

MODELACIÓN DE LA PERCEPCIÓN SUBJETIVA DEL ESPACIO PÚBLICO MEDIANTE TÉCNICAS DE MACHINE LEARNING Y MODELOS DE ELECCIÓN DISCRETA

Tomás Rossetti ^a, Hans Lobel ^{bc} y Ricardo Hurtubia ^{cd}

^a Centro de Desarrollo Urbano Sustentable e Instituto de Sistemas Complejos de Ingeniería,
terosset@uc.cl

^b Departamento de Ingeniería de Transporte y Logística, Pontificia Universidad Católica de Chile

^c Departamento de Ciencia de la Computación, Pontificia Universidad Católica de Chile

^d Escuela de Arquitectura, Pontificia Universidad Católica de Chile

Resumen

La manera en que las personas perciben el espacio construido es relevante para entender cómo lo utilizan y se mueven por él. La mayor parte de las contribuciones en el campo de modelar dichas percepciones usando datos recolectados a un nivel masivo ocupan técnicas de *machine learning*, logrando resultados significativos. Sin embargo, estos modelos resultan poco informativos debido a la alta dificultad asociada a identificar analíticamente la forma en que los distintos atributos de los espacios públicos contribuyen a las percepciones de los usuarios.

Este trabajo plantea estimar modelos que cuantifiquen percepciones mediante modelos basados en la teoría de utilidad aleatoria, usando como variables explicativas segmentaciones semánticas de imágenes de espacios públicos generadas mediante algoritmos de *machine learning*. Utilizando el set de datos *Place Pulse*, se estimaron modelos con buenas capacidades predictivas y que permiten un análisis de la relevancia de los atributos, aunque con ajustes menores a los logrados con métodos directos de *machine learning*. A partir de esto, planteamos una línea de investigación centrada en la estimación de modelos que logren cuantificar percepciones al nivel de las técnicas de *machine learning*, y que alimenten el proceso de diseño de políticas públicas y la toma de decisiones en cuanto a inversión en infraestructura y espacios públicos.

Palabras clave: percepciones; modelos de elección discreta; *machine learning*

Keywords: perceptions; discrete choice models; machine learning

1. Introducción

La forma en que las personas perciben el entorno construido es relevante para comprender la forma en que este se utiliza. Por ejemplo, si un lugar se percibe como deprimente o ruidoso, podría ser infrautilizado, incluso si existen amenidades físicas que podrían tener el potencial para atraer personas. En ese sentido, es de interés comprender cómo se perciben distintos espacios para identificar oportunidades de intervención que puedan mejorar la calidad de vida de sus usuarios mediante el diseño de infraestructura y espacios públicos que resulten más atractivos y socialmente rentables.

Durante los últimos años han habido esfuerzos significativos por cuantificar y sistematizar estas percepciones, particularmente con la recolección masiva de datos y el uso de técnicas de *machine learning*. Esto ha permitido, por ejemplo, generar mapas perceptuales para ciudades enteras, como es el caso de StreetScore (Naik et al., 2014), facilitando encontrar áreas prioritarias para intervención, rediseño o aplicación de políticas públicas. Sin embargo, a pesar de las ventajas que presentan las metodologías de *machine learning*, generalmente no permiten interpretar los parámetros obtenidos y su relación con los atributos del espacio, dificultando su utilidad para guiar el proceso de generación de políticas públicas. En otras palabras, la mayoría de estos modelos no entregan información que dé a entender cuáles son los motivos por los cuales las personas perciben un espacio de cierta forma.

Este estudio propone cuantificar percepciones del espacio público en una escala masiva utilizando una metodología novedosa. En específico, se utilizaron los indicadores perceptuales del set de datos *Place Pulse*¹, que contiene casi 1.000.000 observaciones para más de 100.000 imágenes de distintos lugares del mundo obtenidos mediante una encuesta online. Estos fueron utilizados por modelos basados en la teoría de utilidad aleatoria que tuvieron como *input* principal segmentaciones semánticas de dichas imágenes. Los modelos obtenidos logran reproducir las percepciones de las personas encuestadas de buena manera, y además entregan parámetros interpretables.

Estos resultados, que creemos son prometedores, son el inicio de un camino de investigación que tiene tres objetivos principales: en primer lugar, lograr modelos basados en la teoría de utilidad aleatoria que cuantifiquen percepciones tan bien como lo hacen los métodos de *machine learning*; en segundo lugar, comparar estos resultados con datos reales de uso de la ciudad; y en tercer lugar, cuantificar percepciones a partir de una encuesta dirigida específicamente a modos de transporte sustentables.

La Sección 2 de este documento presenta una breve revisión bibliográfica de esfuerzos anteriores por cuantificar percepciones de atributos cualitativos de espacios públicos a una escala masiva. Luego, la Sección 3 describe la fuente de los indicadores perceptuales utilizados, el análisis de imágenes que se realizó, y la metodología utilizada para estimar los modelos que se presentan. Posteriormente, la Sección 4 muestra los resultados obtenidos y presenta algunos breves análisis. Finalmente, la Sección 5 presenta algunas conclusiones y caminos a seguir.

¹<http://pulse.media.mit.edu/>

2. Revisión bibliográfica

Desde el trabajo seminal de Kevin Lynch (1960) que existe interés por comprender cómo las personas perciben el entorno que las rodea, particularmente en contextos urbanos. En efecto, este trabajo fue de los primeros que explícitamente plantearon que las anteojerías con las que las personas procesan la información de su entorno es relevante para entender el uso que se le da. Estudios posteriores le han dado la razón, encontrando, por ejemplo, que la comodidad y facilidad de acceso aumentan el uso de espacios públicos (Khisty, 1994; Shriver, 1997) o incentivan el uso de ciertos modos de transporte (Antonakos, 1995; Hyodo et al., 2000; Zacharias, 2001; Hunt y Abraham, 2007). Otros estudios han incluso concluido que barrios que se perciben como más desordenados e inseguros tienen un relación relevante con la prevalencia de depresión (Latkin y Curry, 2003), obesidad (Ellaway et al., 2005), alcoholismo (Kuipers et al., 2012) y enfermedades de transmisión sexual (Cohen et al., 2000).

Algunos estudios iniciales buscaron sistematizar y cuantificar estas percepciones, logrando aportes significativos (ver, por ejemplo, Schroeder et al., 1984; Weber et al., 2008). Sin embargo, al momento en que se realizaron estos estudios resultaba bastante costoso recolectar indicadores de percepción para un gran número de contextos y cantidad de usuarios. Sin embargo, durante los últimos años han existido varios estudios que aprovechan los avances en el campo de *machine learning* para procesar indicadores obtenidos mediante métodos novedosos de recolección masiva de datos.

Todos los estudios encontrados que caen dentro de esta clasificación consideran encuestas en línea que buscan que, ante un par de imágenes correspondientes a espacios públicos, los usuarios elijan cuál se ajusta más a cierto atributo cualitativo. Una vez obtenidos estos indicadores, todos los estudios revisados los agruparon para generar un único puntaje que indique qué tan bien se ajusta una imagen a este adjetivo. Mientras que Quercia et al. (2014), Ordonez y Berg (2014) y Porzi et al. (2015) construyen este puntaje en base a la cantidad de votos que recibió cierta imagen, Dubey et al. (2016) y Naik et al. (2014) lo hacen generando un ranking que considera qué imágenes fueron elegidas por sobre otras.

Una vez obtenidos estos puntajes, todos los estudios ocuparon métodos de *machine learning* que clasifican una imagen dentro de algún adjetivo. Esto se hizo a través de redes neuronales convolucionales (Krizhevsky et al., 2012) o regresión de máquinas de vectores (Drucker et al., 1997). Mientras que para el primer caso los datos de entrada son la imagen en sí, para el segundo se deben seleccionar los atributos a considerar. Tanto Naik et al. (2014) como Ordonez y Berg (2014) seleccionaron atributos de bajo nivel para estas regresiones (es decir, aquellos que se basan solamente en información contenida en la imagen, como su contraste, saturación o nitidez). Ambos casos eligieron atributos de bajo nivel representados por vectores de grandes dimensiones, lo que hace que no puedan ser fácilmente comparables ni interpretables.

A pesar de las ventajas que presentan estos modelos en su procesamiento de datos y habilidad predictiva, estos no entregan información que explique los motivos detrás de las elecciones de los usuarios. En efecto, ni las regresiones ni modelos de redes neuronales presentados permiten hacer inferencia de los parámetros estimados, por lo que no son capaces de indicar qué elementos tienen una incidencia significativa sobre los procesos de percepción de los usuarios ni guiar políticas

públicas. La única excepción es el caso de Porzi et al. (2015), que logran identificar qué elementos de una imagen inciden más sobre el puntaje de esta, lo que mediante un análisis *ex post* permite realizar ciertas inferencias. Sin embargo, este proceso difícilmente puede ser sistemático dado el volumen de imágenes necesarias para la estimación de este modelo.

Creemos que esta debilidad puede ser subsanada mediante el uso de modelos econométricos, con un sólido fundamento en teorías de comportamiento, que actualmente no existe en la literatura. En específico, proponemos llenar esta vacío mediante el uso de modelos fundado en la teoría de utilidad aleatoria (McFadden, 1974; Domencich y McFadden, 1975), que permite no solo predecir adecuadamente el comportamiento de los usuarios, sino que también informar qué elementos de las imágenes presentadas inciden sobre las percepciones de los usuarios.

3. Metodología y recolección de datos

En esta sección se detallará la metodología seguida para obtener los modelos mostrados en la Sección 4. En primer lugar se detallará la fuente de los indicadores de percepción. Luego, se presentará la metodología que se siguió para analizar las imágenes y obtener información de ellas que sirva como *input* para los modelos estimados. Finalmente, se describirá brevemente la base metodológica de los modelos basados en la teoría de utilidad aleatoria, para además mostrar la especificación de los modelos estimados.

3.1. Indicadores de percepción: *Place Pulse 2.0*

Para obtener información sobre percepciones en entornos urbanos, se utilizó la base de datos *Place Pulse 2.0* (Dubey et al., 2016). Esta plataforma busca sondear percepciones de personas a lo largo de todo el mundo mediante *crowdsourcing* en el sitio web <http://pulse.media.mit.edu/>. Aquí, a cada sujeto encuestado se le presentan dos imágenes tomadas al azar de 56 ciudades, y se le pide que decida cuál se ajusta mejor a algún adjetivo. Este set de datos contempla seis adjetivos, que son “aburrido”, “adinerado”, “animado”, “bonito”, “deprimente” y “seguro”.

El set de datos, de libre disposición para cualquiera que quiera utilizarlos, contempla más de 1.200.000 elecciones en total. Para cada elección, se informan las coordenadas geográficas de cada alternativa además de si se eligió a la alternativa de la derecha, de la izquierda, o si el usuario manifestó que ambas son similares. Un ejemplo de la encuesta se ve en la Figura 1.

3.2. Análisis de imágenes

Para traducir la información contenida en las imágenes en atributos fácilmente interpretables, se obtuvieron atributos de bajo y alto nivel. Mientras que los de bajo nivel se refieren a información



Figura 1: Ejemplo de experimento de elección presentado en *Place Pulse*

contenida en los píxeles en sí, tales como brillo, saturación o contraste, los de alto nivel son segmentaciones realizadas mediante técnicas de *machine learning* para predecir a qué elemento del espacio, tales como vereda, vegetación o vehículos, corresponde cada píxel. En las secciones que siguen se explicará cada uno en detalle.

3.2.1. Atributos de bajo nivel

Estos atributos se caracterizan por ser conseguidos con un costo muy bajo y dar información bastante gruesa de las imágenes. Cada variable se define como se indica a continuación:

1. Bordes: El porcentaje de píxeles de la imagen que se detectan como un borde (esto es, donde existe un cambio brusco en luminosidad). Un ejemplo se muestra en la Figura 2a.
2. Blob: El número de blobs (*binary large objects*), definidos como un grupo de píxeles conectados con características homogéneas, tales como color y textura, que se encuentra en una imagen. Un ejemplo se muestra en la Figura 2b.
3. Promedios de canales HSL: valores promedio para los canales de tinte, saturación y luminosidad, del espacio de colores HSL, con valores entre 0 y 255. Se eligió este espacio por estar basado en la percepción del ojo humano al color, además del bajo costo computacional que conlleva trabajar con él (Lissner y Urban, 2012).
4. Desviación estándar de canales HSL: desviación estándar para los canales de tinte, saturación y luminosidad. Esta última representa adecuadamente el contraste de una imagen (Moulden et al., 1990).

En un set de prueba aleatorio con más de 8.000 elecciones válidas observadas se estimaron modelos como los que se muestran más adelante que, además, incluyeron histogramas de color para las imágenes en los canales HSV. Los parámetros relacionados a estos histogramas resultaron ser no significativos, por lo que se descartó su uso para la prueba con el set de datos completo. También se descartaron otros atributos de bajo nivel utilizados por estudios anteriores (Naik et al., 2014; Ordonez y Berg, 2014) debido a su alta dimensionalidad, lo que imposibilita hacer una adecuada interpretación de los parámetros obtenidos.



Figura 2: Ejemplos de atributos de bajo nivel obtenidos

3.2.2. Atributos de alto nivel: Segmentación semántica de imágenes

El objetivo principal de los modelos que se estiman en esta investigación es definir qué elementos de una calle hacen que esta se ajuste más o menos a cierto atributo cualitativo. Debido a esto, se hace necesario inferir qué elementos de una calle están presentes en una imagen y cuál es su peso relativo. Para esto, se ocupó segmentación semántica de imágenes, que consiste en la asignación de etiquetas a cada uno de los píxeles de una imagen, donde estas etiquetas pertenecen a categorías previamente definidas. En este trabajo se utilizó el modelo de segmentación semántica *SegNet* (Badrinarayanan et al., 2015), el cual se basa en una red neuronal convolucional profunda². Para este caso específico, se entrenó el modelo con imágenes etiquetadas entregadas por los creadores de *SegNet* que consideran 11 clases: auto, calzada, ciclista, cielo, edificio, peatón, poste, señalética, valla, vegetación y vereda.

A pesar de que este modelo entrega muy buenos resultados para las imágenes de entrenamiento, que corresponden a un video tomado en las cercanías de Cambridge (Reino Unido), los resultados son variables para las imágenes utilizadas en *Place Pulse*. Como se ve en la Figura 3, a pesar de que a grandes rasgos los resultados parecen ser razonables, existen bastantes errores en la salida del algoritmo.

En el set de prueba aleatorio mencionado anteriormente, se estimaron modelos con varias especificaciones para la información contenida en la salida del modelo de segmentación semántica. La que entregó un mejor ajuste fue la que se muestra en la Ecuación (1), donde p_{ics} es la proporción de píxeles de la clase c en la mitad s de la imagen i . En otras palabras, se representa el efecto que tiene cada atributo de alto nivel con un efecto decreciente sobre la elección, y además se separa este efecto en la mitad superior e inferior de la imagen para representar de manera aproximada aquellos elementos sobre y bajo la línea del horizonte.

$$HLF_{ics} = \log(100p_{ics} + 1) \quad (1)$$

²Debido a la limitada extensión que debe tener este artículo, lamentablemente no podemos ahondar en el funcionamiento de los modelos de segmentación semántica en general. Se recomienda al lector revisar los apuntes del curso “Convolutional Neural Networks for Visual Recognition” (Li et al., 2017), que se encuentran en <http://cs231n.github.io/>.

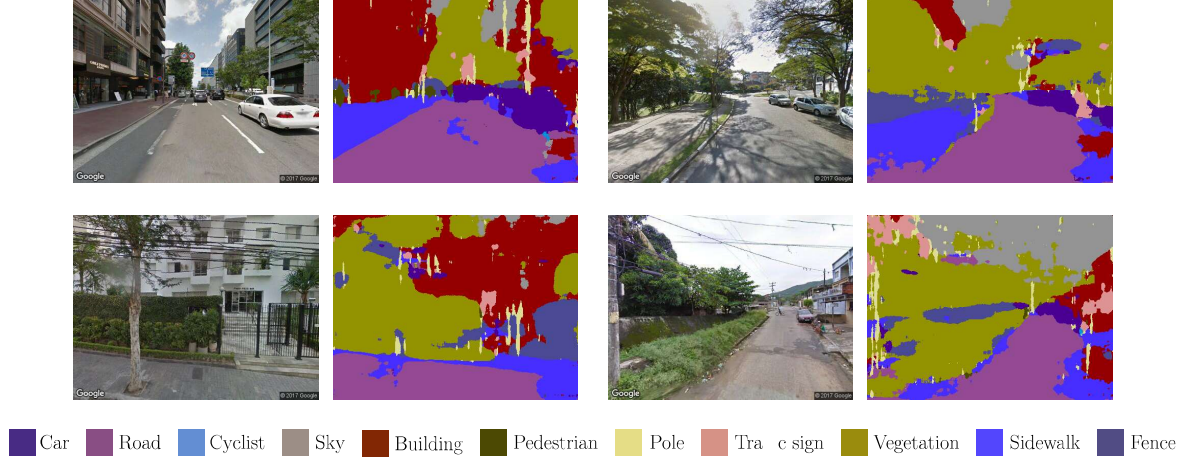


Figura 3: Ejemplos de cuatro imágenes utilizadas por *Place Pulse* (izquierda) segmentadas con el uso de *SegNet* (derecha)

3.3. Metodología de estimación

Para poder lograr resultados que tengan un buen poder explicativo además de predictivo, proponemos utilizar modelos que se funden sobre la teoría de utilidad aleatoria (McFadden, 1974; Domencich y McFadden, 1975). Esta postula que las personas perciben cierta utilidad al elegir una alternativa por sobre otra, y que eligen la que les proporciona la mayor utilidad. El modelo también supone que esta utilidad tiene una parte estocástica y otra determinística, por lo que solo se puede inferir la probabilidad de que cierto sujeto elija cierta alternativa.

Uno de los modelos más utilizados dentro de esta familia es el logit, que postula que la probabilidad de que el individuo n elija la alternativa i , P_{in} , está dada por la Ecuación (2), donde V_{in} es la parte determinística de la utilidad descrita anteriormente y C_n es el conjunto de alternativas disponibles para el individuo n . V_{in} es una función, usualmente lineal, de los atributos de las alternativas y características de los usuarios, X_{in} , que se relacionan con la utilidad a través de parámetros β por estimar. El parámetro μ se confunde perfectamente con los parámetros β , por lo que cuando V_{in} es lineal se fija en uno.

$$P_{in} = \frac{\exp \mu V_{in}(X_{in}; \beta)}{\sum_{j \in C_n} \mu V_{jn}(X_{in}; \beta)} \quad (2)$$

Para este caso específico, se construyeron funciones de utilidad como la que se muestra en la Ecuación (3), donde LLF_i son los atributos de bajo nivel (*Low Level Features*) y $HLLF_{ics}$ son los atributos de alto nivel (*High Level Features*), como se describieron en la sección anterior. Los parámetros β se estiman a través de maximización de verosimilitud, que en este caso específico se realizó mediante el software Biogeme (Bierlaire, 2003). Debido a que podría existir un sesgo sistemático en favor de las imágenes presentadas a la izquierda o derecha, se agregó una variable

muda o constante I_i que es igual a uno si la imagen i se presenta a la izquierda, y cero en caso contrario.

$$V_i = \beta_0 I_i + \sum_k \beta_k^{LLF} LLF_{ik} + \sum_c \sum_s \beta_c^{HLF} HLF_{ics} \quad (3)$$

En la estimación se omitieron aquellas imágenes que se detectaran como *outliers*, que incluyen imágenes distorsionadas y escenas nocturnas o de interior. En total, se dejaron fuera 3.729 imágenes, que corresponden al 3,37 % del total disponible.

4. Resultados

La Tabla 2 muestra los resultados de los modelos de elección discreta calibrados para cada uno de los seis atributos cualitativos considerados en *Place Pulse*. Todos muestran una mejora significativa con respecto a un modelo que contiene solo un intercepto (la constante asociada al lado donde se presenta la imagen), y contienen una buena cantidad de parámetros altamente significativos ($p < 0,01$). A pesar de esto, el ajuste de estos modelos es relativamente bajo, y son capaces de predecir correctamente tan solo entre un 60 % y un 65 % de las elecciones observadas (kappas de Cohen (1960) entre 0,2 y 0,3). A pesar de que estas cifras son menores a las logradas por Dubey et al. (2016) con el mismo set de datos utilizando metodologías de *machine learning* (entre 63 % y 74 % de las elecciones, kappas de Cohen (1960) entre 0,26 y 0,48), se debe tener en cuenta que la función objetivo de estos modelos es la maximización de la verosimilitud, y no la del porcentaje de elecciones correctamente observadas.

Los resultados muestran que la presencia de actividad humana, ya sea a través de autos, peatones o ciclistas, hacen que una calle se sienta más animada y segura. Por otra parte, la presencia de elementos naturales, como vegetación o que se vea el cielo, hacen que las imágenes se perciban como más aburridas y menos animadas. Esto puede deberse a la presencia de imágenes en zonas rurales que son más bien bucólicas pero que no evocan ideas de actividad o entretenimiento.

Con los estimadores máximos verosímiles obtenidos, se calculó la utilidad promedio de cada atributo para cada imagen. Al analizar las correlaciones entre estos valores, que se muestra en la Tabla 1, se ve que los atributos cualitativos están bastante correlacionados entre sí. Por ejemplo, los atributos “Deprimente” y “Bonito” están casi perfectamente representados por una recta con pendiente negativa, mientras que algo similar sucede con “Seguro” y “Adinerado” con una pendiente positiva. Esto sucede probablemente por las relaciones que existen entre estos conceptos no son independientes. “Animado” y “Aburrido”, por ejemplo, pueden ser considerados como antónimos, por lo que es esperable que estos atributos tengan una fuerte correlación negativa.

En la Figura 4 se muestran las imágenes con menores (izquierda) y mayores (derecha) atributos cualitativos del total de imágenes analizadas. En general, estas imágenes representan de manera razonable los distintos adjetivos, y permiten hacer una interpretación más rápida de la Tabla 2. Se ve, por ejemplo, que las imágenes menos aburridas y más animadas son las que tienen mayor presencia

Tabla 1: Matriz de correlación para atributos estimados

	Aburrido	Adinerado	Animado	Bonito	Deprimente	Seguro
Aburrido						
Adinerado	-0,66					
Animado	-0,95	0,77				
Bonito	-0,39	0,88	0,45			
Deprimente	0,61	-0,95	-0,68	-0,96		
Seguro	-0,52	0,96	0,67	0,91	-0,95	

de actividad, mientras que las más aburridas muestran calles desoladas y sin mucha presencia de actividad humana. Las más seguras tienden a ser más verdes, mientras que las menos son más bien grises y oscuras. En general, estos casos extremos parecieran representar relativamente bien los atributos cualitativos, por lo que probablemente tienen más errores en casos donde las diferencias son más sutiles.

5. Conclusiones

Este trabajo propuso una metodología novedosa para cuantificar atributos cualitativos de espacios públicos a una escala masiva. Utilizando complementariamente las ventajas de los modelos basados en la teoría de utilidad aleatoria con modelos de segmentación semántica basados en técnicas de *machine learning*, se lograron modelos que, a pesar de tener un menor poder predictivo, presentan mejoras sustantivas para explicar el comportamiento de las personas y realizar recomendaciones de política pública. Creemos que estos resultados son prometedores y presentan un camino de investigación que puede ser un aporte significativo para la comunidad científica y la toma de decisiones en cuanto a diseño de infraestructura y espacios públicos. En particular, este trabajo abre varias preguntas que nos gustaría abordar a futuro.

En primer lugar, los modelos de elección discreta presentados aquí no toman en cuenta el error asociado a la segmentación que se hizo de las imágenes para extraer atributos de alto nivel, lo que podría provocar un sesgo en los parámetros obtenidos. La solución definitiva está en mejorar la especificación del modelo de elección. Sin embargo, si se mejora la capacidad predictiva del modelo de segmentación semántica es posible que este sesgo se reduzca a un mínimo. Para lograr esto, probablemente sea necesario clasificar un set más grande de imágenes para reentrenar el modelo, lo que podría ser bastante costoso. Además de esto, se puede reentrenar la red considerando solo imágenes que provengan tan solo de un país o continente. A pesar de que esta reducción de variabilidad contribuiría a corregir posibles sesgos, reduciría la capacidad del modelo de capturar percepciones a una escala más bien universal.

En segundo lugar, creemos que los atributos de alto nivel utilizados aquí no son suficientes como para poder estimar los modelos con suficiente precisión. Existen algunas variables que no fueron consideradas, como presencia de graffiti, nivel de detalle en las fachadas de los edificios, o cantidad de ventanas que dan a la calle, que podrían ser relevantes en la evaluación que hacen las

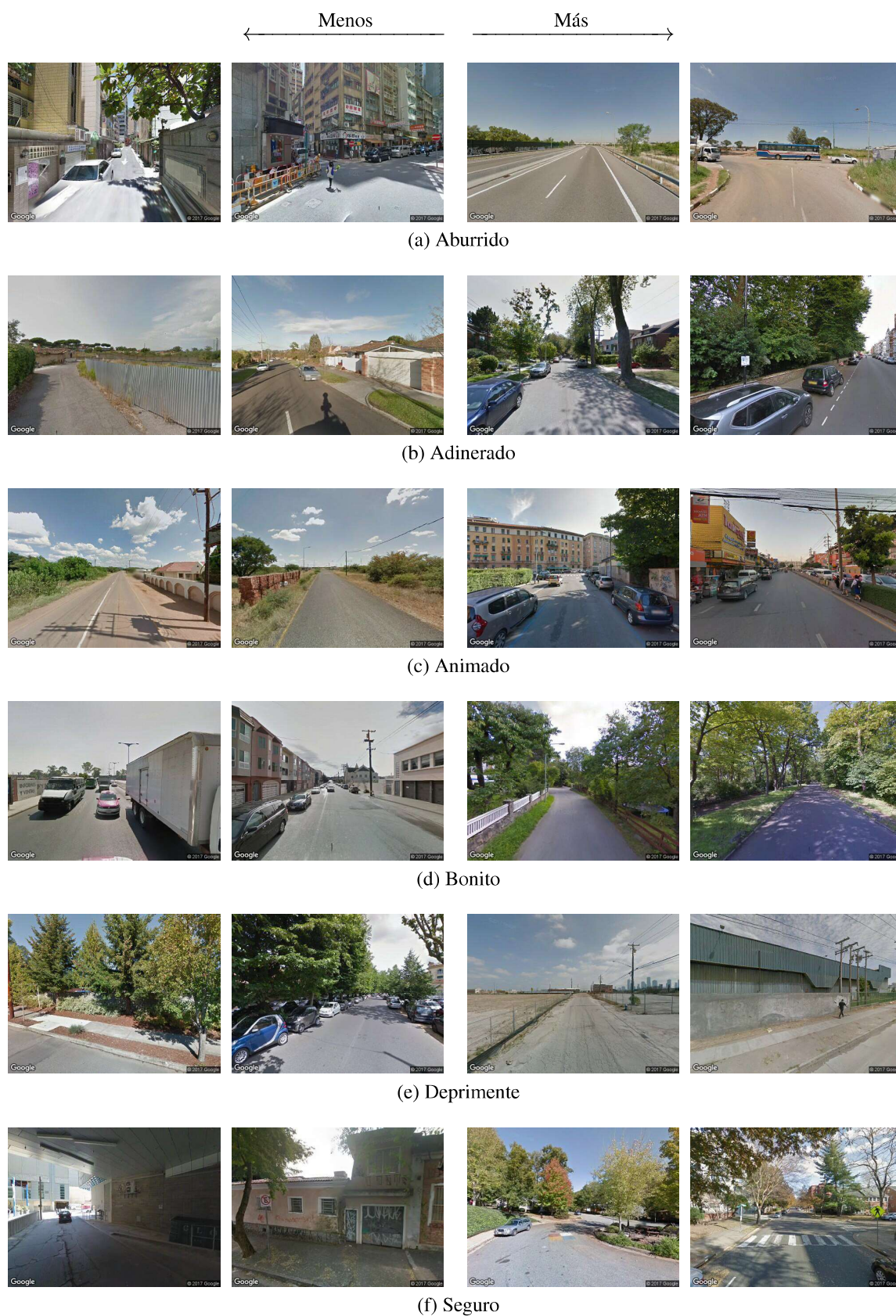


Figura 4: Imágenes con menor y mayor probabilidad de representar los atributos cualitativos considerados



Figura 5: Imágenes con similar utilidad para el modelo “Adinerado”

personas (ver, por ejemplo, Porta y Renne, 2005). También podría ser relevante identificar algunos elementos arquitectónicos que pudieran ser indicadores de atributos cualitativos. Por ejemplo, la Figura 5 muestra dos imágenes con utilidades muy similares para el modelo del atributo “Adinerado”. Sin embargo, es altamente probable que la mayoría de las personas considere que la Figura 5a corresponde a un lugar más adinerado que el de la Figura 5b. Un entrenamiento adecuado del modelo de redes neuronales podría hacer una clasificación más refinada de los elementos presentes en las imágenes, pudiendo reproducir de mejor forma las elecciones realizadas por los usuarios que respondieron la encuesta *Place Pulse*.

Una vez superados estos obstáculos, nuestra intención es comparar estos modelos con otros estimados con métodos de *machine learning*. Haciendo esto, va a ser posible evaluar si las ventajas en capacidad explicativa de los modelos sustentados en la teoría de utilidad aleatoria sacrifican capacidad predictiva.

Si los poderes predictivos de estos modelos son adecuados, nuestro objetivo posteriormente es generar mapas sensoriales para distintas ciudades utilizando los modelos calibrados e imágenes tomadas de Google Street View. Con esto, es posible analizar las relaciones que existan entre estos atributos cualitativos y estadígrafos geográficos, como podría ser ingreso medio, criminalidad, u otros.

Finalmente, una vez validada esta metodología, nuestra intención es aplicar una encuesta similar a *Place Pulse* pero cuyo objetivo sea medir atributos cualitativos relacionados al uso de modos de transporte sustentable. Esto podría servir a gobiernos locales para que identifiquen puntos conflictivos dentro de sus territorios y mejoren la situación de los usuarios de estos modos.

A pesar de que estos modelos aún tienen bastante espacio para mejorar, creemos que abren una línea de investigación interesante y que puede aportar a la comprensión de la forma en que las personas perciben y procesan el espacio que las rodea.

Agradecimientos

Esta investigación fue parcialmente financiada por el Instituto de Sistemas Complejos de Ingeniería (ICM: P-05-004-F, CONICYT:FBO16) y el Centro de Desarrollo Urbano Sustentable (CEDEUS,

CONICYT/FONDAP 15110020).

Referencias

- Antonakos, C. L. (1995). Environmental and travel preferences of cyclists. *Transportation Research Part A: Policy and Practice*, 29(1):85.
- Badrinarayanan, V., Kendall, A., y Cipolla, R. (2015). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. pages 1–14.
- Bierlaire, M. (2003). BIOGEME: a free package for the estimation of discrete choice models. In *Swiss Transport Research Conference*, pages 1–27, Ascona, Switzerland.
- Cohen, D., Spear, S., Scribner, R., Kissinger, P., Mason, K., y Wildgen, J. (2000). “Broken Windows” and the Risk of Gonorrhea. *American Journal of Public Health*, 90(2):230–236.
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, XX(1):37–46.
- Domencich, T. y McFadden, D. (1975). *Urban Travel Demand: A Behavioral Analysis*. North-Holland Publishing Company, Amsterdam.
- Drucker, H., Burges, C. J., Kaufman, L., Smola, A., y Vapnik, V. (1997). Support vector regression machines. *Advances in neural information processing systems*, 9(x):155–161.
- Dubey, A., Naik, N., Parikh, D., Raskar, R., y Hidalgo, C. A. (2016). Deep learning the city: Quantifying urban perception at a global scale. *Lecture Notes in Computer Science*, 9905 LNCS:196–212.
- Ellaway, A., Macintyre, S., y Bonnefoy, X. (2005). Graffiti, greenery, and obesity in adults: secondary analysis of European cross sectional survey. *British Medical Journal*, 331(7517):611–612.
- Hunt, J. D. y Abraham, J. E. (2007). Influences on bicycle use. *Transportation*, 34(4):453–470.
- Hyodo, T., Suzuki, N., y Takahashi, K. (2000). Modeling of Bicycle Route and Destination Choice Behavior for Bicycle Road Network Plan. *Transportation Research Record: Journal of the Transportation Research Board*, 1705(1):70–76.
- Khisty, C. (1994). Evaluation of pedestrian facilities: beyond the level-of-service concept. *Transportation Research Record*, 1438:45–50.
- Krizhevsky, A., Sutskever, I., y Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances In Neural Information Processing Systems*, pages 1–9.
- Kuipers, M. A. G., van Poppel, M. N. M., van den Brink, W., Wingen, M., y Kunst, A. E. (2012). The association between neighborhood disorder, social cohesion and hazardous alcohol use: A national multilevel study. *Drug and Alcohol Dependence*, 126(1-2):27–34.

- Latkin, C. A. y Curry, A. D. (2003). Stressful Neighborhoods and Depression: A Prospective Study of the Impact of Neighborhood Disorder. *Journal of Health and Social Behavior*, 44(1):34–44.
- Li, F.-F., Johnson, J., y Yeung, S. (2017). CS231n: Convolutional Neural Networks for Visual Recognition.
- Lissner, I. y Urban, P. (2012). Toward a unified color space for perception-based image processing. *IEEE Transactions on Image Processing*, 21(3):1153–1168.
- Lynch, K. (1960). *The image of the city*. MIT Press.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics*, pages 105–142. Academic Press.
- Moulden, B., Kingdom, F., y Gatley, L. F. (1990). The standard deviation of luminance as a metric for contrast in random-dot images. *Perception*, 19(1):79–101.
- Naik, N., Philipoom, J., Raskar, R., y Hidalgo, C. (2014). Streetscore - predicting the perceived safety of one million streetscapes. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 793–799.
- Ordonez, V. y Berg, T. L. (2014). Learning high-level judgments of urban perception. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8694 LNCS(PART 6):494–510.
- Porta, S. y Renne, J. L. (2005). Linking urban design to sustainability: formal indicators of social urban sustainability field research in Perth, Western Australia. *Urban Design International*, 10(1):51–64.
- Porzi, L., Buló, S. R., Lepri, B., y Ricci, E. (2015). Predicting and Understanding Urban Perception with Convolutional Neural Networks. *Proceedings of the ACM International Conference on Multimedia (ACM-MM)*, pages 139–148.
- Quercia, D., Schifanella, R., y Aiello, L. M. (2014). The Shortest Path to Happiness: Recommending Beautiful, Quiet, and Happy Routes in the City. In *Proceedings of the 25th ACM conference on Hypertext and social media*, Santiago, Chile.
- Schroeder, H. W., Anderson, L. M., y Daniel, T. C. (1984). Perception of personal safety in urban recreation sites. *Journal of Leisure Research*, 16(2):178–194.
- Shriver, K. (1997). Influence of environmental design on pedestrian travel behavior in four Austin neighborhoods. *Transportation Research Record: Journal of the Transportation Research Board*, 1578(1):64–75.
- Weber, R., Schnier, J., y Jacobsen, T. (2008). Aesthetics of Streetscapes: Influence of Fundamental Properties on Aesthetic Judgments of Urban Space. *Perceptual and Motor Skills*, 106(1):128–146.
- Zacharias, J. (2001). Pedestrian Behavior and Perception in Urban Walking Environments. *Journal of Planning Literature*, 16(3):3–18.

Tabla 2: Resultados para la modelación de atributos cualitativos

Parámetro (estadístico t)	Aburrido	Adinerado	Animado	Bonito	Deprimente	Seguro
Constante izq.	-0.0101 (-1.54)	-0.0327** (-5.48)	-0.0427** (-9.64)	-0.0446** (-7.88)	-0.0221** (-3.38)	-0.0316** (-8.39)
Auto (sup.)	0.0398** (4.24)	-0.106** (-12.24)	-0.0555** (-8.71)	-0.0828** (-10.28)	0.0814** (8.69)	-0.0541** (-9.96)
Auto (inf.)	-0.0546** (-9.35)	0.0797** (14.25)	0.119** (28.52)	-0.0149** (-2.83)	-0.0595** (-10.04)	0.0699** (19.66)
Calzada (sup.)		-0.0396* (-2.8)	-0.0217* (-2.1)			
Calzada (inf.)	-0.0292* (-2.44)	0.211** (18.5)	0.14** (16.31)	0.121** (11.21)	-0.133** (-10.96)	0.212** (29.39)
Ciclista (sup.)		0.206** (3.02)		0.275** (4.28)	-0.138 (-1.82)	0.145** (3.42)
Ciclista (inf.)	-0.0558* (-2.33)	-0.04 (-1.85)	0.0832** (5.18)	-0.0969** (-4.8)	0.0464 (1.95)	-0.0258 (-1.88)
Cielo (sup.)	0.0345** (5.71)	-0.0334** (-5.98)	-0.0153** (-3.8)	-0.0533** (-10.38)	0.0467** (8)	-0.0183** (-5.14)
Cielo (inf.)	-0.0196* (-2.2)	0.0165 (1.94)				-0.0196** (-3.69)
Edificio (sup.)	-0.0613** (-10.53)	-0.018** (-3.48)	0.063** (16.01)	-0.103** (-21.04)	0.0451** (8.47)	-0.0235** (-6.88)
Edificio (inf.)	0.0377** (4.99)	-0.0528** (-7.39)	-0.0445** (-8.25)	-0.0804** (-11.7)	0.0637** (8.04)	-0.0362** (-7.89)
Peatón (sup.)	-0.0659* (-2.32)	0.106** (4.03)	0.061** (3.2)		-0.0625* (-2.21)	0.102** (6.11)
Peatón (inf.)	-0.196** (-13.58)	-0.033* (-2.44)	0.148** (14.82)		-0.0484** (-3.27)	0.0233* (2.73)
Poste (sup.)		0.0337* (2.54)				-0.0237* (-2.79)
Poste (inf.)		-0.208** (-16.13)	-0.0926** (-9.33)	-0.105** (-8.44)	0.0666** (4.55)	-0.166** (-20.31)
Señalética (sup.)	-0.0492** (-4.87)	0.0883** (8.85)	0.0597** (8.49)	0.102** (11.32)	-0.0732** (-7.07)	0.0539** (8.56)
Señalética (inf.)			0.032** (3.92)	-0.0316** (-3.04)	0.0484** (4.08)	
Valla (sup.)	-0.0578** (-7.37)	0.102** (14.46)	0.0785** (14.76)	0.111** (16.4)	-0.0885** (-11.37)	0.0844** (18.43)
Valla (inf.)	-0.066** (-8.55)		0.0328** (6.18)	0.0528** (7.85)	-0.0713** (-9.31)	0.015** (3.35)
Vegetación (sup.)	-0.0234* (-2.32)		0.0328** (4.77)	0.0183* (2.02)		0.0216** (3.61)
Vegetación (inf.)	0.0505** (7.28)	0.0658** (10.02)	-0.0452** (-9.32)	0.136** (22.22)	-0.0948** (-13.65)	0.0824** (19.86)
Vereda (sup.)	0.0679** (5.17)	-0.151** (-11.18)	-0.115** (-11.5)	-0.122** (-10.69)	0.128** (9.72)	-0.166** (-21.4)
Vereda (inf.)		0.0621** (6.27)	0.0584** (8.01)	-0.0492** (-5.27)		0.0627** (9.97)

Continúa en la siguiente página. ** $p < 0, 01$; * $p < 0, 05$.

Tabla 2: Resultados para la modelación de atributos cualitativos

Parámetro (estadístico t)	Aburrido	Adinerado	Animado	Bonito	Deprimente	Seguro
Bordes	-3.96** (-11.89)	6.85** (21.77)	5.28** (23.12)	7.77** (27.04)	-7.62** (-22.24)	5.12** (25.81)
Blobs	-0.00886** (-8.47)	0.00734** (7.76)	0.00837** (11.9)		-0.00233* (-2.24)	0.00519** (8.76)
Media tinte	-0.00173** (-3.08)	0.00398** (7.45)	0.000312 (0.77)	0.00393** (7.59)	-0.00339** (-6.12)	0.000713* (2.07)
Media luminosidad	0.00264** (5.87)	0.00122** (2.98)	0.0000638 (0.21)	0.00111** (3.11)	-0.00206** (-4.66)	0.00295** (11.31)
Media saturación	0.00525** (5.92)	0.000811 (0.99)	-0.00273** (-4.6)	0.00238** (3.13)	-0.00156 (-1.81)	0.00193** (3.72)
Desv. est. tinte	-0.0014 (-1.28)	0.00901** (8.96)	0.00577** (7.84)	0.0165** (17.58)	-0.00885** (-8.24)	0.012** (18.92)
Desv. est. luminosidad	-0.00502** (-6.71)	0.00609** (9)	0.00365** (7.19)	0.00745** (11.55)	-0.00636** (-8.53)	0.00384** (8.92)
Desv. est. saturación	-0.00683** (-6.94)	0.00482** (5.36)	0.00656** (9.96)	0.00147 (1.75)	-0.00261* (-2.69)	0.00156* (2.76)
Observaciones	98.304	121.569	217.475	140.770	102.257	297.930
Número de parámetros	25	28	28	25	26	29
log-verosimilitud	-65.370,46	-79.430,62	-143.510,22	-89.135,80	-66.603,73	-198.679,58
Test de razón de verosimilitud ^a	5.535,88	9.642,62	14.373,64	16.825,58	8.542,00	15.594,48
rho cuadrado	0,041	0,057	0,048	0,086	0,060	0,038
Probabilidad elegida sobre 0,5 ^b	60,0 %	61,6 %	60,7 %	64,5 %	62,0 %	59,5 %

Continuada desde la página anterior. ** $p < 0, 01$; * $p < 0, 05$.

^a Tests comparando con un modelo que solo considera el intercepto. Todos indican un rechazo a la hipótesis nula con $p < 0, 01$.

^b Definido como la proporción de las elecciones observadas cuya probabilidad estimada de ocurrencia es mayor que 0,5.