Lógica y Probabilidad

en

Representación de Conocimiento

Leopoldo Bertossi

Carleton University
School of Computer Science
Ottawa, Canada

Probabilidad

- En nuestros estudios usualmente nos enfrentamos con las probabilidades a través de los axiomas que rigen la teoría matemática de probabilidad

- Esos axiomas nos dicen qué propiedades matemáticas tiene una distribución de probabilidad; y de ahí parte la teoría

- Esta teoría matemática es amplia, profunda, rica en problemas y desarrollos, y útil en aplicaciones

- Sin embargo, esta teoría matemática no entra en terrenos ontológicos ni epistemológicos

- En particular, no dice qué es una aseveración probabilística, ni cuál es su naturaleza, ni de dónde emerge una medida de probabilidad, ni por qué esos axiomas tienen que ser aceptados, etc.

Los Fundamentos

- El área de los “fundamentos de la probabilidad” toca precisamente preguntas como las anteriores

- Se puede entender como una visión filosófica de la probabilidad; y se puede estudiar también en forma matemática

- Como es usual en filosofía, hay varias respuesta a todas estas preguntas en torno al concepto de probabilidad

  Todas susceptibles de ser desarrolladas matemáticamente

- Una situación similar se da en Estadística; por varias razones, en particular por sus bases probabilísticas

- Ha habido y hay diversas “escuelas de probabilidad”

- Probabilidad subjetiva (B. De Finetti, ...)
- Probabilidad frecuentista (R. von Mises, ...)
- Probabilidad lógica (R. Carnap, ...)
- Probabilidad fáctica (K. Popper, ...) Etc.

- No necesariamente excluyentes; se dan a distintos niveles; y pueden converger [9, 1]
- Estas pueden conducir a los axiomas de probabilidad
- E.g., un agente puede asignar probabilidades en forma subjetiva, y "corregirlas" por condicionalización (à la Bayes)

  Si el agente es "racional" al asignar probabilidades, la medida de probabilidad generada va, demostradamente, a satisfacer los axiomas de probabilidad [4]
- Algo similar ocurre con la interpretación lógica [2] Etc.

Rolando Chuaqui

- Rolando trabajó en lógica matemática y en fundamentos de probabilidad

- Se interesó sistemáticamente en los fundamentos de los métodos estadísticos

  Estos y la probabilidad juegan un rol fundamental en los procesos de diagnóstico, en los cuales tenía un interés especial (y natural)

- Propuso y desarrolló una concepción fáctica de probabilidad

  Informalmente, una medida de probabilidad emerge como posible sobre la base de (o está determinada por) un conjunto de simetrías del fenomeno -físico, material- aleatorio en cuestión

  ... como las probabilidades asociadas a un dado simétrico ...

- También estableció las conexiones con las otras concepciones de la probabilidad, en especial con la concepción lógica

- Desarrolló la probabilidad fáctica con la visión y los métodos de la lógica matemática

- Su investigación en torno a los fundamentos de la probabilidad y métodos estadísticos (FPME) está presente en su libro:

R. Chuaqui. *Truth, Possibility and Probability: New Logical Foundations of Probability and Statistical Inference*. North-Holland, 1991.

- Creo que la investigación en torno a estos temas era parte fundamental, substancial de su vida

- Como buen científico persiguió y desarrolló su proyecto personal

  No fue un seguidor ni un investigador incremental

  Su investigación respondió a su propio plan científico/filosófico

- Por eso creó escuela, y lo recordamos en estas Jornadas

- Yo me interesé en su investigación en FPME, algo hice en el área, incluso trabajamos juntos

  Primero como alumno, luego como colaborador

  ¡Y fue una extraordinaria experiencia científica!

- El interés que él despertó en mí por estos temas me acompaña hasta hoy

  Y se lo agradezco ...

  Y está en las bases de algunas de mis actividades de investigación actuales

  ¿Por qué?

La Conexión con AI y un momento personal ...

- Terminado mi doctorado, me dediqué a la investigación en ciencia de computación, en especial, a representación de conocimiento (KR) en inteligencia artificial (AI)

- El dia 22 de Abril de 1994, Rolando y yo almorzamos juntos, en compañía de Jacques Stern, quien lo visitaba

  Por iniciativa de Rolando ...

- El creía -correctamente- que su obra en FPME tenía un rol importante que jugar en AI

  Me invitaba a colaborar para hacer trascender su investigación en el área hacia la comunidad de AI y tener un impacto en ella

- Lamentablemente no pudimos dar ni el primer paso en esa dirección

Un Poco de KR (clásico)

- KR tiene que ver con: [20]
  - Representar conocimiento en computadores
  - Habilitar computadores para "razonar" con ese conocimiento
- Típicas áreas -tradicionales- de KR (entre otras):
  1. Diagnóstico automatizado
  2. Razonamiento en presencia de incertidumbre [7] (uncertain reasoning) : conocimiento ambiguo, conocimiento incompleto, conocimiento inconsistente, toma de decisiones, etc.
  3. Razonamiento con sentido común (puede caer en 2.)
  4. Razonamiento sobre causalidad (en particular, acción/efecto)
  5. Razonamiento probabilístico
- 1.-4. no necesaria o intrínsecamente de naturaleza probabilística

Ejemplo 1: Razonamiento de sentido común

Conocimiento: *“Piolín es un pájaro”* (*)

Conclusión: *“Piolín vuela”*

Ejemplo 2: Razonamiento de diagnóstico

Conocimiento: *Gripe* $\Rightarrow$ *Fiebre*

Fiebre (evidencia, observación)

Conclusión: *Gripe*

- ¿Cómo se interpreta el conocimento explícito?
  ¿Cómo se infiere a partir de él?
- Muchas aristas, problemas, complicaciones, desafíos, ...

Un enfoque “lógico”:

- 1. - 4. han sido abordados usando métodos de “lógica” simbólica
- No son lógicas clásicas, porque la forma de razonamiento asociado es “no-clásico”
- Posible interpretación del uso de (*):

  Se le aplica una “regla por defecto” (default rule)

“En ausencia de información contraria, concluya que el pájaro vuela ...”

“En ausencia de información de que el pajaro es excepcional, concluya que el pájaro vuela ...”

En términos más "lógicos":

$$KB = \{\, \forall x(Pajaro(x) \wedge \neg Anormal(x) \rightarrow Vuela(x)),\;\; Pajaro(piolin)\, \}$$

"Como no hay evidencia de que Piolín sea anormal, conclúyase que vuela ..."

- ¿Qué lógica da cuenta de algo así? Definitivamente no-clásica ...

  Es una lógica "no-monótona": Nuevo conocimiento inserto en una base de conocimiento puede invalidar conclusiones previas

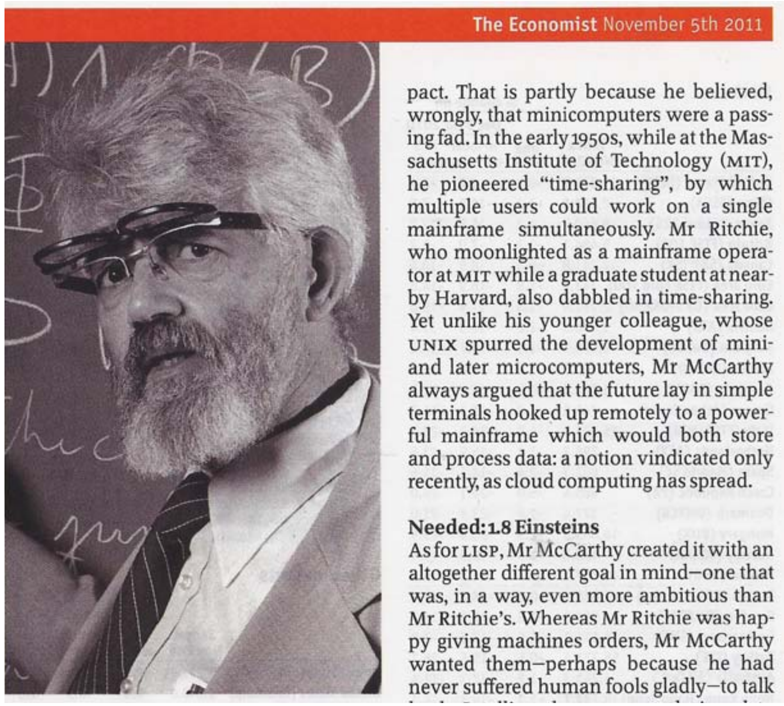
  Por ejemplo, que Piolín es un emu (y éstos son considerados anormales) ...

  En contraste con la lógica clásica, que es monótona

- ¡El (conocimiento y razonamiento de) sentido común nos permite funcionar en ausencia de conocimiento completo!

- De acuerdo con John McCarthy (1927 - 2011, Turing Award 1971), “el padre de la AI”, es posible usar métodos lógico-formales para atacar estos problemas

  Y el problema de razonamiento de sentido común es el gran problema de la AI (common-sense reasoning) [12]



The Economist November 5th 2011

pact. That is partly because he believed, wrongly, that minicomputers were a passing fad. In the early 1950s, while at the Massachusetts Institute of Technology (MIT), he pioneered “time-sharing”, by which multiple users could work on a single mainframe simultaneously. Mr Ritchie, who moonlighted as a mainframe operator at MIT while a graduate student at nearby Harvard, also dabbled in time-sharing. Yet unlike his younger colleague, whose UNIX spurred the development of mini- and later microcomputers, Mr McCarthy always argued that the future lay in simple terminals hooked up remotely to a powerful mainframe which would both store and process data: a notion vindicated only recently, as cloud computing has spread.

Needed: 1.8 Einsteins

As for LISP, Mr McCarthy created it with an altogether different goal in mind—one that was, in a way, even more ambitious than Mr Ritchie's. Whereas Mr Ritchie was happy giving machines orders, Mr McCarthy wanted them—perhaps because he had never suffered human fools gladly—to talk

- De hecho, él mismo propuso una lógica formal para estos efectos: “circumscription”

  Esa lógica captura la idea de que “el predicado de anormalidad debe ser minimizado (circunscrito) al momento de razonar, haciendo que contenga sólo aquello que está obligado a contener”

  ... Por eso Piolín queda fuera de él ...

- Un enfoque lógico alternativo fue propuesto por Raymond Reiter (1939-2002, IJCAI Award 1993), quien inventó la "default logic" [17] (con enorme impacto, incluso indirecto, en KR)

Regla de deducción: $$\frac{Pajaro(x) \ : \ Vuela(x)}{Vuela(x)}$$

"Dado que x es un pájaro, si se puede suponer consistentemente que x vuela, conclúyase que x vuela"

- El problema de diagnóstico en AI es también susceptible de ser tratado en forma lógico-formal (con enormes contribuciones de Reiter [18])

  Por ejemplo, a través de un enfoque abductivo:

  Dado un cuerpo de conocimiento K, y una observacion O, se busca una explicación E, expresada en términos lógicos, tal, que:

  $$K \wedge E \Rightarrow O$$ (usualmente con una condición de minimalidad sobre E)

- John McCarthy y Ray Reiter fueron los principales representantes de la “escuela lógico-formal” en AI
- Otras escuelas? Alternativas?

Un enfoque “probabilístico”:

- Los problemas 1.-4. también pueden tener elementos probabilísticos o ser tratados como problemas probabilísticos, aun sin probabilidades explícitas iniciales (cf. pag. 9)
- Por ejemplo, una regla por defecto puede ser tratada como una aseveración probabilística o estadística (con alta probabilidad)

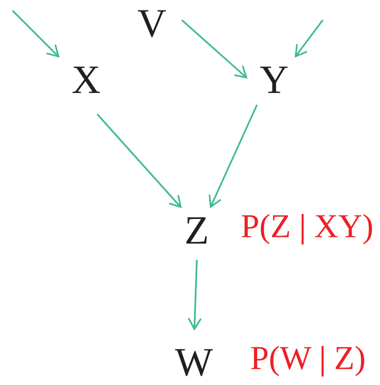
  Y las consequencias también pueden ser probabilísticas
- O como probabilidades condicionales: $P(vuela \mid pajaro) = 0.95$

  “la probabilidad de volar dado que se es pájaro es 0.95”

  ¿Cómo se representan las probabilidades condicionales?

- Aparecen las “redes Bayesianas” [3]

  Cómo representar conocimiento probabilístico; en particular, sobre propagación de probabilidades condicionales



  El modelo probabilístico surge de la estructura de la red más probabilidades condicionales, y otras suposiciones

  - El modelo es una distribución de probabilidad conjunta P sobre todas las variables
  - La estructura puede proveer elementos causales? fácticos? ...
  - Las probabilidades condicionales pueden ser subjetivas
  - Las suposiciones adicionales del tipo: la probabilidad condicional de un nodo dados sus nodos-padre no depende de sus nodos no-descendientes (red localmente Markoviana)

- Y redes de Markov (MNs): como arriba, pero no dirigidas; y representan probabilidades conjuntas (join probabilities)

- Incluso, las probabilidades pueden ser simbólicas, cualitativas, y su propagación y comportamiento en el límite son lo más relevante: $P(Vuela \mid Pajaro) \geq 1 - \epsilon$

- Diagnóstico abductivo también puede ser probabilístico/estadístico

  Dada una distribución de probabilidad, podemos, por ejemplo, buscar una explicación E de máxima verosimilitud:

$$\text{arg max}_E\, P(O \mid K \wedge E)$$

  (más ejemplos más adelante ...)

La Importancia de las Escuelas

- Tradicionalmemte ha habido una “escuela probabilística” en AI (con diferencias de enfoque internas)

- En cierto modo, competidora de la escuela lógica por un cierto tiempo

- Máximo exponente de la “escuela probabilística” ha sido Judea Pearl (*1932, Turing Award 2011)

  Su enfoque es más bien Bayesiano ... [14]

  El mayor contribuyente a la investigación sobre causalidad [16]

  Con impacto en AI, estadística, economía, etc. etc.

  Pearl adopta enfoques lógicos y probabilísticos, y los relaciona

- Sin embargo, el enfoque probabilístico de KR naturalmente genera “sub-escuelas”

- Distintas escuelas probabilísticas puede dar distintas respuestas a preguntas como:

  - ¿De dónde salen las probabilidades?
  - ¿Asignadas en forma subjetiva?
  - ¿Emergen de la estructura del problema?
  - ¿Se obtienen de experimentos? Etc.

- En esta dirección surge un tema fundamental: Aprendizaje de máquina (ML) ... ya que, al final, genera probabilidades ...

  Por ejemplo, una red Bayesiana puede ser “aprendida” (creación de un modelo probabilístico) Etc.

- Sin embargo, ML no se sustrae a las visiones de diversas escuelas probabilísticas
- Todo esto redunda en que las escuelas probabilísticas influyen en la forma en que se hace KR y AI
- En AI no basta con usar la teoría matemática de probabilidad (aquella que parte de los axiomas)

  ¡La introducción de probabilidad en AI, por la naturaleza del área, no ha podido sustraerse a la concepción de la probabilidad que puede tener un agente inteligente, ni a la forma en que la usa!

  Los FPME están en las raíces mismas del uso de probabilidad en AI, y tienen presencia directa y práctica

- La visión filosófica de probabilidad y estadística afecta el enfoque adoptado en AI y los métodos aplicados, e.g. en:
  - diagnóstico
  - toma de decisiones
  - aprendizaje de máquina
  - causalidad
  - sentido común
  - razonamiento probabilístico, etc.
- En todas ellas la discusión filosófica ilumina y determina:
  - qué significa una probabilidad asignada
  - de dónde sale
  - cómo cambia
  - cómo se aprende
  - qué significa una conclusión probabilística
  - etc.

- Los trabajos que unifican o conectan distintos enfoques probabilísticos son de crucial importancia
- Tradicionalmente, las escuelas “lógica” y “probabilística” han sido alternativas y competidoras en AI
- En los últimos años se han convertido en complementarias

  Hoy se ataca los problemas de KR con modelos matemáticos y técnicas que involucran simultáneamente a la lógica (más bien clásica) con la probabilidad

Lógica $+$ Probabilidad en AI

- Diversas formas de KR que combinan lógica y probabilidad

  Cada una con un dominio de aplicaciones subentendidas

  Sólo un par de enfoques, para dar la idea ...

Distribuciones de ME:

- ¿Consecuencias lógicas a partir de una base de conocimiento KB?
- En KB hay conocimiento "duro" (e.g. $emu \rightarrow pajaro$) y reglas condicionales "blandas", probabilísticas, de la forma r: $(\alpha|\beta)[p]$,

  E.g. $r_v: \ (vuela|pajaro)[0.9]$
- Hay una colección $\mathcal{C}$ de mundos, $\mathcal{W}$, asociados a KB que satisfacen el conocimiento duro, pero no necesariamente $\beta \rightarrow \alpha$ (condicionales como implicaciones clásicas)

- También hay varias distribuciones de probabilidad (candidatas) P sobre $\mathcal{C}$: $\mathcal{W} \mapsto P(\mathcal{W})$

  Nos restringimos a la colección $\mathcal{P}$ de distribuciones P que satisfacen la KB, en particular, las reglas condicionales r: $(\alpha|\beta)[p]$

  Por ejemplo, debe cumplirse:[1] $P(\alpha|\beta) = p$ (y $P(\beta) > 0$)

- Las consecuencias lógico-probabilísticas de KB son aquellas con alta probabilidad

  ¿Pero con cuál de las distribuciones en $\mathcal{P}$?

  Puede haber algunas mejores que otras ...

[1] $P(\alpha|\beta) := \frac{P(\alpha \wedge \beta)}{P(\beta)} := \frac{P(\{\mathcal{W} \mid \mathcal{W} \models \alpha \wedge \beta\})}{P(\{\mathcal{W} \mid \mathcal{W} \models \beta\})}$

- Elegimos una que tenga máxima entropía:

$$P^\star := \text{arg max}_{P \in \mathcal{P}} \ Entropy(P)$$
$$= \text{arg max}_{P \in \mathcal{P}} \ -\sum_{\mathcal{W} \in \mathcal{C}} P(\mathcal{W}) \times ln(P(\mathcal{W}))$$

La distribución con menos suposiciones arbitrarias o máximo desorden, menos estructura arbitraria, máxima independencia, menos información gratuita: [8, 13]

- Las consecuencias lógico-probabilísticas de la KB son aquellas que tienen alta probabilidad bajo $P^\star$

$$KB \ \models_{\text{l-p}} \ \varphi \quad :\Longleftrightarrow \quad P^\star(\varphi) \geq 1 - \epsilon$$

- El cómputo de consecuencias de una KB usando una distribución de máxima entropía puede automatizarse [10]

- Y es una elección es filosóficamente interesante [15]; y en la dirección de la probabilidad lógico-fáctica [1]

Markov Logic Networks: (MLNs) [19]

- Las MLNs combinan lógica de primer orden y MNs en la misma representación

- Se tiene una base de conocimiento KB en lógica de primer orden

- Las interpretaciones (modelos, mundos, ...) tienen probabilidades asociadas

- Puede haber “modelos” más probables que otros

- Un mundo que viola una fórmula no es inválido (no-modelo), sino sólo menos probable

- Se asigna pesos w_i a las fórmulas F_i en KB $\quad (i = 1, \ldots, F)$

  A mayor peso, mayor es la diferencia entre un mundo que satisface la fórmula y uno que no lo hace (con todo lo otro igual)

Formula	Weight
$\forall x : Steal(x) \Rightarrow Prison(x)$	3
$\forall x \forall y : CrimePartners(x, y) \wedge Steal(x) \Rightarrow Prison(y)$	1.5

- El peso de una fórmula captura la forma en que la probabilidad decrece cuando una "instancia" de la fórmula es violada

  Hay un dominio fijo, finito, de constantes, con el cual se puede instanciar los predicados y fórmulas, e.g. $Dom = \{bob, anna, ...\}$

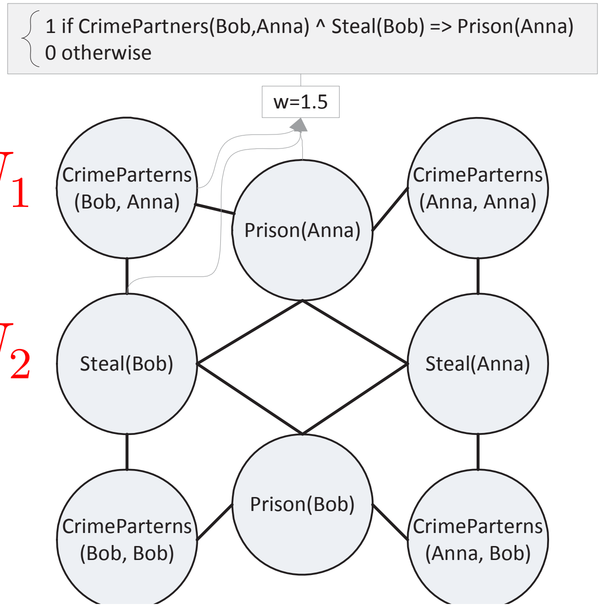
  Produciendo átomos y fórmulas instanciadas (de tipo proposicional), e.g. $CrimePartners(bob, anna)$, $Steal(bob) \rightarrow Prison(bob)$, resp.

- Un mundo es un conjunto de átomos (una estructura de Herbrand), e.g. $\mathcal{W}_1 = \{CrimePartners(bob, anna), Steal(bob)\}$

$\mathcal{W}_2 = \{CrimePartners(bob, anna), Steal(bob), Prison(anna)\}$

N_1

N_2



Los átomos instanciados son los M nodos de la red (no dirigida)

Cada nodo N toma valor 0 o 1 dependiendo de si es falso o verdadero (en un mundo $\mathcal{W}$)

Cada nodo N es una variable aleatoria de Bernoulli X^N

$\mathcal{W}_1$ representado por vector aleatorio: $\mathcal{X} = \langle X^{N_1}, X^{N_2}, X^{N_3}, \ldots \rangle$
$= \langle 1, 1, 0, \ldots, 0 \rangle$

- Arista entre dos nodos si aparecen en una misma fórmula instanciada

- Como en ML, la representación tiene “features” medibles asociadas
- Cada instanciación de una fórmula genera una “feature”, una variable medible, con valor 1 si es verdadera en el mundo $\mathcal{W}$, y 0 si no
- Todos estos ingredientes más una “función de potencial” (para generar distribuciones de probabilidad conjuntas) dan la probabilidad del mundo $\mathcal{W}$ asociado a $\mathbf{x} \in \{0,1\}^M$:

$$P(\mathcal{X} = \mathbf{x}) = \frac{1}{Z} exp(\sum_{i=1}^{F} w_i \times n_i(\mathbf{x}))$$

$n_i(\mathbf{x})$: número de instanciaciones de F_i verdaderas en mundo $\mathbf{x}$

Z normaliza considerando todos los mundos posibles:

$$Z = \sum_{\mathbf{z} \in \{0,1\}^M} exp(\sum_{i=1}^{F} w_i \times n_i(\mathbf{z}))$$

Aquí sólo comienza la diversión ...

- ¿Cómo inferir usuando MLNs?
- ¿Cómo aprender una MLN? (ML) [5]
- ¿Cómo hacer “lifted” inference? Sin instanciar ... [11, 6]

Conclusiones

- Los problemas y discusiones en torno al concepto y al uso de la probabilidad son de larga data

  En algunos casos, de siglos (aprendizaje inductivo, causalidad, etc.)

  Ellos permean toda las actividades científicas y tecnológicas

  Y han desafiado la curiosidad humana y la necesidad de comprender

- Rolando Chuaqui consagró gran parte de su vida científica a investigar los fundamentos filosóficos, lógicos, y matemáticos de la probabilidad y los métodos estadísticos

  Y tuvo conciencia de su relevancia en la AI

- La concepción y uso de la probabilidad y los métodos estadísticos tienen relevancia directa y práctica en áreas con la AI, y KR, en particular

  Hay modelos y métodos que parten de ciertas posiciones específicas al respecto

- Las escuelas lógica y probabilística en AI siguieron cursos paralelos por largo tiempo

  Hoy vemos una convergencia y combinación de los enfoques lógico y probabilístico en KR

  Y el futuro está lleno de desafíos científicos interesantes ...

http://www.scs.carleton.ca/∼bertossi

Email: bertossi@scs.carleton.ca

References

[1] Timothy Childers. *Philosophy & Probability*. Oxford U.P., 2013.

[2] Richard T. Cox. *The Algebra of Probable Inference*. The John Hopkins Press, 1961.

[3] Adnan Darwiche. Bayesian Networks. *Communications of the ACM*, 2010, 53(12):80-90.

[4] Bruno De Finetti. Foresight: Its Logical Laws, Its Subjective Sources. En *Studies in Subjective Probability*, H. Kyburg (ed.), Krieger Pub Co. 1980.

[5] Pedro Domingos. A Few Useful Things to Know about Machine Learning. *Communications of the ACM*, 2012, 55(10):78-87.

[6] Eric Gribkoff, Guy Van den Broeck and Dan Suciu. Understanding the Complexity of Lifted Inference and Asymmetric Weighted Model Counting. CoRR abs/1405.3250 (2014)

[7] Joseph Y. Halpern. *Reasoning about Uncertainty*. The MIT Press, 2003.

[8] E.T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge U.P., 2003.

[9] Richard Jeffreys. *Probability and the Art of Judgment*. Cambridge U.P., 1992.

[10] Gabriele Kern-Isberner, Christoph Beierle, Marc Finthammer and Matthias Thimm. Probabilistic Logics in Expert Systems: Approaches, Implementations, and Applications. Proc. DEXA'11, Springer LNCS 6860, 2011, pp. 27-46.

[11] Kristian Kersting. Lifted Probabilistic Inference. Proc. ECAI 2012, pp. 33-38.

[12] John McCarthy. Generality in Artificial Intelligence. *Communications of the ACM*, 1987, 30(12):1029-1035.

[13] Jeff B. Paris. *The Uncerratin Reasoner's Companion: A Mathematical Perspective*. Cambridge U.P., 1994.

[14] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, revised 2nd printing, 2009.

[15] Jeff B. Paris. Common Sense and Maximum Entropy. *Synthese*, 1999, 117, pp. 75-93.

[16] Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge U.P., 2nd ed., 2009.

[17] Raymond Reiter. A Logic for Default Reasoning. *Artificial Intelligence*, 1980, 13(1-2):81-132.

[18] Raymond Reiter. A Theory of Diagnosis from First Principles. *Artificial Intelligence*, 1987, 32(1):57-95.

[19] Matthew Richardson and Pedro Domingos. Markov Logic Networks. *Machine Learning*, 2006, 62(1-2):107-136.

[20] Frank Van Harmelen, Vladimir Lifschitz and Bruce Porter (eds.). *Handbook of Knowledge Representation*. Elsevier Science, 2008.