



Filter variables

Description

When did you decide to grow a beard? Not every event over which a person assumes agency has a singular moment or an origin. Nor does it, of necessity, involve a decision. A slightly less gendered scenario is that of strolling through the neighbourhood â?? perhaps as exercise during pandemic restrictions.

Sometimes we just [wander](#) â?? a process often accountable in terms of habit, custom, happenstance, curiosity, and instant responses to environment. At times we may overlay this unselfconscious phenomenological ([Heideggerian](#)) engagement in the world with goals, plans, criteria and constraints. At least we deploy the talk, methods and practices of such â??rational decision making.â?•

Online shopping

These deliberative practices show up when shoppers select from a range of goods and services online. The variety may be illusory as products exhibit ever diminishing differentiation, but developers configure their e-commerce platforms so that customers can filter according to criteria. Here, I take a *criterion* as another term for a *variable*.

When I book a hotel room I am presented with a series of filters. I can select according to my budget, whether breakfast is included, the number of room occupants, double or single bed, kitchenette, bath, Internet, proximity to the town centre, leisure facilities, and customer rating scores. These are *criteria*, i.e. variables tabulated in a database for each accommodation option. Some variables are hidden from view. Here are some categories of hidden variables.

1. Unavailable variables. There are candidate variables that may or may not be stored in the vendorâ??s databases. Some may be hidden to simplify the customer decision making process, or they are just too much trouble to catalogue accurately. To ward off competition and preserve reputation the vendors may want also to keep some variables secret.

Here are some variables that until now Iâ??d never thought to explore as a customer in search of a hotel room, but are likely to be somewhere in the organisationâ??s record system: floor area of the

rooms, ceiling height, fire safety records, the health certification of the restaurant, the kind of shampoo placed in the bathrooms, the noise rating of the ventilation system, whether the windows can be opened, level of soundproofing between rooms, number of people who died while in the hotel.

2. Derivable variables. Some variables are unavailable but could be derived simply by algorithm or by the customer's own calculation. From publicly available map information you can derive proximity to shops, public transport, beaches or gyms. With less certainty, from the number of rooms and the price you can estimate the level of personal service the hotel is likely to offer.

3. Latent variables. These are variables to which you may not even be able to assign a name. We may have difficulty putting these variables into words. They may be apprehended by looking at photographs and reading between the lines of customer comments. In instrumental terms, the character of such variables comes to the fore in statistical analysis and machine learning algorithms. Here, the concept of the variable merges with ideas about classification. The floor area of a room is a variable. It also helps define a class: e.g. all bedrooms under 20 square metres.



Conceptual clusters

In a much earlier post ([Brain scans and creativity](#)) I discussed a theory about *conceptual clustering* as a way of understanding how we categorise things. The example I used there also serves to illustrate the presence of hidden variables.

Here's the example. A particular room contains a cooking stove, a cupboard, a refrigerator, and a toaster. Other rooms contain items such as easy-chairs, beds, and coffee tables. Many examples of such real-world combinations (as lists of words) are fed into a neural network, an algorithmic system contrived to model aspects of human cognition, or at least brain function.

After processing all the learning examples as inputs the system is presented with a single word such as 'toaster' as input. The system then produces as output a set of other descriptors strongly associated with 'toaster', eg 'cooking stove', 'refrigerator', 'dishwasher'. In other words the system presents a description of the contents of a typical kitchen.

There's no concept of room type in the network, just clusters of components and weighted relationships between them derived from algorithms that derive weightings from hundreds of examples

of actual room content lists. When so “trained” the system behaves as if the various attributes are listed under room headings. The system won’t generate the names *kitchen*, *bedroom*, *bathroom*, etc, but acts as if it has. Room type is a hidden variable. It’s latent in the system.

When it doesn’t have a name

Conceptual clustering is a common explanation of the many cases where we lack the vocabulary of the expert, but we know what we expect to see if we are in a *vestibule*, *atrium*, *ante-room*, *conservatory*, *sacristy*, even if we lack the vocabulary to identify those spaces.

The process is analogous to the way that people learn their first language. It is possible to be a competent language speaker while knowing nothing about grammatical or syntactic categories: nouns, verbs, prepositions, clauses, etc. These categories are effectively the hidden variables of language competence. In fact, some automated [speech-to-text](#) translation systems bypass grammar and “learn” the appropriate grammar through myriad instances of well-formed sentences from a training set.

Person, woman, man, camera, tv

But there’s a further sophistication to the concept of conceptual clustering. Some categories derived in this way are fluid and unnamable. The interesting aspect of the room content experiment is the emergence of new categories, new implied instantiations of the latent room class variable. If you identify a new room that has a toaster and a bedside table then in the automated neural network system the components (attributes) reconfigure themselves to present an arrangement that looks a bit like a bed sitting room. The combination of a chandelier and a wash basin might introduce a wardrobe and a bed into the ensemble producing a new rogue category, or something like a bordello in a raunchy period drama.

In so far as this simple explanation of conceptual clustering matches human cognition, it’s apparent that the world is full of hidden variables. In the case of the room example, the variable of room type is unnamed and hidden, as are the candidate instantiations. Returning to the hotel booking example, instead of room type, conceptual clustering may yield other vague variables, such as *pleasantness*, *conducive to study*, *secure*, *cheerful*, *bright*, *healthy* and other variables unnamed and invisible but that we can presume influence our decision-making processes.

One of the challenges of statistical and machine learning methods, especially those that try to establish correlations and causes is the identification and influence of *hidden variables* “unknown factors latent in the data, or *confounding variables* that are invisible because we don’t have the right kind of data that brings them to light.

Bibliography

- Lakoff, George. 2003. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press
- Ngai, Sianne. 2012. *Our Aesthetic Categories: Zany, Cute, Interesting*. Cambridge, MA: Harvard University Press

- Rumelhart, D. E., and J. L. (eds) McClelland. 1987. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge, Mass.: MIT Press

Note

- The picture is a faded bedroom in Brodsworth Hall, near Brodsworth, 5 miles north-west of Doncaster in South Yorkshire.

Category

1. Architecture

Tags

1. classification
2. conceptual clustering
3. variable

Date Created

January 2, 2021

Author

rcoyne99

default watermark