

BAYESIAN MODELS OF OBJECT PERCEPTION

DANIEL KERSTEN

ALAN YUILLE

2003

A pesar de la ambigüedad en la información visual, el sistema visual es capaz de llegar con certeza a perceptos únicos, lo que disfraza la complejidad del proceso.

Los modelos Bayesianos permiten dividir los problemas de percepción en una clase limitada de categorías que se ajustan a un marco de referencia teórico que puede lidiar con las complejidades de las imágenes naturales en estudios de visión de computadoras.

El marco de referencia Bayesiano se basa en la noción de Helmholtz de “inferencia inconsciente”. Utiliza la teoría de probabilidad Bayesiana, en la cual el conocimiento previo de escenas visuales se integra con las características de la escena para inferir la interpretación más probable de la imagen. Con ella se derivan modelos de “observadores ideales”, con los que se puede comparar el desempeño humano con la información requerida para realizar una tarea visual. Los observadores humanos se comportan de forma muy cercana a la óptima en varias tareas visuales.

La aproximación Bayesiana a la percepción humana de objetos ha avanzado en dos frentes: el análisis de la estadística del mundo real, y la categorización y mejor entendimiento de los problemas de inferencia. La inferencia Bayesiana de propiedades de los objetos depende de descripciones probabilísticas de las características de las imágenes en función de sus causas en el mundo, y también depende de descripciones a priori de estas mismas causas, independientemente a las imágenes.

REAL-WORLD STATISTICS

Son necesarias regularidades estadísticas para manejar la complejidad de las imágenes. Propiedades como la forma, material o iluminación llevan a regularidades estadísticas (*e.g.*, píxeles adyacentes suelen)tener la misma intensidad luminosa. Se ha encontrado que la estadística puede dar una caracterización eficiente de texturas homogéneas, contribuir al reconocimiento de escenas, encontrar fronteras en los objetos o delimitar el contexto del reconocimiento de objetos.

Relacionar la intensidad de las imágenes con estadística de profundidad de superficie medida ha llevado a soluciones en visión de computadoras del reconocimiento facial. Aun no es claro cómo el sistema visual humano aprende los prior estadísticos apropiados, pero algunos parecen estar anclados en la genética.

BASIC BAYES

La fórmula de Bayes de inferencia inversa es

$$p(S|I) = \frac{p(I|S)p(S)}{p(I)}$$

donde $p(I|S)$ es el modelo para formar la imagen I , y $p(S)$ es el conocimiento previo sobre las estructuras que ocurren naturalmente S . Tanto $p(S)$ como $p(I|S)$ se pueden aprender de estadística del mundo real.

Los observadores ideales infieren el valor más probable para S dado I , es decir, el valor de S que da el máximo valor de $p(S|I)$. El desempeño de los observadores ideales es determinado por la información de la tarea visual y puede usarse como comparativo para el desempeño humano.

Como consecuencia a la fórmula de Bayes, la percepción es un *trade-off* entre la confiabilidad de las características de la imagen ($p(I|S)$) y el prior $p(S)$. Algunas percepciones se basarán más en los priors y otras en los datos.

Para profundizar el entendimiento se categorizan los problemas Bayesianos con gráficas o diagramas de influencia. Se descompone el mundo natural S en componentes S_1, S_2, S_3, \dots , y la imagen I en características I_1, I_2, I_3, \dots , y se expresa la distribución conjunta como $p(S, I) = p(s_1, S_2, \dots, I_1, I_2, \dots)$. Esto permite caracterizar cómo las variables del mundo se influyen entre sí. En el peor caso todas las variables están conectadas. Estos diagramas de influencia permiten categorizar a los problemas de inferencia visual en términos de cómo las características de la imagen se generan. Se describen tres categorías importantes de problemas de inferencia visual.

DISCOUNTING

¿Cómo es que hay en la inferencia de los objetos a pesar de variables de confusión, como cambio en el punto de vista, iluminación, etc? Estas variables son análogas al ruido en la teoría de detección de señales, pero son más complicadas de modelar y afectan a la formación de imágenes de forma no lineal.

En perspectiva Bayesiana, para modelar estos problemas se debe definir una distribución conjunta $p(S_1, S_2, I_1)$ donde S_1 sea el objeto blanco (*i.e.*, una bicicleta), S_2 sea la variable de confusión (*i.e.*, el punto de vista), y I sea la imagen. Se descuentan las variables de confusión al sacarlas mediante integración:

$$p(S_1, I_1) = \sum_{S_2} p(S_1, S_2, I_1)$$

Descontar la iluminación puede reducir la ambigüedad con respecto a la localización o la forma de un objeto.

Descontar una variable es equivalente a asumir que tiene tan poca utilidad que no vale la pena estimarla. Una alternativa es estimar con precisión la variable de confusión que explica una imagen particular. Esto puede llevar al fenómeno “*explainig away*” discutido más adelante. La elección de las variables a descontar dependerá de la naturaleza de la tarea en cuestión.

CUE INTEGRATION

La visión integra claves distintas sobre la forma, movimiento y profundidad de los objetos. Éstas se pueden modelar desde una perspectiva Bayesiana.

Cuando se tienen dos claves con estimados \hat{S}_1 y \hat{S}_2 para cada una (es decir, \hat{S}_1 es el mejor estimador para $p(S|I_1)$). Entonces, la integración óptima (el valor más probable) de los dos estimados toma en cuenta la incertidumbre causada por el ruido en la medición (las desviaciones estándar, σ_1 y σ_2), y está dada por el promedio ponderado,

$$\hat{S}_{optimal} = \hat{S}_1 \frac{r_1}{r_1 + r_2} + \hat{S}_2 \frac{r_2}{r_1 + r_2}$$

donde r_i , la confiabilidad, es el recíproco de la varianza ($\frac{1}{\sigma_i^2}$). Este modelo ha sido usado para estudiar si el sistema visual humano combina las claves de forma óptima. Un modelo Bayesiano de integración óptima de movimiento ha permitido dar explicación de múltiples ilusiones y datos experimentales.

PERCEPTUAL ‘EXPLAINING AWAY’

La ambigüedad visual puede reducirse por medidas auxiliares que pueden estar disponibles en una imagen o se pueden buscar activamente. Estas medidas pueden alterar el concepto al explicar los efectos de las variables de confusión. El cambio en la probabilidad de una hipótesis que compite altera la probabilidad de otra.

“*Explaining away*” está relacionado con una situación con modelos que computen para explicar los mismos datos.

NEURAL IMPLICATIONS

Estos modelos pueden tratarse como modelos puramente funcionales de percepción. Serán más plausibles si se les enlaza con mecanismos neuronales y se usan para hacer predicciones experimentales.

Una clase de modelos Bayesianos se puede implementar mediante redes paralelas con interacciones locales. En estos modelos el conocimiento previo y la función de verosimilitud se implementan mediante los pesos sinápticos. Estos modelos son, a grandes rasgos, consistentes con mediciones de electrodos.

Hay mecanismos neuronales propuestos para representar la incertidumbre, lo que daría un mecanismo para la combinación ponderada de claves. El candidato más plausible es el *population encoding*.

La naturaleza gráfica de los modelos Bayesianos es similar a las conexiones feedforward y feedback que conectan las áreas visuales de los primates.

CONCLUSIONS

La aproximación Bayesiana permite diseñar modelos de observadores ideales para experimentos psicológicos y diseñar sistemas de visión de computadoras prácticos. Al aprovechar las regularidades estadísticas en las imágenes, esta aproximación ofrece una forma de lidiar con la ambigüedad y complejidad del mundo natural. La representación gráfica de las dependencias entre las variables de las tareas sugiere modos en que estos modelos pueden implementarse en términos de mecanismos neurales.