

UNIVERSIDAD DE LA LAGUNA

Trabajo de Fin de Grado de Psicología

**APLICACIÓN DE MODELOS COMPUTACIONALES
AL RAZONAMIENTO HIPOTÉTICO EN COGNICIÓN
AFECTIVA**

Departamento de Psicología Cognitiva, Social y Organizacional

Facultad de Psicología y Logopedia

Curso académico: 2021-2022

Autora: Ainhoa de la Rosa García
Tutor: Carlos Santamaría Moreno

ÍNDICE

Resumen/Abstract.....	3
Introducción.....	4
Método.....	9
Resultados.....	10
Conclusiones.....	12
Bibliografía.....	14

RESUMEN

El objetivo del presente estudio es contemplar la posibilidad de la aplicación de modelos computacionales a inferencias cognitivas afectivas relativas al razonamiento hipotético. Los seres humanos realizamos inferencias prácticamente de manera automática sobre los estados mentales propios y de los demás. En este trabajo planteo la posibilidad de predecir comportamientos posibles mediante un modelo probabilístico que captura conocimiento abstracto y conceptual mediante programas estocásticos. Estos comportamientos posibles se deben de predecir teniendo en cuenta emociones hipotéticas. Es decir, se trata de un trabajo que implica inferencias afectivas que no han ocurrido, ya que son imaginadas.

Palabras clave: emoción, modelos computacionales, teoría de la mente, razonamiento hipotético

ABSTRACT

The aim of the present study is to consider the possibility of applying computational models to affective cognitive inferences concerning hypothetical reasoning. Humans make inferences about their own and others' mental states almost automatically. In this paper I propose the possibility of predicting possible behaviours by means of a probabilistic model that captures abstract and conceptual knowledge by means of stochastic programs. These possible behaviours must be predicted taking into account hypothetical emotions. In other words, it is a task that involves affective inferences that have not occurred, as they are imagined.

Key words: emotion, computational models, theory of mind, hypothetical thinking

INTRODUCCIÓN

Los seres humanos tenemos un conjunto de ideas sobre la realidad fenomenológica tanto propia como de las personas que nos rodean. Estas ideas constituyen observaciones de nuestro propio comportamiento y de nuestros estados mentales, y cómo éstos influyen en el entorno, observando no sólo las causas de los mismos sino también las consecuencias (observables o no observables).

Esto ocurre porque no tenemos un acceso privilegiado a nuestros estados mentales. Razonamos sobre nuestra experiencia de manera contextualizada mediante teorías intuitivas, de la misma manera que razonamos sobre los estados mentales de los demás. Por ejemplo, en un contexto en el que uno se presenta a un examen sin estudiar, razonamos sobre nuestros estados mentales de manera que conscientemente inferimos que tenemos el deseo de aprobar el examen, sin embargo tenemos la creencia de que no lo vamos a conseguir. Esta inferencia sobre creencias y deseos constituye una teoría intuitiva sobre los estados mentales propios.

El hecho de inferir estados mentales en los demás existe como constructo teórico y lleva el nombre de teoría de la mente. Por ejemplo, en un contexto en el que la pareja emocional del sujeto se muestra poco cariñosa, el sujeto realiza una inferencia sobre sus estados mentales, concluyendo que pueda estar enfadada/o.

La teoría de la mente permite realizar inferencias sobre las motivaciones de otros y así explicar su comportamiento. Consiste en una ontología estructurada de conceptos como son la personalidad, las metas o el comportamiento; así como también las relaciones causales de esos conceptos. Por ejemplo, el sujeto mencionado anteriormente, sabe que su pareja es una persona cariñosa, y explica este comportamiento a partir de regularidades de sus conductas observables. Es decir, lo normal es que se comporte de forma cariñosa.

Siguiendo con el ejemplo anterior, vamos a imaginar que nosotros sabemos que Sally no ha estudiado para un examen, por tanto inferimos que Sally cree que no aprobará el examen y que probablemente tenga el deseo de aprobarlo. Estas inferencias las realizamos por el hecho de que nosotros somos conscientes de que Sally tiene una mente igual que la nuestra, y que por tanto funcionará de forma parecida.

Los recursos que utilizamos en la teoría de la mente son, por un lado lo observable del

ambiente, y por otro lado el comportamiento del sujeto del que inferimos estados mentales. Estas teorías nos ayudan a dar sentido a los demás.

Ong, (2017), pretende formular una teoría similar aplicada a las emociones, esto es, realizar inferencias de emociones en tercera persona, a grandes rasgos. La cognición afectiva comprende procesos de razonamiento cognitivo de dominio general aplicado al conocimiento de dominio específico sobre emociones (Ong, 2017). Este conocimiento abarca las teorías intuitivas de los observadores. Éstos, poseen una teoría intuitiva de emociones que incluye un conocimiento conceptual de los estados emocionales y cómo estos se relacionan con sus causas y efectos. En función de esta teoría que nos formulamos, decidimos cómo responder en contextos sociales (Ong, 2017).

Estas teorías incluyen las creencias que tiene el observador sobre cómo funcionan las emociones del observado. Estas creencias forman la base de cómo el observador entiende a los demás y cómo interactúa con ellos. Es decir, creencias sobre qué son las emociones, cómo se producen y cómo las personas se comportan en respuesta a ellas.

La experiencia emocional empieza con un sujeto experimentando un estímulo, recoge a su vez información del contexto así como señales viscerales de valencia y activación cortical (arousal). El sujeto integra estas informaciones en un proceso de conceptualización donde se etiqueta la emoción usando conceptos relativos a ésta: una teoría intuitiva de la emoción (Barrett, 2006; Lindquist & Barrett, 2008).

El análisis a nivel cognitivo que se realiza en los contextos en los que realizamos inferencias sobre estados mentales ajenos se denomina como “appraisal”. Este análisis consiste en una extracción de datos de un evento o situación, en el que como resultado se elicitaba una emoción (o una respuesta emocional).

La teoría del appraisal refleja las conexiones que existen entre emociones y actividad cognitiva. La evaluación de los eventos dados juega un papel crucial en identificar las emociones que han sido resultado de esos eventos (Wu, Baker, Tenenbaum, & Schulz, 2018).

Los deseos y creencias de un sujeto afectan a la evaluación del evento, y por tanto, influye en la reacción emocional que se genera. La teoría del appraisal también interviene en cómo se generan las emociones (Wu, Baker, Tenenbaum, & Schulz, 2018). Se infieren estados mentales desde respuestas emocionales observables implicando también la evaluación (appraisal) que se hace de los eventos.

Por otro lado, la teoría intuitiva refleja cómo los individuos piensan sobre las causas de sus propios estados emocionales. En cuanto a la teoría de emociones, se infieren los estados mentales no observables de los otros a partir de su respuesta emocional.

Se asume que la respuesta emocional está influenciada por las creencias y deseos que se tengan sobre los eventos:

- *se cree que el evento ocurrirá (creencia)*
- *se quiere/ no se quiere que el evento ocurra (deseos)*
- *se intervino/ no se intervino para provocar el evento (comportamiento intencional)*

Las atribuciones de deseo, creencia y responsabilidad causal explican una proporción grande de la varianza de las reacciones emocionales a los eventos.

Las creencias y deseos sobre un evento:

- *Son causas probables de su reacción emocional (si el sujeto es observador y no responsable del evento)*
- *Son causas probables de sus acciones (si el sujeto es responsable del evento)*

Las creencias y los deseos dan lugar a acciones, y se espera que el sujeto actúe persiguiendo sus deseos dadas sus creencias.

Ong, D. C., Zaki, J., & Goodman, N. D. (2019) generaron una taxonomía de inferencias cognitivas afectivas que incluye un amplio rango de razonamiento sobre estados mentales utilizando tanto comportamiento observable como contextos, (también una mezcla de ambos) para la inferencia de estados mentales.

Se incluyen dos tipos de relaciones causales:

- *Conexión emoción-causa (de la misma)*
- *Efecto de la emoción*

El observador, mediante el appraisal en tercera persona, razona sobre:

- *El resultado del evento que el observado experimenta.*
- *Los estados mentales del observado (sus creencias y deseos).*

Juntos resultan en que el observado experimente una emoción. Esta emoción provoca una reacción emocional y un comportamiento intencional para perseguir nuevos estados emocionales, metas y emociones.

¿Cómo entendemos el hecho de que eventos y estados mentales (del observado) den lugar a emociones?

El appraisal en tercera persona que realiza el observador depende de las inferencias que él mismo hace sobre las creencias y los deseos del observado, lo cual puede diferir de los estados mentales reales del observado.

La taxonomía generada sería la siguiente:

1. *Reconocimiento de emociones*: $P(e/x)$; $P(e/a)$. Inferencia de emociones dada una reacción emocional o una acción. (Siendo e = emoción; x = expresión emocional; a = acción/ comportamiento observable)
2. *Appraisal en 3ª persona*: $P(e/o)$; $P(e/o,d,b)$; $P(e/b,d)$. Inferencia de emociones dado/s un evento y/o unas creencias y deseos. (Siendo o =resultado; b = creencias; d = deseos).
3. *Inferencia de causas de emociones*: $P(o/e)$; $P(o/e,b,d)$. Razonamiento sobre los eventos que puedan haber causado una emoción
4. *Integración de causas*: $P(e/o,x)$; $P(e/o,a)$; $P(e/a,x)$; $P(e/o,b,d,a,x)$. Dadas múltiples causas, razonar sobre cómo éstas afectan a la emoción
5. *Appraisal Inverso*: $P(d,b/e,o)$; $P(b,d/e)$. Inferencia de deseos y creencias dada una emoción
6. *Predicciones (razonamiento hipotético)*: $P(a/e)$; $P(x/e)$. Inferencia de un posible comportamiento dada una emoción hipotética
7. *Razonamiento contrafactual*: $P(e/\text{not } o)$; $P(a/\text{not } e)$. Dado un hecho y unas emociones, razonar sobre emociones hipotéticas dado un hecho contrafactual

El fin último de este estudio sería contemplar la posibilidad de que podamos aplicar un modelo probabilístico y computacional a estas inferencias afectivas que se han generado, y en concreto, al razonamiento hipotético (apartado de predicciones, $P(a/e)$; $P(x/e)$). El razonamiento hipotético implica la imaginación de posibilidades y la exploración de sus consecuencias mediante un proceso de simulación mental (Evans,

2007). En este caso se realizaría la inferencia de un posible comportamiento dada una emoción hipotética.

Por otro lado, los modelos computacionales y probabilísticos ofrecen una visión detallada de cómo debe ser resuelta una determinada tarea, dados unos supuestos establecidos sobre la estructura de la misma.

Proporcionan predicciones cuantitativas graduadas que pueden ser probadas con datos empíricos. Un paradigma utilizado en este modelo es el análisis de nivel computacional (Marr, 1982). En este nivel la tarea se realiza con un alto grado de abstracción: ¿cuál es el objetivo de estas operaciones? ¿Qué información requiere el observador para completar las operaciones?

Si la tarea trata de explicar el comportamiento del observado, se le aplican los conceptos de “creencia-deseo-acción” y se asume que el observador tiene una teoría intuitiva del comportamiento, la tarea del observador se reduce a inferir los deseos y creencias del observado desde sus acciones. Si luego aplicamos un lenguaje de programación probabilístico, la tarea se reduce a calcular la probabilidad de sus creencias y deseos dadas sus acciones. Los modelos probabilísticos codifican la incertidumbre en los juicios humanos, y han demostrado éxito en el modelado de la cognición humana y en particular, en la cognición social. Los modelos de cognición social a menudo suponen que tanto el observado como el observador actúan racionalmente, cosa que viene de la aplicación de derivaciones matemáticas estrictas.

Los modelos de observadores ideales describen el razonamiento sin ser conscientes del proceso por el cual se extraen conclusiones y proporcionan hipótesis cuantitativas precisas a través de las cuales se explora la cognición humana.

Con todo lo mencionado anteriormente, el objetivo final de este trabajo sería contemplar la posibilidad de aplicar modelos computacionales con el propósito de predecir comportamientos posibles dadas unas emociones hipotéticas

MÉTODO

Diseño

Se realizó una revisión bibliográfica donde se buscaron diferentes artículos científicos dedicados al estudio, investigación o revisión de modelos computacionales aplicados a la cognición afectiva y/o teoría de la mente.

Estrategia de búsqueda

La búsqueda de la información se llevó a cabo a través de Google Académico y PuntoQ-ULL (portal de búsqueda de información de la Universidad de La Laguna). Se hizo tanto en español como en inglés, y sin limitar el año de publicación, con la finalidad de obtener la mayor cantidad de literatura científica posible acerca de este ámbito.

Para ello, se emplearon diferentes combinaciones de búsqueda: “computational models of emotion” y “theory of mind”, “computational models of emotion” y “hypothetical thinking”, “Appraisal theory” y “computational models of emotion”, “computational models of emotion” y “emotion inference”.

Una vez seleccionados los artículos de interés, se analizaron los abstract y las referencias bibliográficas de estos, con el objetivo de determinar la literatura relevante para esta revisión. Nuevamente, los artículos de interés se localizaron a través de Google Académico y PuntoQ-ULL.

Finalmente, tras un análisis, se seleccionaron 4 artículos científicos (Tabla 1).

Tabla 1
Artículos seleccionados para la revisión bibliográfica

Autores	Artículos
D.C Ong (2017)	<i>Computational Affective Cognition: Modeling Reasoning About Emotion</i> (Doctoral dissertation, Stanford University)
Ong, D. C., Soh, H., Zaki, J., & Goodman, N. D. (2019)	Applying probabilistic programming to affective computing. <i>IEEE Transactions on Affective Computing</i> , 12(2), 306-317.

Ong, D. C., Zaki, J., & Goodman, N. D. (2019)	Computational models of emotion inference in theory of mind: A review and roadmap. <i>Topics in cognitive science</i> , 11(2), 338-357.
Wu, Y., Baker, C. L., Tenenbaum, J. B., & Schulz, L. E. (2018)	Rational inference of beliefs and desires from emotional expressions. <i>Cognitive science</i> , 42(3), 850-884.

Análisis de datos

La información se estructuró en función de la presencia de información acerca de modelos computacionales aplicados a cognición afectiva. Se analizó la información de manera que los criterios de selección se basaban en el hecho de que el constructo a modelar eran inferencias afectivas en tercera persona.

RESULTADOS

Las teorías intuitivas se pueden formalizar utilizando modelos bayesianos en modelos de informadores ideales. Un análisis de observador ideal describe las conclusiones óptimas que un observador sacaría dada la evidencia observada y las suposiciones del observador sobre el mundo (su conocimiento previo). Este enfoque se ha utilizado con éxito para modelar el desempeño humano en una amplia gama de fenómenos: la visión, la memoria, la toma de decisiones, y más recientemente en la cognición social (Ong, 2017). Los modelos de observadores ideales describen el razonamiento sin hacer afirmaciones sobre el mecanismo o proceso por el cual los observadores humanos extraen estas conclusiones, es decir, realizan análisis a "nivel computacional", (Marr, 1982), y proporcionan hipótesis cuantitativas precisas a través de las cuales explorar la cognición.

Programas probabilísticos

Ong, D. C., Soh, H., Zaki, J., & Goodman, N. D. (2019), proponen un enfoque probabilístico por el que, mediante programación informática podamos generar modelos

relacionados con el procesamiento de emociones. Este enfoque recibe el nombre de computación afectiva. Para esto, se revisan enfoques probabilísticos que integran el razonamiento sobre emociones con el razonamiento sobre otros estados mentales.

Los programas probabilísticos permiten capturar conocimiento abstracto y conceptual mediante programas estocásticos. Los programas probabilísticos producen muestras de una distribución de posibles resultados. Esto permite modelar la incertidumbre, tanto si la incertidumbre viene sobre un conocimiento incompleto sobre el mundo y los estados mentales de los demás, como si viene de una teoría incompleta, o la aleatoriedad inherente de un proceso generativo.

Otro aspecto crucial es la modularidad. En un paradigma de programación probabilística, los programas probabilísticos modulares están dispuestos de manera secuencial o jerárquica para producir fenómenos más complejos, o simplemente para captar el razonamiento sobre estos fenómenos. Este aspecto es importante en el modelado de las emociones y de los procesos que las causan así como de sus consecuencias. Esto se debe a que el modelado de emociones requiere la modelación de una amplia gama de procesos de elicitación de emociones (a través del appraisal) y de los efectos conductuales de la emoción. La modularidad es esencial para definir estos procesos.

Aplicabilidad de programas probabilísticos a inferencias afectivas

En el estudio de Wu et al., (2017), la habilidad de distinguir estados mentales basándonos en expresiones emocionales varió en los 4 experimentos, sin embargo, las inferencias sí se ajustaron al modelo (todas las correlaciones tenían un nivel de $r=.95$ correspondiendo al menos el 90% de la varianza explicada). El modelo tuvo la capacidad de representar las inferencias que se hicieron sobre reacciones emocionales que fueron resultado de un hecho observado (experimento 1 y 2) o de un hecho del que el sujeto es responsable (experimento 3 y 4). Así mismo, también representó adecuadamente las dos reacciones emocionales que se propusieron (ante lo que se espera y ante lo que ocurre). El trabajo mencionado sugiere que la estructura causal de estas relaciones es suficiente para producir modelos cuantitativos inversos de inferencia de estados mentales de manera precisa. Parece que nuestra teoría intuitiva de emociones está estructuralmente y causalmente entrelazada con nuestra teoría de la mente, de una manera que permite la predicción hacia adelante de las creencias y deseos de un sujeto desde sus expresiones emocionales, y hacia atrás, de inferencia de expresiones

emocionales hacia creencias y deseos, con un grado de coherencia interna que sugiere mecanismos de inferencia probabilística altamente optimizados (Wu et al., 2017).

Modelado del razonamiento hipotético en programas probabilísticos

El razonamiento hipotético se puede modelar simplemente permitiendo que las variables inferidas sean hipotéticas, por ejemplo, estimar el comportamiento dada una emoción hipotética $P(a|e)$. Dentro del modelado bayesiano, esto se puede operativizar como muestreo de "predicciones posteriores" bajo el modelo (es decir, predicciones condicionales de nuevos datos), y se ha aplicado para modelar predicciones en otros dominios, pero aún no en cognición afectiva. El trabajo futuro debe centrarse en la aplicación de tales técnicas para modelar predicciones de la gente común sobre emociones (hipotéticas) (Ong, D. C., Zaki, J., & Goodman, N. D., 2019).

CONCLUSIONES

El presente estudio se centra en recopilar información acerca de la inferencia de emociones en terceras personas, modelos computacionales y su aplicabilidad en campos cognitivos, más específicamente afectivos en este caso.

La literatura revisada apunta a que el modelado de la cognición afectiva es todavía un mundo a explorar entre los investigadores del campo, y resulta ser un tema susceptible de trabajar en las líneas en las que se ha confeccionado el trabajo.

La evidencia recopilada ha tenido una serie de repercusiones, entre las que se encuentran, por ejemplo, la aplicación de teorías matemáticas a la comunicación cooperativa. Las simulaciones computacionales respaldan y elaboran los resultados teóricos, así como demuestran su ajuste al comportamiento humano. Los resultados de este estudio, muestran que la comunicación cooperativa probablemente permite una transmisión de creencias efectivas que se requieren para explicar el aprendizaje humano y mejorar la interacción hombre-máquina (Wang, P., Wang, J., Paranamana, P., & Shafto, P., 2020). Estos resultados indican el futuro desarrollo de los modelos computacionales aplicados al comportamiento humano.

Por otro lado, también ha tenido repercusión en temas de sistemas computacionales de aprendizaje social. El aprendizaje del valor de los estímulos y acciones de otros (aprendizaje social) contribuye adaptativamente a la supervivencia y juega un papel clave para la evolución cultural. El aprendizaje social sigue los mismos principios que el

aprendizaje experiencial basado en valores, incluyendo los modelos computacionales. Olsson, A., Knapska, E., & Lindström, B., (2020) integran perspectivas neuronales y computacionales del aprendizaje social, comprendiendo el comportamiento complejo y no complejo.

Resulta interesante mencionar que la literatura propuesta, ha servido también para el desarrollo de trabajos sobre interpretación y modelado de emociones en la administración de robots autónomos, usando paradigmas de control basado en una variable de programación (Kowalczyk, Z., Czubenko, M., & Merta, T., 2020) y en trabajos sobre contribución de claves visuales y lingüísticas a la sincronización fisiológica y la empatía exacta (Jospe, K., Genzer, S., Klein Selle, N., Ong, D., Zaki, J., & Perry, A., 2020).

Con todo lo mencionado, se puede concluir que la aplicación de modelos computacionales a inferencias afectivas es todavía un campo en el que existe un gran desconocimiento, y sobre el cual se podría profundizar a nivel investigación, sobretodo teniendo en cuenta la amplia repercusión que tienen los trabajos destinados a este fin.

BIBLIOGRAFÍA

Barrett, L. F. (2006). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review*, 10 (1), 20–46.

Evans, J.S.B.T. (2007). *Hypothetical Thinking: Dual Processes in Reasoning and Judgement* (1st ed.). Psychology Press. <https://doi.org/10.4324/9780203947487>

Jospe, K., Genzer, S., Klein Selle, N., Ong, D., Zaki, J., & Perry, A. (2020). The contribution of linguistic and visual cues to physiological synchrony and empathic accuracy. *Cortex*, 132, 296-308.

Kowalczyk, Z., Czubenko, M., & Merta, T. (2020). Interpretation and modeling of emotions in the management of autonomous robots using a control paradigm based on a scheduling variable. *Engineering Applications of Artificial Intelligence*, 91, 103562.

Lindquist, K. A., & Barrett, L. F. (2008). Constructing emotion the experience of fear as a conceptual act. *Psychological Science*, 19 (9), 898–903.

Olsson, A., Knapska, E., & Lindström, B. (2020). The neural and computational systems of social learning. *Nature Reviews Neuroscience*, 21(4), 197-212.

- Ong, D. C. (2017). *Computational Affective Cognition: Modeling Reasoning About Emotion* (Doctoral dissertation, Stanford University).
- Ong, D. C., Soh, H., Zaki, J., & Goodman, N. D. (2019). Applying probabilistic programming to affective computing. *IEEE Transactions on Affective Computing*, *12*(2), 306-317.
- Ong, D. C., Zaki, J., & Goodman, N. D. (2019). Computational models of emotion inference in theory of mind: A review and roadmap. *Topics in cognitive science*, *11*(2), 338-357.
- Wang, P., Wang, J., Paranamana, P., & Shafto, P. (2020). A mathematical theory of cooperative communication. *Advances in Neural Information Processing Systems*, *33*, 17582-17593.
- Wu, Y., Baker, C. L., Tenenbaum, J. B., & Schulz, L. E. (2018). Rational inference of beliefs and desires from emotional expressions. *Cognitive science*, *42*(3), 850-884.