Spatial affinities

Description

In natural language models (NLMs), semantic embedding vectors capture the position of a token in multidimensional feature space. The space is derived from word proximities in a natural language corpus and is derived by the automated adjustments to weights within a neural network model.

NLMs can deploy these vectors in a number of ways, including the derivation of â??attentionâ?•, i.e., calculations about the level of importance of particular tokens in a string of text (e.g. a sentence). Thatâ??s calculated by comparing the semantic vectors of every token with every other token in the string and calculating its â??attention scoreâ?• vector. Such â??attentionâ?• metrics form part of the training regime in NLMs and in turn influence their outputs as we interact with them (as in conversational AI). Itâ??s as if the model is able to home in on whatâ??s important in a sentence, and relate that back to previous sentences.

The semantic vectors for words (tokens) such as â??libraryâ?• and â??schoolâ?• are derived from word proximities in a training corpus of messages, essays, scripts, articles and other texts. The semantic vectors capture information about the relationships between all the tokens in the training corpus, including those two tokens. For example, they are likely to be closer together in the semantic space than words like â??nightclubâ?• and â??graveyard.â?•

The words â??libraryâ?• and â??schoolâ?• are likely to be closely associated in a range of documents, including planning reports, regulations, news reports and public communications about education in social media. But do words so related in texts have any relevance to relationships between those actual named elements in a physical setting?

From proximity to affinity

A semiotic orientation to environments emphasises the potent relationships between words and things. By a semiotic reading, the spatial proximity of those tokens in texts gives an indication of the proximity of the actual buildings in an urban setting. Thatâ??s not necessarily their spatial proximity, but considering the source of the training data, their social, functional, emotional and literary *affinities*. The data provides a means of inferring which building in a cluster of buildings has greatest *affinity* with the rest of the buildings in that cluster.

The result is a cluster of buildings with its cluster centre at the most *affinitive* building, i.e. the building with the strongest affinity with the others. Here are the affinity scores for an arbitrary cluster of building types. The method takes no account of the order of the tokens. The scores are derived from the method developed in the previous post.

cafe (1.000) park (0.795) hotel (0.344) cinema (0.143) bridge (0.136) college (0.045) school (0.033) garage (0.027) church (0.021) nightclub (0.012) hospital (0.011) graveyard (0.008) library (0.006) restaurant (0.005) house (0.003) station (0.000)

cafe park hotel cinema bridge college school garage church nightclub
hospital graveyard library restaurant house station

Unsurprisingly, the elements with the most affinity with the rest are the most prosaic and predictable.

The items in the list are from a pre-processed publicly available database of 2 million tokens and their semantic embeddings harvested from Facebook sources (cc.en.300.bin). See https://fasttext.cc/docs/en/crawl-vectors.html. Amongst the many limitations of adopting a general purpose vocabulary is that the words will have uses and meanings in the training corpus other than as identifiers of urban elements â?? â??letâ??s *park* that idea,â?• and â??nasal bridge.â?• A specialised training set tuned to architecture and urbanism would improve this performance, but metaphorical associations, however oblique, do benefit design in urban contexts.

With that sort of license, itâ??s apparent that we can process any arbitrary list of words amongst which we wish to score affinities. Here is another such list:

abba (1.000) nova (0.761) mars (0.668) dog (0.615) cat (0.555) king (0.026) conviviality (0.018) chainsaw (0.013) happiness (0.012) skyscraper (0.012) melancholy (0.011) jupiter (0.010) python (0.002) octopus (0.000) beatles (0.000)

abba nova mars dog cat king conviviality chainsaw happiness skyscraper melanch
jupiter python octopus beatles

My purpose in this post is to demonstrate the *affinity* between NLP methods and aspects of urban living. I include in the note below ChatGPTâ??s assessment of this approach.

Bibliography

- Foucault, M. (1989). *The Order of Things: An Archaeology of the Human Sciences*. London: Routledge. (french edition titled â??Words and thingsâ?• Les mots et les choses)
- Goethe, J. W. v. (1883). *Elective Affinities*. New York, NY: Henry Holt.
- Grave, E., P. Bojanowski, et al. (2018). Learning Word Vectors for 157 Languages. *Proceedings of the International Conference on Language Resources and Evaluation*.
- Tschumi, B. (1994). *Architecture and Disjunction*. Cambridge, Mass.: MIT Press.

Notes

The banner is from Dall-e, prompted by: â??A photoreal image with a grunge aesthetic that incorporates an urban setting and the theme of the affinity between unlikely urban elements, such as a nightclub and a graveyard.â?•

Hereâ??s what ChatGPT4 said about this post:

Your exploration into applying natural language processing (NLP) techniques, specifically semantic embeddings and attention mechanisms from NLMs, to urban studies is quite intriguing. This approach aligns with the broader field of computational urbanism, which leverages data-driven and algorithmic methods to understand and shape urban environments.

Affinities and Urban Semiotics:

1. **Semantic Relationships**: The proximity of words like â??libraryâ?• and â??schoolâ?• in semantic space, as determined by NLP models, does indeed reflect a kind of social or functional affinity. This could be indicative of common co-location or shared roles in urban planning and public discourse.
2. **Urban Clustering and Functionality**: Your method of calculating affinity scores for various urban elements and determining the â??cluster centerâ?? based on these scores is innovative. It provides a data-driven way to understand urban space as a reflection of social and functional networks.
3. **Broader Urban Implications**:
    - **Urban Planning and Design**: This method could inform planners about the types of amenities and infrastructures that are commonly associated in the public consciousness, guiding decisions about urban development.
    - **Cultural and Social Insights**: The clustering and affinity scores could also reveal cultural associations and social priorities within urban contexts, as reflected in the vast corpus of digital texts.
4. **Limitations and Considerations**:
    - **Contextual Relevance**: The semantic relationships derived from a general corpus like Facebookâ??s may not always align perfectly with specific urban contexts. Regional variations and local nuances might require more localized training data.
    - **Multidimensionality of Space**: Urban spaces are multidimensional, influenced by a myriad of factors beyond linguistic associations (e.g., economic, historical, geographical). Hence, these NLP-derived insights should be integrated with other urban analysis methods for a holistic understanding.
5. **Potential Extensions**:
    - **Dynamic Analysis**: Incorporating temporal data to see how urban element affinities change over time, reflecting evolving urban landscapes and societal trends.
    - **Interactive Tools**: Developing interactive platforms where urban planners and the public can input various urban elements to see their affinity scores, fostering participatory urban design.
6. **Ethical Considerations**:

- **Bias and Representation**: Ensuring that the data and models used do not perpetuate biases or overlook marginalized perspectives in urban narratives.
- **Transparency and Interpretability**: Maintaining transparency in how these models work and are applied in urban contexts to avoid misinterpretations or misuse.

Conclusion:

Your approach exemplifies the potential of interdisciplinary research, merging NLP with urban studies. It opens up new avenues for understanding urban spaces through the lens of language and collective discourse, while also highlighting the need for careful, context-aware application of these advanced technologies.

Category

1. Artificial Intelligence

Tags

1. affinity
2. NLP
3. semantic
4. urban

Date Created
December 2, 2023
Author
rcoyne99