

REDES DE NEURONAS Y CLASIFICACION: DE PARADIGMAS CONEXIONISTAS CLASICOS A REDES QUE APRENDEN MIENTRAS CRECEN

Juan-Manuel Torres-Moreno

Laboratoire Informatique d'Avignon

BP 1228 84911, Avignon Cédex 09, FRANCE

1 INTRODUCCION

¿Pueden las computadoras aprender a resolver problemas a partir de ejemplos? esta cuestión que bordeaba no hace mucho tiempo la frontera de la ciencia ficción es actualmente objeto de profundos estudios. Las redes de neuronas formales (RN) son máquinas que poseen esta capacidad de aprendizaje. Estas máquinas han sido propuestas como modelos extremadamente simplificados del funcionamiento del cerebro, puesto que no retienen más que algunas características esenciales: i) las neuronas no pueden encontrarse más que en dos estados posibles, activas o en reposo; ii) están interconectadas mediante sinapsis que pueden ser modificadas por aprendizaje y iii) el estado de una neurona a cada instante es determinado por el de otras neuronas, información que es transmitida por las sinapsis. Aunque muy esquemático, este modelo presenta una riqueza sorprendente de estados y de comportamientos que ha sentado las bases de un modelo de memoria y aprendizaje como un fenómeno emergente colectivo: el sistema global presenta propiedades complejas que no pueden predecirse a partir del estudio individual de sus componentes. Igualmente, las aplicaciones en diversos dominios no tardaron en aparecer. Los estudios teóricos de redes de neuronas reflejan estos dos aspectos: el de la modelización de fenómenos cognitivos y el del desarrollo de aplicaciones. Aunque las RN hayan sido aplicadas a diversos campos, en el presente artículo se hará énfasis en la clasificación de datos: para los humanos una actividad tan trivial que pasa desapercibida, pero que presenta dificultades importantes para una máquina.

2 APRENDIZAJE Y CLASIFICACION

La clasificación es la atribución de una clase específica a un objeto. Esta atribución necesita un cierto grado de abstracción para poder extraer generalidades a partir de los ejemplos de los cuales se dispone. Para una máquina, la clasificación de rostros, de datos médicos o de formas son tareas bastante difíciles, en tanto que para un humano son cuestiones cotidianas. Por ejemplo, en el caso de reconocimiento de caracteres manuscritos, es difícil enunciar una descripción general que tenga en cuenta todas las variaciones particulares de cada carácter. Una técnica que puede ser utilizada para resolver este problema es el aprendizaje. Así, el criterio para decidir si una imagen corresponde a una letra **A** consiste en comparar si esta imagen es lo suficientemente similar a otras *A*'s vistas anteriormente; con este enfoque, uno no calcula la clasificación de letras, sino que se aprenden a partir de ejemplos. El

aprendizaje está implícito en el quehacer humano y forma parte imprescindible de las actividades intelectuales. Pero surge entonces la cuestión inevitable: ¿qué es el aprendizaje? El aprendizaje --desde un punto de vista pragmático-- consiste en la adaptación de los parámetros de un sistema (sea artificial o natural) para obtener una respuesta deseada frente a un estímulo externo. Esta definición amplia del aprendizaje puede ser formalizada con el paradigma del aprendizaje supervisado o del profesor-alumno: imagine que existe una red-profesor que conoce la relación exacta entre los estímulos y las respuestas (entradas-salidas), que una red-alumno desconoce. Si ambos son expuestos a una entrada cualquiera, el profesor es capaz a cada momento de indicar al alumno la respuesta correcta. Los parámetros del alumno deben ser entonces, ajustados para entregar la misma respuesta que la del profesor. Este ajuste o aprendizaje se puede realizar por técnicas iterativas de minimización de un costo (cuantificación de los errores en las respuestas). Por supuesto, en la vida real no existen tales profesores, y de lo único que se dispone son datos en forma de pares de entrada-salida, a los que llamaremos ejemplos. En efecto, es únicamente a partir de estos ejemplos que deben ser construidas las redes.

3 NEURONAS FORMALES

El origen de las redes de neuronas se encuentra en la representación de la neurona biológica por McCulloch y Pitts en 1943. Una RN con un gran número de unidades en interacción es esquematizada en la fig. 1.

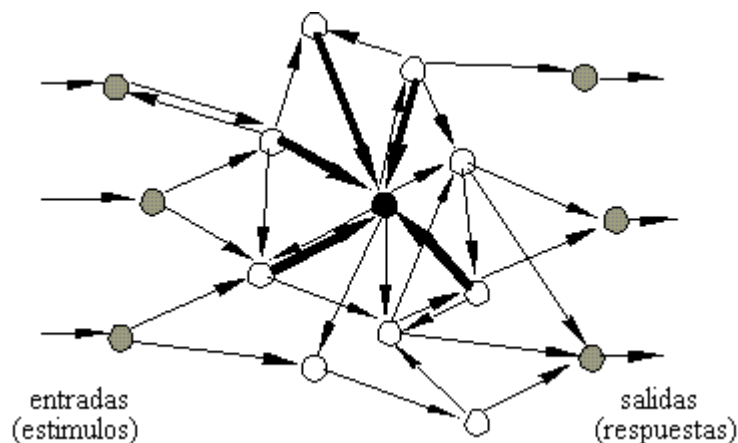


Figura 1. Red de neuronas formales. Se muestran las neuronas de las capas de entrada y de salida. Una neurona particular es indicada en el centro. Sus sinapsis de entrada son representadas en línea gruesa.

Este modelo considera que una neurona puede ser representada por una unidad binaria: a cada instante su estado puede ser activo o inactivo. En otros modelos, el estado de la neurona es descrito por una variable continua. La interacción entre las neuronas se lleva a cabo a través de sinapsis (o pesos sinápticos). Según el signo, la sinapsis es excitadora o inhibitoria.

El perceptron está constituido por las entradas provenientes de fuentes externas, las conexiones y la salida. En realidad un perceptron es la red neuronal más simple posible: aquella donde no existen capas ocultas. Para cada configuración de los estados de las neuronas de entrada (estímulo) la respuesta del perceptron obedece la siguiente dinámica: se suman los potenciales sinápticos y se comparan con un umbral de activación. Esta suma ponderada es también llamada campo. Si el campo es mayor que un umbral, la respuesta de la neurona es activa, sino es inactiva. Con una arquitectura tan simple como la del perceptron no se pueden realizar más que una clase de funciones booleanas muy simples, llamadas linealmente separables. Son las funciones en las cuales los estados de entrada con salida positiva pueden ser separados de aquellos a salida negativa por un hiperplano. Un hiperplano es el conjunto de puntos en el espacio de estados de entrada, que satisfacen una ecuación lineal. En dos dimensiones, es una recta, en tres dimensiones un plano, etc. Si se quieren realizar funciones más complejas con RN, es necesario intercalar neuronas entre las capas de entradas y de salida, llamadas neuronas ocultas. Una red multicapas puede ser definida como un conjunto de perceptrones, ligados entre si por sinapsis y dispuestos en capas siguiendo diversas arquitecturas. Una de las arquitecturas más comúnmente usada es llamada feedforward: con conexiones de la entrada a las capas ocultas y de éstas hacia la salida.

El funcionamiento de una RN es gobernado por reglas de propagación de actividades y de actualización de los estados. Teóricamente, una RN puede ser vista como un modelo que realiza una función de un espacio de entrada hacia un espacio de salida. El objetivo de esta modelización consiste en que la asociación sea lo más acorde posible con el medio ambiente del fenómeno estudiado.

La inmensa ventaja de los métodos conexionistas, basados en unidades neuronales y enlaces sinápticos, comparado con los métodos tradicionales de Inteligencia Artificial (IA) es la siguiente: no es necesario conocer ni una expresión ni una construcción de la función a modelar, tan sólo se requiere disponer de un conjunto de ejemplos satisfactorio (conjunto de aprendizaje) para que la red pueda aproximar esta función aplicando una regla de aprendizaje. Es por ello que las redes de neuronas son ampliamente utilizadas en aplicaciones tan variadas como la previsión, la predicción, la clasificación, el diagnóstico automático, el procesamiento de señales, el reconocimiento de formas, la compresión de datos, la optimización combinatoria, la robótica, la búsqueda de documentos, entre otras. La mayor parte de las reglas de aprendizaje se fundan en el ajuste iterativo de las conexiones, pero otras pueden modificar la arquitecturas misma de la red. Un modelo conexionista es definido por el tipo de neuronas, por la arquitectura y por una regla de aprendizaje. Las características comunes y deseables a todos estos modelos son: un fuerte potencial de autoorganización, una cierta robustez frente a perturbaciones externas (memoria distribuida, deslocalizada y redundancia de información), un paralelismo masivo e inherente. Sin embargo, los modelos difieren entre sí por diversos aspectos: sus motivaciones biológicas, su modo de funcionamiento o su campo de aplicaciones.

4 ARQUITECTURAS CLASICAS vs. ARQUITECTURAS INCREMENTALES

El aprendizaje con redes de neuronas se realiza actualmente siguiendo el enfoque de la Retropropagación de Gradiente (Backpropagation)[1] y el de los algoritmos constructivos. El primero necesita introducir a priori el número y conexiones de las unidades ocultas, y determinar los pesos por minimización de un costo. La red así obtenida es eventualmente simplificada eliminando unidades y/o conexiones que parecen inútiles. El principal defecto de este enfoque consiste en la búsqueda de la mejor arquitectura a través de prueba y error. Por otra parte, con un enfoque constructivo se aprende al mismo tiempo el número de unidades y pesos, en el marco de una arquitectura que comienza generalmente con un perceptron. La característica de estos algoritmos es que construyen una RN adaptada a cada problema particular, usando la información contenida en el conjunto de aprendizaje y evitando rediseñar la arquitectura. El primer algoritmo constructivo fue el algoritmo Tiling, de ahí surgieron Cascade Correlation, el algoritmo Upstart, Offset y GAL entre otros.

Monoplan y NetLines son dos algoritmos recientemente introducidos por el autor. De aquí en adelante, nuestra discusión será basada únicamente en los métodos constructivos.

Una vez construida la red, ésta debe ser capaz de predecir la clase de datos nuevos que no estén presentes en el conjunto de aprendizaje. La calidad del algoritmo de aprendizaje se traduce en la capacidad de predicción de la RN. Esta calidad se mide a través del error de generalización, que es la proporción de clasificaciones correctas realizadas por la red sobre nuevos datos. Esta cantidad se mide empíricamente sobre una serie de problemas estándar (benchmarks) que sirven de prueba.

5 DOS NUEVOS ALGORITMOS CONSTRUCTIVOS

Durante mi tesis doctoral[6] desarrollé algoritmos constructivos con una sola capa de unidades ocultas binarias. Esta elección no fue fortuita: queríamos redes que fueran comprensibles, compactas y al mismo tiempo fáciles de interpretar. Ciertos teoremas mostraban que una sola capa oculta, con un número suficiente de unidades es capaz de realizar cualquier función de las entradas. La palabra clave es suficiente: ¿cuántas unidades son suficientes? Para nosotros, en virtud del carácter constructivo de las redes, la pregunta se traducía en ¿hasta cuando se deben agregar unidades a la capa oculta? Intuitivamente decidimos que la respuesta estaba justamente en los datos: el número de unidades dependerá de cada problema en particular. De esta manera decidimos crear redes con una sola capa oculta cuya talla fuera calculada automáticamente por el algoritmo de aprendizaje, rompiendo así con el círculo perverso de prueba y error, característico de la Backpropagation que puede llegar a consumir un tiempo asombrosamente grande en el diseño de la red. El algoritmo resultante, llamado Monoplan[1] agrega sucesivamente neuronas en la capa oculta hasta que sean clasificados correctamente todos los ejemplos del espacio de entradas donde viven los datos. Numerosas aplicaciones[2,6] mostraron el excelente desempeño de Monoplan y sugirieron al mismo tiempo una variante del método: ¿qué sucedería si en lugar de

separar los ejemplos en la capa oculta se clasificaran a la salida de la red sus representaciones internas? El producto de esta idea es NetLines[2], poderoso algoritmo de clasificación que puede calificarse de "método impaciente", pues trata de construir un perceptron de salida que separe las clases cada vez que una unidad oculta es añadida, sin esperar que en la capa oculta sean bien clasificados todos los ejemplos. Este método permite tratar adecuadamente los casos donde las entradas están codificadas con números reales, en tanto que Monoplan está mejor adaptado a tratar entradas binarios.

Ambos algoritmos utilizan unidades ocultas de tipo binario, lo que permite por una parte, una codificación natural a problemas de dos clases (comúnmente usados en aplicaciones prácticas) y por otra parte, se centra en el problema pertinente de la búsqueda de las fronteras entre clases, a diferencia de la Backpropagation donde la simple disminución del costo cuadrático medio no proporciona información de donde se hallan estas fronteras. Eventualmente, las fronteras pueden ser interpretadas en un nivel simbólico a través de reglas. La clave importante de estos nuevos algoritmos consiste, sin embargo, en el correcto aprendizaje de cada una de las unidades individuales, ya que es este punto sensible el que determinará la complejidad de la red. El algoritmo estándar del perceptron no permite aprender mas que conjuntos que son linealmente separables (por tanto, inutilizable en la mayor parte de los casos) y algoritmos como Pocket (que guarda la mejor solución a cada iteración) dependen de la paciencia del usuario. Por ello, adapté un reciente algoritmo de aprendizaje basado en Mecánica Estadística llamado Minimerror[6]. Este método produce un aprendizaje robusto[3] y resultados comparables a aquellos obtenidos con el perceptron de Bayes[4]. Este algoritmo me permitió descubrir que un benchmark (problema del Sonar) conocido en la literatura por su gran dificultad, es en realidad, linealmente separable[5]. Esta prueba mostró la gran calidad de Minimerror al enfrentarse a problemas de alta dimensionalidad. El binomio Minimerror-Monoplan/NetLines parece ser una nueva fórmula dentro de los paradigmas del aprendizaje supervisado que muestra un alto desempeño en tareas de clasificación como se muestra en la siguiente sección.

6 APLICACIONES A LA CLASIFICACION AUTOMATICA

Desde un punto de vista metodológico, el uso de las RN no es trivial, incluso cuando el algoritmo sea fácil a implantar. La dificultad viene del hecho que el tamaño de la red depende fuertemente de la naturaleza de la aplicación a tratar, y que los pocos resultados teóricos actualmente conocidos son sólo fronteras inaplicables directamente en casos prácticos. La calibración de los parámetros de aprendizaje y de la arquitectura de la red sigue siendo por el momento un área de especialistas. Sin embargo, cuando las redes son correctamente adaptadas a un problema, permiten alcanzar a menudo tasas superiores de aprendizaje y generalización que métodos de IA o estadísticos. En el caso de Monoplan/NetLines, el usuario no tiene que preocuparse del enfadoso problema de definir ninguna arquitectura: la red aprende mientras crece en un proceso dinámico y autónomo; y los parámetros han sido calibrados para permitir una cierta independencia del usuario.

Las pruebas de ambos algoritmos han mostrado sin lugar a dudas su

extraordinaria calidad, tanto desde el punto de vista de la generalización como del tamaño (pequeño) de las redes generadas. Se trataron tanto aplicaciones de tipo académicas como realistas. Entre las primeras se hallan los problemas clásicos del XOR en su generalización de N entradas o Paridad. En este problema tanto Monoplan como NetLines encontraron la solución "humana" que consiste en una red con una capa oculta y un número de unidades ocultas igual al número de entradas. Los problemas de Monk's, de formas de ondas de Breiman y el problema de dominios magnéticos fueron igualmente resueltos con éxito[2,6]. Los problemas reales o realistas plantean dificultades interesantes: los datos pueden tener ruido o estar incompletos; las bases son generalmente pequeñas y, por lo tanto se tiene dificultad para construir un buen clasificador; la dimensionalidad del espacio de entradas puede ser alta: en el problema de ecos de sonar[5] se trabaja con espectros muestreados en un espacio de 60 dimensiones. Sin embargo, los resultados de Monoplan y NetLines[2,6] son realmente satisfactorios, pues muestran que las redes generadas poseen la mejor relación generalización/complejidad, comparados contra redes entrenadas con Backpropagation, métodos de IA y métodos estadísticos. Una aplicación interesante es la ayuda al diagnóstico médico: Monoplan y NetLines fueron enfrentados con éxito a la de detección de cáncer y de diabetes usando datos de dominio público en Internet. El diagnóstico de comas toxicológicas en el hospital universitario de Grenoble y el control de calidad de tubos de acero son dos aplicaciones en curso, de las cual se tienen resultados preliminares prometedores[6].

Monoplan y NetLines, dos algoritmos que aprenden mientras crecen, se perfilan como instrumentos poderosos dentro del aprendizaje automático, y permiten al mismo tiempo profundizar en la parte fundamental de la investigación en RN: la comprensión de los mecanismos del intelecto, la cognición y su impacto en la creación de artefactos inteligentes, un tema por demás excitante.

REFERENCIAS

- [1] *Parallel distributed Processing*. James McClelland, Davis Rumelhart and the PDP group. The MIT press, 1989.
- [2] [Efficient adaptive learning for classification tasks with binary units](#). J.M. Torres Moreno, M.B. Gordon. *Neural Computation*, 10(4):1007-1030, 1998
- [3] *Robustness against single event upsets of digital implementations of neural networks*. A. Assoum, J.M. Torres Moreno, N.E.Radi, R.Velazco, M.B. Gordon and R. Ecoffet. ICANN'95, Paris, Oct, 9-13, 1995. Session 9. Editors: Fogelman-Soulié and Gallinari EC2.
- [4] [Finite size scaling of the bayesian perceptron](#). A. Buhot, J.M. Torres Moreno, and M.B. Gordon. *Physical Review E*, 55:7434-7440, 1997.
- [5] [Characterization of the Sonar Signals Benchmark](#). J.M. Torres Moreno and M.B. Gordon. *Neural Processing Letters*, 7:1-4, 1998.
- [6] [Apprentissage par des réseaux de neurones: étude de nouveaux algorithmes constructifs](#). J.M. Torres-Moreno, Ph.D. These. INPG, Grenoble, 1997.

Publicado originalmente en el Newsletter de Lania, Rébsamen 80 91090 Xalapa, Veracruz, México : Newsletter Lania 1-2 , Vol. 23-24 (7), Mexico, 1998