Research Problem

Mauricio Gonzalez Soto

8 de febrero de 2018

Índice

1. Introducción

Mientras jugamos un videojuego, o aprendemos a conducir, estamos obligados a tomar decisiones y a intervenir de manera activa en el mundo. Este proceso activo de toma de decisiones modifica el estado del mundo, lo cual nos permite aprender cómo nuestras acciones cambiaron el mundo y qué fue causado por estas. De esta manera, a través de la interacción con el entorno, podemos descubrir relaciones causales que luego serán utilizadas en posteriores decisiones.

Las intuiciones sobre aprendizaje causal por interacción provienen de estudios realizados en Psicología Computacional y Ciencias Cognitivas, de aquí se sabe que los seres humanos obtienen conocimiento causal al encontrarse en situaciones en las cuales deben tomar decisiones de manera secuencial para lograr un objetivo futuro. Se sabe también que el conocimiento causal adquirido puede ser utilizado para mejorar las futuras decisiones tomadas.

Notando que los modelos causales tradicionales requieren de intervenciones y experimentación en el mundo, y que estas intervenciones son una forma de interacción y alteración del mundo, salta a la vista la relación entre Aprendizaje por Refuerzo y aprendizaje causal. Entonces, resulta natural preguntar cómo pueden aprenderse relaciones causales de manera interactiva de modo que el conocimiento causal adquirido permita a un agente tomar mejores decisiones.

Lake et al. (2017) argumentan que los sistemas inteligentes actuales, a pesar de sus impresionantes logros, aun están lejos de alcanzar una verdadera inteligencia; uno de los requisitos para esto es que puedan construir

modelos causales que sean interpretables, y no sólo reconocer patrones estadísticos.

La pregunta se convierte entonces en cómo obtener e incorporar conocimiento causal en un problema de aprendizaje por refuerzo. Además, uno se pregunta si es posible utilizar este conocimiento causal obtenido al interactuar sobre el mundo para realizar otro tipo de inferencias causales en ciertos dominios de interés.

2. Círculos del State of The Art

2.1. Outer Circle

- Holland (1986) hace un recuento de varias definiciones de causalidad a lo largo de la historia, desde Aristóteles, pasando por David Hume, a Patrick Suppes, Granger.
- Sutton and Barto (1998) es el libro clásico en el cual se define como aprendizaje por refuerzo a cualquier técnica orientada a resolver problemas mediante la interacción con un entorno y utilizando *feedback*.
- Joyce (1999) estudia la toma de decisiones bajo incertidumbre utilizando información causal; es decir; sé que A *causa* B, entonces puedo predecir qué pasará si A es escogido, u observado.
- Spirtes et al. (2000) junto con Pearl (2009) estudian los modelos gráficos causales y exponen la definición de causalidad aquí utilizada.
- El libro de Koller and Friedman (2009) tiene un capítulo sobre modelos gráficos causales, y explica que estos consisten en dotar de una semántica causal amodelos gráficos probabilistas.
- Libro de Teoría de la Decisión bajo incertidumbre Gilboa (2009).
- Pearl (1988), Libro clásico sobre uso de grafos para representar distribuciones.
- van Otterlo and Wiering (2012) presenta el e stado del arte de aprendizaje por refuerzo hasta esa fecha
- Danks (2014) explica cómo es posible representar procesos mentales como operaciones sobre gráficos probabilistas; en particular, el problema de inducción causal.
- Gershman (2015). Aprendizaje por Refuerzo en el Cerebro.
- Lopez-Paz et al. (2015). Hacia una teoría de aprendizaje para relaciones causa-efecto.
- Lake et al. (2017). Condiciones que, idealmente, deberían cumplir algoritmos para que se pueda decir que piensan *como* los humanos,

2.2. Middle Circle

- Dawid (2002) estudió problemas de decisión y modelos causales, y mostró cómo las redes bayesianas pueden ser *aumentadas* con nodos de decisión que permitan utilizar información causal.
- Hagmayer and Sloman (2009) muestra que los humanos entendemos nuestras acciones en el mundo como *intervenciones* sobre este.
- On-line learning in causal models(Wellen and Danks (2012))): Proponen un modelo de aprendizaje causal *on-line* en el cual se añaden o borran arcos en un modelo causal según un ciclo de predicción-error.
- Lopez-Paz et al. (2015)
- Krynski and Tenenbaum (2007) Human reasoning under uncertainty naturally operates over causa mental models and statistical data supports correct Bayesian inference only when they can be incorporated into a causal model.

- Goudet et al. (2017) aprendizaje de modelos gráficos a través de redes neuronales generativas.
- Hernandez-Leal et al. (2017) estudia la no-estacionareidad de un entorno.
- Li (2017). Overview general de Deep Reinforcement Learning

2.3. Inner Circle

- Eberhardt (2008) plantean el problema de *descubrimiento causal* como un juego entre un científico vs la naturaleza en el cual la naturaleza intenta mantener ocultos sus secretos y el científico intenta descubrirlos mientras minimiza un costo.
- Hagmayer and Meder (2013) proponen explorar el uso de información causal en problemas de toma de decisiones secuenciales en seres humanos.
- Ortega and Braun (2014) exploran el muestreo de Thompson para el descubrimiento de estructura causal en problemas de decisión para los cuales la política está especificada.
- Bramley et al. (2015) presenta heurísticas para el aprendizaje causal en el caso de decisiones secuenciales.
- Alon et al. (2015) Estudian problemas de aprendizaje on-line en los cuales en cada ronda un jugador escoje una acción y recibe un *loss* previamente especificado. Estudian las *feedback graphs* que básicamente dicen si al ejecutar acción i vemos el loss asociado a acción j y estudian cómo la estructura de este grafo afecta el aprendizaje. Es principalmente un paper de estructura de grafos y el *aprendizaje* se define en términos de Teoría de Juegos.
- Lattimore et al. (2016) buscan descubrir la mejor intervención, dado un modelo causal, para maximizar un retorno.
- Garnelo et al. (2016) proponen un algoritmo de aprendizaje por refuerzo que construye una representación simbólica del ambiente, la cual es usada para escoger la siguiente acción.
- Innes et al. (2018) proponen un algoritmo para que un tomador de decisiones infiera posibilidades no vistas durante el entrenamiento así como las relaciones causales ente datos.

3. Motivación

Tomando el cuenta los trabajos mencionados en la sección anterior, surge la pregunta de si es posible diseñar un agente que aprenda a realizar una tarea por medio de interacción con su entorno y mediante el uso de información causal recolectada a través de esa interacción. Este agente buscará aprender un modelo causal de su entorno al mismo tiempo que aprende una *política*. Con cada acción que toma, el agente aprenderá un poco sobre la *estructura causal* de su entorno y utilizará este conocimiento para la futura toma de decisiones.

Cada decisión que el agente toma (acción que lleva a cabo) altera el estado del mundo. La información obtenida a partir de cada acción, será utilizada para actualizar el conocimiento causal que se tenga hasta el momento, esto con el fin mejorar las decisiones que toma y para aprender otros aspectos estructurales (causales) sobre su ambiente. La intuición detrás de esto es que al aprender mediante una toma de decisiones repetida, decisiones que intervienen directamente sobre el entorno, se aprenden también relaciones causales; por ejemplo, al aprender a conducir se aprende implícitamente que ciertos movimientos del volante *causan* ciertos cambios en el estado del automovil (Danks (2014)). Otro ejemplo es en el contexto de videojuegos, en el cual al tomar ciertas decisiones se conoce el impacto que estas provocan en el estado del juego, conocimiento que a su vez es utilizado para tomar mejores decisiones en un futuro y que maximicen las recompensas u objetivos a largo plazo. No sólo hay intuición detrás de esto, pues se sabe que los seres humanos utilizan el conocimiento causal que la experiencia les provee para realizar predicciones para intervenciones en el mundo no realizadas aun (Meder et al. (2008), Hagmayer and Sloman (2009))

Notemos que al hablar de modelos gráficos causales que se actualizan y adaptan a la luz de nueva evidencia estamos tocando terreno Bayesiano, pues tenemos creencias actuales sobre la estructura de un grafo, observaremos evidencia, y actualizaremos estas creencias. Existen trabajos de inferencia Bayesiana sobre datos estructurados como grafos (Acar et al. (2007)), además de grafos probabilistas (Pearl (1988), Koller and Friedman (2009), Sucar (2015)). Hacer inferencia bayesiana sobre grafos implicaría utilizar distribuciones de probabilidad sobre un espacio de grafos. Por eso, utilizar inferencia Bayesiana no-paramétrica (Phadia (2015), Müller et al. (2016), Ghosal and van der Vaart (2017)) puede ser útil pues admite modelos de alta complejidad.

El trabajo de Wellen and Danks (2012) muestra cómo un modelo gráfico causal podría ser aprendido en línea al tomar en cuenta las decisiones tomadas y las recompensas obtenidas por un tomador de decisiones. Por otro lado, los modelos de Sloman and Hagmayer (2006), Meder et al. (2008), Hagmayer and Sloman (2009) y Hagmayer and Meder (2013) muestran que estas intuiciones no son erradas y que los seres humanos incorporamos y adquirimos conocimiento causal al enfrentarnos en problemas de toma de decisión simples, aunque aun no existen muchos estudios en los cuales los escenarios de decisión no sean hipotéticos sino que se intente maximizar una recompensa real a largo plazo. De la misma manera, Hagmayer and Meder (2013) mencionan que faltan estudios en donde las personas tengan otro objetivo más allá de revelar directamente la estructura causal del entorno y en los cuales vean directamente las consecuencias de sus acciones. Por lo tanto, es interesante preguntar si un agente que busque maximizar una recompensa pueda aprender modelos causales mediante la interacción con su entorno.

Un agente que interactua con su entorno con el fin de maximizar una recompensa efectivamente puede aprender aspectos estructurales sobre el entorno, como nos muestra Garnelo et al. (2016), quien además señala que el agente logra mejorar su desempeño. Además, muestran la importancia de tener una semántica que permita hablar de la estructura del entorno. Ellos utilizaron lógica de primer orden, pues están más interesados en hacer razonamientos *de alto nivel* sobre el entorno, que involocura más que las relaciones causales, y por esto ellos se comprometen con lógica de primer orden, que al restringirnos al caso causal se queda corta pues los modelos gráficos probabilistas son más expresivos, además de modulares y montónicos.

Mientras que Garnelo et al. (2016) utiliza una representación lógica para escoger una acción, Lattimore et al. (2016) explora la pregunta de cómo escoger *la mejor intervención* dado un modelo causal a través de una serie de intervenciones y observaciones secuenciales; por ejemplo, en el caso en el que un granjero quiera maximizar sus cosechas, y sabe que ese rendimiento está afectado por el uso de fertilizante, la humedad y la exposición al sol, pero sólo tiene los recursos para llevar a cabo sólo una de estas modificaciones en cada temporada. El algoritmo de Lattimore le dirá, después de T observaciones, cuál es la mejor intervención a llevar a cabo con el fin de maximizar esa cantidad. Este podría ser un primer paso hacia responder la pregunta *¿dado un modelo causal, qué acción tomar?*

Por otro lado, en el caso inverso Ortega and Braun (2014) exploran un esquema de muestreo llamado Muestreo de Thompson, en el cual un agente muestrea una de entre varias acciones de acuerdo a la probabilidad subjetiva que tenga el agente de que esta sea la mejor acción. Los autores conjeturan que esto pueden utilizarse para hacer inducción causal sobre una serie de acciones establecidad (política conocida).

4. Problema de investigación

4.1. Enunciado

Aprendizaje de modelos causales a través de la interaccion con el entorno y el uso de este conocimiento para la toma de mejores decisiones en problemas de decisiones secuenciales.

4.2. Relevancia

Los métodos actuales de Aprendizaje por Refuerzo (RL) son puramente reactivos (Garnelo et al. (2016)), además de necesitar de mucho tiempo de entrenamiento, pues los métodos usuales requieren de observar muchas parejas acción-estado (Sutton and Barto (1998)). Las representaciones del entorno normalmente utilizadas, en términos de Procesos de Decisión Markovianos no permiten hacer razonamiento de alto nivel sobre este entorno (Innes et al. (2018), Garnelo et al. (2016)), por lo que aprender estructuras causales permitiran a un agente tomar mejores decisiones pues podrá considerar consecuencias de sus acciones en vez de sólo estar reaccionando a recompensas.

4.3. Factibilidad

Del trabajo desarrollado en el área de RL sabemos que es posible que un agente autónomo aprenda una enorme variedad de tareas mediante interacción; ejemplos de esto los tenemos en los trabajos de Mnih et al. (2013), Mnih et al. (2015), Silver et al. (2016), Silver et al. (2017), Espeholt et al. (2018).

Por otro lado, sabemos que los seres humanos adquieren conocimiento causal de su entorno en el caso de problemas de decisión secuenciales (Hagmayer and Sloman (2009), Hagmayer and Meder (2013)).

Además, gracias al trabajo de Lattimore et al. (2016) se sabe que es posible escoger acciones dado un modelo causal, mientras que el trabajo de Wellen and Danks (2012) muestra que un modelo causal puede ser aprendido *on-line*; es decir, mediante interacción con el ambiente.

Por lo tanto, los diversos elementos existen (en sus ideas de fondo) y pueden ser unidos, aunque no será una tarea trivial.

5. Grupos principales

5.1. Causalidad

- Carnegie Mellon (Spirtes)
- UC Los Angeles (Pearl)
- MIT (Josh Tenenbaum)
- Max Planck Institute for Intelligent Systems (Schölkopf)
- University of Copenhagen (Jonas Peters)
- Edinburgh (Albrecht, Subramanian Ramamoorthy)

5.2. Reinforcement Learning

- Google DeepMind (David Silver, Alex Graves, V. Mnih, Murray Shanahan, Marta Garnelo)
- Oxford (Nando De Freitas, Shimon Whiteson)
- Cambridge (Ghahramani)
- Alberta University (Richard Sutton, Csaba Szepesvári)
- UC Berkeley (Pieter Abbeel)
- Carnegie Mellon (Salakhutdinov)
- Imperial College (Murray Shanahan, Marta Garnelo)

- University College London Gatsby Unit (Peter Dayan, Watkins, Dawid)
- Edinburgh (Craig Innes)

5.3. Ciencias Cognitivas

- Carnegie Mellon (David Danks)
- MIT (Josh Tenenbaum)
- UC Berkeley (Tom Griffiths)
- Edinburgh University (Andy Clark)
- Göttingen Institut für Psychologie (York Hagmayer)

Referencias

Acar, U. A., Ihler, A. T., Mettu, R. R., and Sümer, Ö. (2007). Adaptive bayesian inference. In *Proceedings of the 20th International Conference on Neural Information Processing Systems*, pages 1441–1448. Curran Associates Inc.

Alon, N., Cesa-Bianchi, N., Dekel, O., and Koren, T. (2015). Online learning with feedback graphs: Beyond bandits. *arXiv preprint arXiv:1502.07617*.

Bramley, N., Dayan, P., and Lagnado, D. A. (2015). Staying afloat on neurath's boat-heuristics for sequential causal learning. In *CogSci*, volume 37. Cognitive Science Society.

Danks, D. (2014). *Unifying the mind: Cognitive representations as graphical models*. Mit Press.

Dawid, A. P. (2002). Influence diagrams for causal modelling and inference. *International Statistical Review*, 70(2):161–189.

Eberhardt, F. (2008). Causal discovery as a game. In *Proceedings of the 2008th International Conference on Causality: Objectives and Assessment-Volume 6*, pages 87–96. JMLR. org.

Espeholt, L., Soyer, H., Munos, R., Simonyan, K., and Volodimir, M. (2018). Impala: Scalable distributed deep-rl with importance weighted actor-learner architectures. *arXiv preprint arXiv:1801.03331v1*.

Garnelo, M., Arulkumaran, K., and Shanahan, M. (2016). Towards deep symbolic reinforcement learning. *arXiv preprint arXiv:1609.05518*.

Gershman, S. J. (2015). Reinforcement learning and causal models. In *The Oxford Handbook of Causal Reasoning*.

Ghosal, S. and van der Vaart, A. (2017). *Fundamentals of nonparametric Bayesian inference*, volume 44. Cambridge University Press.

Gilboa, I. (2009). *Theory of Decision under Uncertainty*. Cambridge University Press.

Goudet, O., Kalainathan, D., Caillou, P., Lopez-Paz, D., Guyon, I., Sebag, M., Tritas, A., and Tubaro, P. (2017). Learning functional causal models with generative neural networks. *arXiv preprint arXiv:1709.05321*.

Hagmayer, Y. and Meder, B. (2013). Repeated causal decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(1):33.

Hagmayer, Y. and Sloman, S. A. (2009). Decision makers conceive of their choices as interventions. *Journal of Experimental Psychology: General*, 138(1):22.

Hernandez-Leal, P., Kaisers, M., Baarslag, T., and de Cote, E. M. (2017). A survey of learning in multiagent environments: Dealing with non-stationarity. *arXiv preprint arXiv:1707.09183*.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960.

Innes, C., Lascarides, A., Albrecht, S. V., Ramamoorthy, S., and Rosman, B. (2018). Reasoning about unforeseen possibilities during policy learning. *arXiv preprint arXiv:1801.03331*.

Joyce, J. M. (1999). *The Foundations of Causal Decision Theory*. Cambridge University Press.

Koller, D. and Friedman, N. (2009). *Probabilistic graphical models: principles and techniques*. MIT press.

Krynski, T. R. and Tenenbaum, J. B. (2007). The role of causality in judgment under uncertainty. *Journal of Experimental Psychology: General*, 136(3):430.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., and Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, 40.

Lattimore, F., Lattimore, T., and Reid, M. D. (2016). Causal bandits: Learning good interventions via causal inference. In Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I., and Garnett, R., editors, *Advances in Neural Information Processing Systems 29*, pages 1181–1189. Curran Associates, Inc.

Li, Y. (2017). Deep reinforcement learning: An overview. *arXiv preprint arXiv:1701.07274*.

Lopez-Paz, D., Muandet, K., Schölkopf, B., and Tolstikhin, I. (2015). Towards a learning theory of cause-effect inference. In *International Conference on Machine Learning*, pages 1452–1461.

Meder, B., Hagmayer, Y., and Waldmann, M. R. (2008). Inferring interventional predictions from observational learning data. *Psychonomic Bulletin & Review*, 15(1):75–80.

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., and Riedmiller, M. (2013). Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533.

Müller, P., Quintana, F. A., Jara, A., and Hanson, T. (2016). *Bayesian nonparametric data analysis*. Springer series in Statistics.

Ortega, P. A. and Braun, D. A. (2014). Generalized thompson sampling for sequential decision-making and causal inference. *Complex Adaptive Systems Modeling*, 2(1):2.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann.

Pearl, J. (2009). *Causality*. Cambridge university press.

Phadia, E. G. (2015). *Prior processes and their applications*. Springer series in Statistics.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al. (2016). Mastering the game of go with deep neural networks and tree search. *nature*, 529(7587):484–489.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., et al. (2017). Mastering the game of go without human knowledge. *Nature*, 550(7676):354.

Sloman, S. A. and Hagmayer, Y. (2006). The causal psycho-logic of choice. *Trends in Cognitive Sciences*, 10(9):407–412.

Spirtes, P., Glymour, C. N., and Scheines, R. (2000). *Causation, prediction, and search*. MIT press.

Sucar, L. E. (2015). Probabilistic graphical models. *Advances in Computer Vision and Pattern Recognition. London: Springer London. doi*, 10:978–1.

Sutton, R. S. and Barto, A. G. (1998). *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge.

van Otterlo, M. and Wiering, M. (2012). Reinforcement learning and markov decision processes. In *Reinforcement Learning*, pages 3–42. Springer.

Wellen, S. and Danks, D. (2012). Learning causal structure through local prediction-error learning. Cognitive Science Society.