

# Aprendizaje y muestreo de la aproximación Kikuchi

Roberto Santana, Pedro Larrañaga, José A. Lozano

Grupo de Sistemas Inteligentes

Departamento de Ciencias de la Computación e Inteligencia Artificial

Universidad del País Vasco, P.O. Box 649, 20080 San Sebastián - Donostia

rsantana@si.ehu.es, Pedro.Larranaga@ehu.es, lozano@si.ehu.es

## Resumen

En este trabajo se presentan un algoritmo para el aprendizaje de la aproximación Kikuchi a partir de datos así como un método de muestreo de la referida aproximación. Se discuten resultados preliminares de la evaluación del algoritmo en el aprendizaje en datos generados a partir de una instancia del modelo Ising de ferromagnetismo. Los resultados obtenidos indican que la aproximación Kikuchi es capaz de representar de manera factible las dependencias existentes en los datos entre las distintas variables.

## 1. Introduction

Los modelos gráficos han devenido herramientas de representación del conocimiento capaces de representar y manejar relaciones de independencia. Una de las características de estas herramientas que han favorecido su generalización es la existencia de algoritmos de aprendizaje automático de modelos gráficos a partir de datos [2, 4, 16]. Tales algoritmos han constituido uno de los pilares fundamentales para el desarrollo de la minería de datos.

Como regla general los modelos gráficos especifican un conjunto de restricciones sobre la estructura gráfica que determinan la clase de factorizaciones que son capaces representar. Estas restricciones especifican a su vez el tipo de aproximaciones de la distribución que pueden obtenerse con el modelo gráfico. Por ejemplo, en las redes Bayesianas existen restricciones que impiden la existencia de ciclos

en el grafo dirigido. Tales restricciones, y la complejidad de algunos de los algoritmos de inferencia basados en modelos gráficos motivan la conveniencia de investigar otros métodos para la aproximación de distribuciones de probabilidad multidimensionales.

Recientes estudios en el desarrollo de algoritmos de inferencia aproximada [17] han revelado la relación entre el problema de la minimización de la energía en determinados sistemas físicos y la propagación de creencias en modelos gráficos. Inicialmente concebidos para modelos gráficos acíclicos, estos algoritmos pueden ser utilizados con resultados positivos en grafos con ciclos. Estos estudios han mostrado igualmente la conveniencia del uso de las aproximaciones de la energía que utilizan descomposiciones de grafos basadas en regiones. La extensión de algunos de estos resultados al campo de la inferencia estadística resulta un tema de interés por las numerosas aplicaciones que de ella pueden derivarse, entre ellas la minería de datos.

En este trabajo revisamos recientes desarrollos en la investigación de aproximaciones de la probabilidad que utilizan descomposiciones de grafos basadas en regiones. El énfasis fundamental está puesto en la aproximación Kikuchi de una distribución de probabilidad y la presentación de un algoritmo de aprendizaje para este tipo de aproximaciones. La aproximación Kikuchi satisface un conjunto de propiedades Markovianas y de descomponibilidad que permiten su utilización para la aproximación de distribuciones [15]. Igualmente, algoritmos evolutivos de estimación de distribucio-

nes [8, 9] que utilizan la aproximación Kikuchi han mostrado su eficiencia en la solución de diferentes problemas [14].

En la siguiente sección se introduce la aproximación Kikuchi y algunas de sus propiedades fundamentales. La Sección 3 presenta un algoritmo para construir una descomposición en cliques válida. En la Sección 4 se presenta el método utilizado para muestrear la aproximación Kikuchi. La Sección 5 describe cómo se realiza el aprendizaje de la aproximación Kikuchi a partir de los datos. Resultados de un experimento que ayuda a ilustrar la aplicación del método se presentan en la Sección 6, y las conclusiones del trabajo se discuten en la Sección 7.

## 2. Aproximación Kikuchi en la física estadística

Sea  $\mathbf{X} = (X_1, \dots, X_n)$  un vector de variables aleatorias. Usaremos  $\mathbf{x} = (x_1, \dots, x_n)$  para denotar una asignación de las variables.  $S$  denotará un conjunto de índices en  $N = \{1, \dots, n\}$ , y  $\mathbf{X}_S$  (respectivamente  $\mathbf{x}_S$ ) un subconjunto de las variables de  $\mathbf{X}$  (respectivamente un subconjunto de valores de  $\mathbf{x}$ ) determinado por los índices en  $S$ . Trabajaremos con distribuciones de probabilidad positivas denotadas como  $p(\mathbf{x})$ . De manera similar,  $p(\mathbf{x}_S)$  denotará la distribución de probabilidad marginal para  $\mathbf{X}_S$ . Usamos  $p(x_i | x_j)$  para denotar la distribución de probabilidad condicional de  $X_i$  dado  $X_j = x_j$ .

En física estadística se tratan con frecuencia modelos de sistemas de múltiples partículas que interactúan entre sí. Las propiedades macroscópicas del sistema son determinadas por las interacciones entre los elementos microscópicos. Normalmente, las interacciones entre las componentes del sistema también determinan la probabilidad de la configuración global correspondiente. Cada estado  $\mathbf{x}$  del sistema estadístico tiene una energía asociada  $E(\mathbf{x})$ . Un resultado fundamental de la mecánica estadística es que, en equilibrio térmico, la probabilidad de un estado está dada por la Ley de Boltzmann:

$$p(\mathbf{x}) = \frac{1}{Z(T)} e^{-\frac{E(\mathbf{x})}{T}}$$

donde  $T$  es la temperatura del sistema, y  $Z(T)$  la correspondiente función de partición.

Yedidia y col. [17] propusieron usar el postulado de la ley de Boltzmann para definir la energía correspondiente a una probabilidad conjunta de un sistema no-físico. A la temperatura se le asigna un valor arbitrario, habitualmente  $T = 1$ , el cual simplemente determina una escala para las unidades en que se mide la energía. Propuestas similares se han usado con éxito en optimización [7, 11].

De cara a obtener una aproximación de la energía, es posible utilizar diferentes técnicas para obtener una aproximación de la energía. Kikuchi et. al [6, 10] desarrollaron el método variacional basado en agrupamientos (CVM). En la base del CVM está la idea de aproximar la energía como una función de un conjunto de diferentes marginales. El conjunto no forma necesariamente una factorización exacta de la distribución, pero la forma en que se escogen los marginales influye críticamente en la calidad de la aproximación.

### 2.1. Descomposición de un grafo en regiones

Nuestro objetivo inicial es utilizar los métodos empleados para aproximar la energía en la aproximación de distribuciones de probabilidad. En una segunda etapa se propone un algoritmo para el aprendizaje de este tipo de aproximación directamente de un conjunto de datos. El primer problema a resolver es cómo escoger cuáles son los marginales que participan en la aproximación. Nos concentramos en el problema de encontrar un conjunto apropiado de marginales que permita obtener una aproximación de la distribución.

Un grafo no dirigido  $G = \langle V, E \rangle$  se define como un conjunto no vacío de vértices  $V$ , y un conjunto de aristas  $E$ . Una arista entre los vértices  $i$  y  $j$  será representada como  $i \sim j$ . Dada una distribución de probabilidad  $p(\mathbf{x})$ , su grafo de independencia es un grafo  $G = (V, E)$  que asocia un vértice a cada variable de  $\mathbf{X}$ , y

donde dos vértices están conectados si las variables correspondientes son condicionalmente dependientes dado el resto de las variables.

Definimos una región  $R$  del grafo de independencia  $G = \langle V, E \rangle$  de una distribución de probabilidad  $p(\mathbf{x})$  como un subconjunto de  $V$ . Una descomposición del grafo en regiones  $(\mathcal{R}, U)$ , es un conjunto de regiones que cubre  $V$  y un conjunto asociado de contadores  $U$ , integrado por un contador  $c_R$  para cada región  $R \in \mathcal{R}$ .  $c_R$  será siempre un entero, que puede ser cero o negativo para algunos  $R$ .

La *aproximación Kikuchi de la energía* [6] se define a partir de una descomposición de la energía basada en las regiones de un grafo. Para ser válida una descomposición, debe satisfacer varias restricciones que relacionan las regiones y los contadores. Inspirados en el trabajo de Yedidia y col. [17], llamamos a este subproblema como *problema de encontrar una descomposición válida del grafo en regiones*. En el método CVM,  $\mathcal{R}$  se forma a partir de un conjunto inicial de regiones  $\mathcal{R}_0$  tal que todos los vértices están en al menos una región de  $\mathcal{R}_0$ , y cualquier otra región en  $\mathcal{R}$  resulta de la intersección de una o más de las regiones en  $\mathcal{R}$ . El conjunto de regiones  $\mathcal{R}$  es cerrado bajo la intersección, y puede ordenarse como un conjunto parcialmente ordenado (poset).

Decimos que un conjunto de regiones  $\mathcal{R}$ , y de contadores  $c_R$  es una descomposición del grafo válida cuando para cada variable  $X_i$ ,

$$\sum_{\substack{R \in \mathcal{R} \\ X_i \subseteq \mathbf{x}_R}} c_R = 1 \quad (1)$$

Yedidia y col. [17] introducen el concepto de regiones para definir una aproximación de la energía libre a partir de una descomposición en regiones dada por grafos de factores. Hemos adoptado su definición para explicar el problema más general de hallar una descomposición de nuestro grafo de independencia en regiones. Un aspecto fundamental en la determinación de una descomposición válida es la selección del conjunto inicial de regiones. El conjunto inicial de regiones usado por la comunidad de física estadística comprende usualmente regiones con el mismo tamaño y topología de con-

xióon.

En este trabajo un método particular para seleccionar el primer conjunto de regiones. La manera particular en que se seleccionan estas regiones determina varias propiedades convenientes de la aproximación de la probabilidad. Dado un grafo no dirigido arbitrario  $G$ , formaremos el conjunto de regiones  $\mathcal{R}_0$  tomando una región por cada clique máximo en  $G$ . Como resultado, todas las posibles regiones  $R \in \mathcal{R}$  serán cliques porque son el resultado de la intersección de dos o más cliques. A este tipo de descomposición del grafo en regiones le denominaremos *descomposición del grafo en cliques*.

*Ejemplo 1: Descomposición en regiones válida correspondiente al grafo mostrado en la figura 1)*

$$\begin{aligned} \mathcal{R}_0 &= \{\{1, 2, 5\}, \{1, 3, 6\}, \{1, 2, 3\}, \{2, 3, 4\}, \\ &\quad \{2, 4, 5\}, \{3, 4, 6\}, \{4, 5, 6\}, \{1, 5, 6\}\}, \\ \mathcal{R}_1 &= \{\{1, 2\}, \{2, 5\}, \{1, 3\}, \{3, 6\}, \{2, 3\}, \\ &\quad \{2, 4\}, \{3, 4\}, \{4, 5\}, \{4, 6\}, \{1, 5\}, \\ &\quad \{1, 6\}, \{5, 6\}\} \\ \mathcal{R}_2 &= \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\} \\ c_R &= 1, \forall R \in \mathcal{R}_0; \\ c_R &= -1, \forall R \in \mathcal{R}_1; \\ c_R &= 1, \forall R \in \mathcal{R}_2 \end{aligned} \quad (2)$$

La *aproximación Kikuchi de una distribución probabilidad  $p(\mathbf{x})$* , construida a partir de una descomposición en cliques del grafo de independencia de  $p(\mathbf{x})$  [14], es un tipo particular de factorización. Los marginales en la factorización están completamente determinados por el grafo de independencia. Dado este grafo, la descomposición basada en cliques está formada por los cliques maximales del grafo y sus intersecciones. Estos cliques constituyen las regiones en este tipo de descomposiciones. Los métodos usados para construir la descomposición del grafo en regiones determinan la aproximación Kikuchi final.

Definimos la aproximación Kikuchi de la distribución de probabilidad  $p(\mathbf{x})$ , denotada  $k(\mathbf{x})$  como:

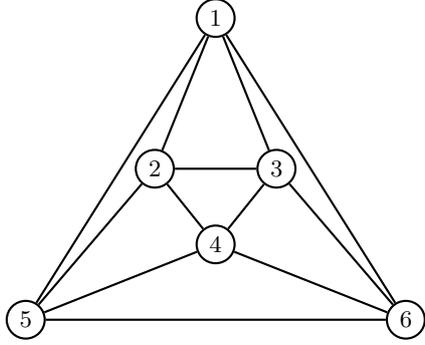


Figura 1: Grafo de independencia

$$k(\mathbf{x}) = \prod_{R \in \mathcal{R}} p(\mathbf{x}_R)^{c_R}, \quad (3)$$

donde  $\mathcal{R}$  se determina a partir de una descomposición del grafo en cliques y los contadores  $c_R$  se calculan utilizando la siguiente fórmula recursiva:

$$c_R = 1 - \sum_{\substack{S \in \mathcal{R} \\ R \subset S}} c_S \quad (4)$$

donde  $c_S$  es el contador de cualquier región  $S$  in  $\mathcal{R}$  tal que  $S$  es un superconjunto de  $R$ . Los valores de  $c_R$  que se corresponden con los cliques maximales iniciales son iguales a 1. Si  $c_R$  es diferente de cero, la región es incluida en la descomposición.

A partir de ahora limitaremos nuestro análisis a las aproximaciones Kikuchi construidas a partir de descomposiciones del grafo en cliques. A menos que se especifique lo contrario, cuando hablamos de una aproximación Kikuchi asumimos que se origina a partir de una descomposición del grafo en cliques.

*Ejemplo 2: Aproximación Kikuchi, construida a partir de la descomposición del grafo en*

*cliques presentada en el ejemplo 1.*

$$L_0 = p(x_1, x_2, x_5) \cdot p(x_1, x_3, x_6) \cdot p(x_1, x_2, x_3) \\ \cdot p(x_2, x_3, x_4) \cdot p(x_2, x_4, x_5) \cdot p(x_3, x_4, x_6) \\ \cdot p(x_4, x_5, x_6) \cdot p(x_1, x_5, x_6)$$

$$L_1 = p(x_1, x_2) \cdot p(x_2, x_5) \cdot p(x_1, x_3) \\ \cdot p(x_3, x_6) \cdot p(x_2, x_3) \cdot p(x_2, x_4) \\ \cdot p(x_3, x_4) \cdot p(x_4, x_5) \cdot p(x_4, x_6) \\ \cdot p(x_1, x_5) \cdot p(x_1, x_6) \cdot p(x_5, x_6)$$

$$L_2 = p(x_1) \cdot p(x_2) \cdot p(x_3) \cdot p(x_4) \cdot p(x_5) \cdot p(x_6)$$

$$k(x) = L_0 \cdot \frac{1}{L_1} \cdot L_2$$

El ejemplo 2 muestra la aproximación Kikuchi que corresponde a la descomposición del grafo en cliques presentada en el ejemplo 1.

Una distribución de probabilidad  $\tilde{p}(\mathbf{x})$  basada en la aproximación Kikuchi puede calcularse normalizando  $k(\mathbf{x})$ .

$$\tilde{p}(\mathbf{x}) = \frac{k(\mathbf{x})}{\sum_{\mathbf{x}'} k(\mathbf{x}')}$$

Sin embargo, la determinación de  $\tilde{p}$  no es una alternativa realista, porque no resuelve el problema de calcular la función de partición  $Z_k = \sum_{\mathbf{x}} k(\mathbf{x})$ .

### 3. Algoritmo para construir una descomposición en cliques válida

Presentamos una panorámica general del algoritmo que construye descomposiciones en cliques. La descomposición final tiene un solo conjunto de regiones  $\mathcal{R}_1$  cuyos contadores  $c_R, R \in \mathcal{R}_1$  son enteros diferentes de cero. El Algoritmo 1 presenta los pasos para encontrar una descomposición en regiones. Sea  $\mathcal{A}$  un conjunto auxiliar de regiones, y  $\Phi(\mathcal{A})$  una función que encuentra todas las intersecciones entre las regiones en  $\mathcal{A}$ .

Primeramente, se construye un conjunto de regiones iniciales  $\mathcal{Q}$ , agregando progresivamente regiones creadas a partir de la intersección entre otras regiones. Cuando todas las posibles intersecciones han sido agregadas,  $\mathcal{Q}$  es ordenado. El orden es determinado por el criterio de inclusión. Si  $R_i \supseteq R_j$  entonces  $R_i$  precede

a  $R_j$  en el orden. Después, cada región  $R$  en  $\mathcal{Q}$  es inspeccionada, y el contador  $c_R$  se calcula como

$$c_R = 1 - \sum_{\substack{S \in \mathcal{R}_1 \\ S \supset R}} c_S \quad (5)$$

donde  $c_S$  es el contador de cualquier región  $S$  en  $\mathcal{R}_1$  tal que  $S$  es un superconjunto de  $R$ . Si  $c_R$  es diferente de cero, la región se agrega a  $\mathcal{R}_1$ .

- 1)  $\mathcal{A} = \mathcal{R}_0$
- 2)  $\mathcal{Q} = \mathcal{A}$
- 3) Do
- 4)  $\mathcal{A} = \Phi(\mathcal{A})$
- 5)  $\mathcal{Q} = \mathcal{Q} \cup \mathcal{A}$
- 6) Until  $\mathcal{A} = \emptyset$
- 7) Encontrar un ordenamiento parcial de  $\mathcal{Q}$
- 8)  $\mathcal{R}_1 = \emptyset$
- 9)  $U = \emptyset$
- 10) For  $R \in \mathcal{Q}$
- 11)  $c_R = 1 - \sum_{\substack{S \in \mathcal{R}_1 \\ S \supset R}} c_S$
- 12) if  $c_R \neq 0$  then
- 13)  $\mathcal{R}_1 = \mathcal{R}_1 \cup R$
- 14)  $U = U \cup c_R$

**Algoritmo 1:** Halla descomposición válida

Todos los cliques maximales en  $\mathcal{R}_0$  estarán en  $\mathcal{R}_1$  con  $c_R = 1$ . El algoritmo encuentra el poset de todas las regiones. Cuando semejante poset es cerrado bajo la intersección de regiones, y los contadores se calculan según (5), la descomposición en regiones es válida [10].

#### 4. Muestreo de Gibbs de la aproximación Kikuchi

Aunque el énfasis fundamental de nuestro trabajo está en el aprendizaje de la aproximación Kikuchi, resulta importante conocer cómo utilizar la información disponible en la aproximación. Una de los modos posibles es partir del muestreo del modelo, el cual permite la generación de soluciones o casos con una distribución similar a la encontrada en el modelo.

Para emplear la aproximación Kikuchi como un generador, haremos un uso conveniente de la propiedad de Markov que satisface

esta aproximación [14, 15]. Como  $k(\mathbf{x})$  no es una distribución de probabilidad, asumimos que los puntos serán generados a partir de su expresión normalizada  $\tilde{p}(\mathbf{x})$ . Para muestrear puntos a partir de  $\tilde{p}(\mathbf{x})$ , usamos el algoritmo de muestreo de Gibbs [3]. Suponemos que aunque  $\tilde{p}(\mathbf{x})$  es demasiado compleja para deducir muestras directamente de ella, sus distribuciones condicionales,  $\tilde{p}(x_i | \mathbf{x} \setminus x_i) = k(x_i | \mathbf{x} \setminus x_i)$  son factibles de calcular.

Para obtener muestras de  $\tilde{p}(\mathbf{x})$ , comenzamos con un estado inicial  $\mathbf{x}^0 = (\mathbf{x}_1^0, \dots, \mathbf{x}_n^0)$ . En cada momento  $t$  del muestreo de Gibbs, una posición  $i$  a ser actualizada será escogida a partir de  $\mathbf{x}^t$ , y el nuevo vector  $\mathbf{x}^{t+1}$  se selecciona usando la regla de transición mostrada en (6).

$$\mathbf{x}^{t+1} = \begin{cases} x_j^{t+1} = x_j^t & \text{for } j \neq i \\ x_j^{t+1} \approx k(x_i^t | bd(x_i^t)) & \text{for } j = i \end{cases} \quad (6)$$

En la ecuación 6,  $bd(x_i^t)$  se refiere a los valores de la variables que forman parte de la frontera de  $X_i$  en el grafo. En el instante  $t + 1$ , el muestreo de  $X_i^{t+1}$  se realiza a partir del muestreo de la aproximación Kikuchi de  $X_i$  dados sus vecinos, de acuerdo a las probabilidades marginales en el instante  $t$ .

Con el objetivo de adaptar la aplicación del muestreo de Gibbs al contexto de la aproximación Kikuchi, definimos  $VS$ ,  $Cy$ , e  $In$  como los parámetros a ser usados por el método.  $VS$  es el tipo de esquema de visita de las variables, y define la manera en que las variables se seleccionan para ser actualizadas. Esquemas de visitas aleatorios ( $VS = 0$ ), o fijos ( $VS = 1$ ) pueden ser utilizados.  $Cy$  es el número de ciclos del muestreo de Gibbs. Un ciclo comprende la actualización de  $n$  variables.  $In$  es un parámetro que determina la manera en que se construye el vector inicial del muestreo de Gibbs. El vector a partir del cual se aplica el puede seleccionarse aleatoriamente ( $In = 0$ ), o muestrearse a partir de una factorización aproximada calculada usando un subgrafo cordal del grafo de independencia ( $In = 1$ ).

## 5. Aprendizaje de la aproximación Kikuchi a partir de los datos

En esta sección se presenta un algoritmo basado en pruebas de independencia para el aprendizaje de la aproximación Kikuchi a partir de los datos.

En las secciones anteriores hemos asumido que los cliques maximales del grafo de independencia, y sus probabilidades marginales son conocidas. En determinados contextos se necesita un algoritmo para aprender la estructura y parámetros del modelo. El Algoritmo 2 es una extensión de un algoritmo introducido en [13] para aprender factorizaciones ordenadas de los datos. Los pasos del uno al tres coinciden con el trabajo anterior, mientras que en los pasos cuatro y cinco se construye la aproximación Kikuchi.

- 1) Aprender un grafo de independencia  $G$  de los datos (el conjunto de las soluciones seleccionadas).
- 2) Si resulta necesario, refinar el grafo.
- 3) Encontrar el conjunto  $\mathcal{C}$  de todos los cliques maximales de  $G$ .
- 4) Construir la descomposición del grafo en cliques.
- 5) Encontrar las probabilidades marginales para las regiones de la descomposición.

**Algoritmo 2:** Aproximación Kikuchi de los datos

La construcción de un grafo de independencia a partir de los datos puede lograrse por medio de pruebas de independencia. Para determinar si una arista pertenece al grafo, se realiza un test de independencia para cada par de variables dado el resto. No obstante, desde un punto de vista algorítmico es importante reducir el costo de las pruebas de independencia. Por esta razón hemos adoptado la metodología seguida previamente por Spirtes [16]. La idea es empezar con un grafo no dirigido completo, y eliminar aristas realizando tests de independencia condicional entre los nodos unidos en el grafo, pero usando conjuntos condicionantes tan pequeños como sea posible.

Para evaluar las relaciones de independencia usamos el test de independencia basado en la distribución Chi-cuadrado. Si dos variables  $X_i$  y  $X_j$  son independientes con un nivel de significación  $\alpha$  previamente especificado, la arista que los une es eliminada del grafo. En el caso general podemos asumir que cada arista  $i \sim j$  en el grafo de independencia inicial es pesada con un valor  $w(i, j)$  que representa la interacción entre el par de variables. Esta información podría estar disponible a partir de información previa, o a partir de las pruebas estadísticas realizadas (el valor del test del Chi-cuadrado). Cuando tal información no está disponible, asumimos que todos los valores de las dependencias son iguales a un parámetro  $w'$  (es decir  $w(i, j) = w', \forall i \sim j \in E$ ).

Cuando el grafo de independencia es muy denso, podemos esperar que la dimensión de los cliques aumentará más allá de un límite factible. Una alternativa para resolver este problema es, en un paso previo al cálculo de los cliques, hacer el grafo menos denso. Hemos seguido esta estrategia poniendo un límite al número máximo de aristas incidentes a cada vértice. Si el vértice tiene más de  $r$  aristas incidentes, aquéllas con los pesos más bajos son eliminadas.

Para encontrar todos los cliques del grafo se usa el algoritmo de Bron y Kerbosch [1]. Este algoritmo usa una técnica de ramificar y acortar para eliminar ramas que puedan conducir a cliques no maximales. Una vez que se han encontrado todos los cliques, se guardan en una lista  $L$ , y sus pesos se calculan a partir de la información sobre las dependencias. El peso de cualquier subgrafo  $G'$  de  $G$  se calcula como  $W(G') = \sum_{i \sim j \in G'} w(i, j)$ . De esta manera se calculan los pesos de los cliques maximales  $w(C_i)$ . La descomposición del grafo basada en cliques se realiza utilizando el algoritmo CVM. Las probabilidades marginales para cada uno de los cliques son calculadas a partir de las frecuencias asociadas a cada configuración, y su posterior normalización. En nuestra implementación, los parámetros del modelo probabilístico pueden ser modificados agregando una perturbación en forma de distribuciones definidas a priori.

## 6. Experimentos

El objetivo de nuestros experimentos es evaluar el comportamiento del algoritmo de aprendizaje de la aproximación Kikuchi y estudiar la exactitud de la aproximación Kikuchi aprendida a partir de los datos. Nos concentramos en un ejemplo de prueba correspondiente a un modelo Ising generalizado [5] definido en una rejilla de dimensiones 4x4, con conexión toroidal. Nuestro problema consiste en aproximar la distribución de un conjunto de estados del modelo Ising. Cada estado se representa por una asignación de valores cero o uno para las 16 variables del sistema.

Un modelo clásico del fenómeno del magnetismo es el *modelo de Ising generalizado*, descrito por la energía funcional (Hamiltoniano) (7) [12]

$$H = - \sum_{i < j \in L} J_{ij} \sigma_i \sigma_j - \sum_{i \in L} h_i \sigma_i \quad (7)$$

$L$  es el conjunto de sitios llamado rejilla. Cada variable  $\sigma_i$  correspondiente al sitio  $i \in L$  toma el valor 1 o  $-1$ , representando el polo sur de un imán pequeño. Una asignación específica de valores para las variables  $\sigma_i$  se llama una configuración. Las constantes  $J_{ij}$  son los coeficientes de la interacción, las constantes  $h_i$  representan las fuerzas del campo magnético externo. El problema de hallar un estado mínimo del modelo de Ising es el problema de hallar cualquier configuración para la cual la energía es mínima.

Cuando todos los  $J_{ij}$  del modelo Ising son no-negativos, el modelo se llama ferromagnético. Más difíciles de tratar son los modelos donde los valores de  $h_i$  y  $J_{ij}$  son arbitrarios. En el caso general este problema es NP-completo.

El problema del modelo Ising generalizado es quizás uno de los ejemplos más claros de la diferencia entre las aproximaciones tradicionales de la probabilidad y la aproximación Kikuchi. Las factorizaciones exactas de funciones descomponibles aditivamente, definidas en rejillas de dimensión 2, requieren el cómputo de distribuciones marginales condicionales de tamaño  $O(\sqrt{n})$ , donde  $n$  es la cantidad de elementos en la rejilla. Como consecuencia, la

complejidad del modelo probabilístico es exponencial en  $n$ .

La aproximación Kikuchi usa los cliques del grafo original. En la rejilla las aristas son los cliques maximales, y tienen tamaño dos. La factorización Kikuchi resulta por lo tanto una manera factible y eficiente de almacenar una factorización de la distribución. Todas las aristas del grafo son usadas en esta factorización.

En nuestro experimento, se evalúa la calidad de la aproximación Kikuchi en un conjunto de datos seleccionados a partir de la evaluación de la energía de los estados del modelo Ising. Se genera un conjunto inicial de 1000 estados (asignaciones de valores) para las 16 variables del problema. Se calcula la energía correspondiente al modelo Ising de estas soluciones, para posteriormente seleccionar las soluciones con mayor energía. Sobre este subconjunto de casos seleccionados se aplica el algoritmo de aprendizaje.

El Cuadro 1 muestra el conjunto de cliques (aristas) aprendidos del conjunto de soluciones seleccionado. Las 19 aristas aprendidas corresponden a interacciones definidas por la topología de conexión de la rejilla, la cual define parcialmente la estructura del problema. Por otro lado, 13 de las 32 interacciones representadas en la estructura de la rejilla no se han capturado en el grafo de independencia. Este hecho puede explicarse porque la existencia de una arista en la estructura no es la única condición para la existencia de dependencias, éstas también se determinan por la fuerza de las interacciones entre las variables.

A partir del grafo de independencia no dirigido, se construyen dos posibles modelos probabilísticos. El primero basado en grafos de cliques y el segundo en la aproximación Kikuchi. Nuevos conjuntos de datos son generados para cada modelo. En el caso de la aproximación Kikuchi, determinamos que el muestreo de Gibbs comience a partir de un punto muestreado de la factorización ordenada ( $In = 1$ ). Para medir la calidad de la aproximación, se calculan los marginales en los dos nuevos conjuntos generados, y se computan los errores en las aproximaciones con respecto a los marginales calculados a partir del conjunto inicial.

$E$	$P$	$W$	$p_{JT}$	$p_{K1}$	$p_{K100}$	$E$	$P$	$W$	$p_{JT}$	$p_{K1}$	$p_{K100}$
(6, 10)	1	19,36	0,028	0,039	0,056	(14, 2)	11	8,55	0,035	0,066	0,059
(6, 5)	2	11,71	0,019	0,091	0,034	(14, 15)	12	18,47	0,034	0,076	0,017
(9, 10)	3	11,22	0,057	0,091	0,056	(3, 2)	13	8,34	0,055	0,029	0,050
(8, 9)	4	17,28	0,019	0,059	0,073	(7, 11)	14	12,77	0,085	0,027	0,049
(8, 12)	5	15,86	0,041	0,009	0,018	(1, 5)	—	9,65	0,339	0,053	0,073
(9, 13)	6	10,12	0,068	0,091	0,033	(3, 0)	—	7,94	0,282	0,084	0,037
(1, 13)	7	15,97	0,029	0,030	0,053	(12, 13)	—	7,32	0,270	0,030	0,051
(1, 2)	8	13,63	0,035	0,025	0,053	(6, 2)	—	6,72	0,283	0,078	0,093
(1, 0)	9	10,55	0,039	0,084	0,053	(4, 8)	—	6,31	0,315	0,062	0,017
(4, 5)	10	9,64	0,013	0,054	0,073	<i>Tot.</i>			2,05	1,09	0,95

Cuadro 1: Errores en la aproximación de los marginales del conjunto seleccionado dados por las aproximaciones válida y Kikuchi.

Sean respectivamente  $p_{ij}^s$  y  $p_{ij}^{new}$  los marginales bivariados de  $X_i, X_j$  calculados a partir del conjunto seleccionado, y de la nueva población. El error en la aproximación se calcula como  $\sum_{x_i, x_j} (p_{ij}^{new} - p_{ij}^s)$ .

En el cuadro 1,  $p_{JT}$  representa el error dado por la aproximación válida para los diferentes marginales y  $P$  es un ordenamiento asociado a las aristas incorporadas al modelo. Los errores son pequeños para los marginales que corresponden a las aristas numeradas, pero notablemente peores para las aristas restantes. Este hecho ocurre porque el árbol de cliques no puede cubrir estas interacciones y usarlas en el muestreo. Después de sólo un ciclo del muestreo de Gibbs, la aproximación Kikuchi consigue resultados ( $p_{K1}$ ) mejores que la aproximación del grafo de cliques ordenado. Para la aproximación Kikuchi no hay diferencias importantes entre los errores de la aproximación para los diferentes cliques. En el cuadro,  $p_{K100}$  muestra el error cuando el número de pasos de muestreo es aumentado a 100. Puede observarse que el error global disminuye. No obstante, el error en la aproximación de algunas aristas puede aumentar, probablemente debido a los errores de muestreo del método de muestreo de Gibbs.

Resumiendo, los resultados del experimento ilustran que cubriendo un mayor número de dependencias que el grafo de cliques, la aproximación Kikuchi puede ser más precisa, y que el algoritmo es capaz de recuperar la estructu-

ra del problema.

## 7. Conclusiones

En este trabajo hemos presentado la aproximación Kikuchi como una alternativa para aproximar distribuciones de probabilidad. Hemos introducido un algoritmo para el aprendizaje de la aproximación Kikuchi a partir de los datos y presentado un ejemplo de aplicación del mismo en datos generados a partir de una instancia del modelo Ising de ferromagnetismo. Los resultados demuestran que la aproximación es capaz de capturar relaciones de dependencias existentes en los datos.

## 8. Agradecimientos

Este trabajo se ha beneficiado de fondos de los proyectos Saiotek S-PEO4UN25 y Etortek-Biolan del Gobierno Vasco.

## Referencias

- [1] C. Bron and J. Kerbosch. Algorithm 457—finding all cliques of an undirected graph. *Communications of the ACM*, 16(6):575–577, 1973.
- [2] L. M. de Campos and J. F. Huete. Algorithms for learning decomposable models and chordal graphs. Technical Re-

- port DECSAI 970213, University of Navarra, 1997.
- [3] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and Bayesian restoration of images. *IEEE transactions on pattern analysis and Machine Intelligence*, (6):721–741, 1984.
- [4] D. Heckerman, D. Geiger, and D. M. Chickering. Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, 20:197–243, 1995.
- [5] E. Ising. The theory of ferromagnetism. *Zeitschrift fuer Physik*, 31:253–258, 1925.
- [6] R. Kikuchi. A theory of cooperative phenomena. *Physical Review*, 81(6):988–1003, 1951.
- [7] S. Kirkpatrick, C. D. J. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, May 1983.
- [8] P. Larrañaga, J. Lozano, and H. Mühlenbein. Algoritmos de estimación de distribuciones en problemas de optimización combinatoria. *Revista Iberoamericana de Inteligencia Artificial*, 19(2):149–168, 2003.
- [9] P. Larrañaga and J. A. Lozano, editors. *Estimation of Distribution Algorithms. A New Tool for Evolutionary Computation*. Kluwer Academic Publishers, Boston/Dordrecht/London, 2002.
- [10] T. Morita. Formal structure of the cluster variation method. *Progress of Theoretical Physics Supplements*, 115:27–39, 1994.
- [11] H. Mühlenbein, T. Mahnig, and A. Ochoa. Schemata, distributions and graphical models in evolutionary optimization. *Journal of Heuristics*, 5(2):213–247, 1999.
- [12] B. Naudts and J. Naudts. Fitness landscapes and problem difficulty: the effect of spin-flip symmetry on the performance of the simple ga. In *Parallel Problem Solving from Nature - PPSN V International Conference*, pages 67–76, Amsterdam, The Netherlands, 1998. Springer Verlag. LNCS 1498.
- [13] R. Santana. A Markov Network based Factorized Distribution Algorithm for optimization. In *Proceedings of the 14th European Conference on Machine Learning (ECML-PKDD 2003)*, volume 2837 of *Lecture Notes in Artificial Intelligence*, pages 337–348, Dubrovnik, Croatia, 2003. Springer-Verlag.
- [14] R. Santana. Estimation of distribution algorithms with Kikuchi approximations. *Evolutionary Computation*, 13(1):67–97, 2005.
- [15] R. Santana, P. Larrañaga, and J. A. Lozano. Markov and decomposability properties of the Kikuchi approximation. Submitted for publication, 2005.
- [16] P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction and Search*, volume 81 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1993.
- [17] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Constructing free energy approximations and generalized belief propagation algorithms. Technical Report TR-2002-35, Mitsubishi Electric Research Laboratories, August 2002.