieure,
24, rue Lhomond, F-75231 Paris Cedex 05, France.

XXVème Journées de Statistiques, Vannes, 1993

L'une des causes du renouveau d'intérêt apparu dans les années 80 pour les *réseaux de neurones formels* est l'utilisation fructueuse des concepts et outils de la physique statistique. Tout d'abord en 1982 J. J. Hopfield montre une analogie entre des neurones formels (neurones binaires de McCulloch et Pitts) fortement interconnectés et un ensemble de moments magnétiques (spins) en interaction. En même temps Hopfield introduit la notion de *calcul par attracteur* : la *réponse* du réseau à un stimulus est l'état stationnaire, ou plus généralement l'attracteur de la dynamique atteint à partir de l'état initial imposé par le stimulus. Ceci permet en particulier de proposer de nouveaux modèles de mémoires associatives.

Dans cet exposé je parlerai de la seconde contribution de la physique statistique, très différente et plus récente, qui est l'étude de l'apprentissage - apprentissage d'associations et surtout apprentissage d'une règle par l'exemple (auquel cas on s'intéresse à la *généralisation*) - réalisable par un réseau de neurones d'une architecture donnée. Ce thème de recherche a été initié par E. Gardner [1] en 1987, qui montra qu'il était possible de faire de la *mécanique statistique dans l'espace des réseaux* : il s'agit en fait de mettre en oeuvre l'approche bayésienne de l'inférence à l'aide des outils propres à la physique statistique. Pour un réseau d'architecture donnée, on se donne une mesure *a priori* sur l'espace des poids synaptiques, et on regarde la mesure obtenue en tenant compte de ce que le réseau doit réaliser un ensemble donné d'associations. On utilise une méthode d'entropie maximale (outil de base en physique statistique), et on s'intéresse au comportement asymptotique des grands réseaux, le nombre de poids tendant vers l'infini : cette limite est l'analogue de la "limite thermodynamique" pertinente en physique. Des calculs effectifs sont réalisés pour des architectures suffisamment simples (par exemple celle du perceptron, réseau en couche sans neurone "caché"), et ceci grâce aux méthodes issues de la physique statistique des systèmes désordonnés. Cette approche permet de calculer des capacités d'apprentissage (donc de calculer les performances optimales en apprentissage), et d'obtenir les performances optimales et typiques en généralisation. On retrouve certains résultats obtenus dans les années 60 par des

*Laboratoire associé au C.N.R.S. (U.R.A. 1306) et aux Universités Paris VI et Paris VII.

méthodes mathématiques toutes différentes, mais la plupart des résultats sont nouveaux. Dans mon exposé je présenterai cette contribution de la physique à la théorie de l'apprentissage et je donnerai les principaux résultats obtenus au cours des cinq dernières années.

Un concept clef en physique statistique est celui d'*entropie*. L'apprentissage est lié à une diminution d'entropie ou, autrement dit, à un gain en *information*: l'*entropie de mélange* des physiciens est en effet identique à la *quantité d'information* de Shannon en théorie de l'information. Il est ainsi plus judicieux de mesurer les performances en terme de quantité d'information stockée [2] plutôt qu'en terme du nombre d'associations réalisées, et il est intéressant de voir l'apprentissage d'une règle par l'exemple comme une diminution d'entropie [3]. Dans mon exposé je montrerai que les liens entre physique statistique et théorie de l'information sont particulièrement intéressant dans le cadre de la théorie de l'apprentissage. En particulier je montrerai que le calcul de la capacité maximale de stockage d'un perceptron [2] a une interprétation dans le cadre de la théorie de la communication, de sorte que cette capacité découle d'un théorème de Fanno obtenu dans les années 50 - un grand classique en théorie de l'information [4]. Je montrerai également que physique statistique et théorie de l'information permettent de relier et comparer, de façon inattendue, l'étude des deux grands paradigmes de l'apprentissage, l'apprentissage supervisé et l'apprentissage non supervisé [5].

References

- [1] Gardner E. Maximum storage capacity in neural networks. *J. Physique (France)*, 48:741–755, 1987.
- [2] Brunel N., Nadal J.-P., and Toulouse G. Information capacity of a perceptron. *J. Phys. A: Math. and Gen.*, 25:5017–5037, 1992.
- [3] Carnevalli P. and Patarnello S. Exhaustive thermodynamical analysis of boolean networks. *Europhys. Lett.*, 4:1199, 1987.
- [4] Blahut R. E. *Principles and Practice of Information Theory*. Addison-Wesley, Cambridge MA, 1988.
- [5] Nadal J.-P. and Parga N. Duality between learning machines: a bridge between supervised and unsupervised learning. *LPSENS preprint, submitted to Neural Computation*, 1992.