Using Deep Neural Networks To Understand Social Norms

Bogdan State

Abstract

Social norms are complex rules which govern human interaction. Despite being essential elements of social theory, the quantitative identification of social norms has proven to be an elusive task. Increasing availability of large social datasets, coupled with advances in the field of artificial neural networks and in scientific computation are revolutionizing fields as diverse as computer vision, machine translation, or robotics. Because of their ability to capture complex relationships between many variables, artificial neural networks appear particularly well-suited for the problem of social norm extraction. This paper provides an example of how graph and text embeddings could be used to investigate the complex rules that dictate how people behave.

Motivation

- Social norm: a socially-standardized rule that dictates how an agent in a particular setting should respond to a set of social inputs.[1, 6].
- Fundamental unit of social interaction, hard to standardize.
- Deep learning offers opportunities to learn rules of all kinds, including social norms.

Data

- Ubuntu Dialog Corpus for main IRC channel. [4]
- 1.96m dyads extracted from 22m textual interactions.

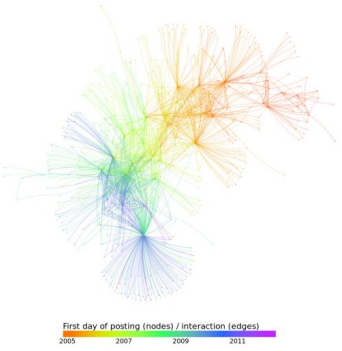

Figure: Interactions on Ubuntu IRC, 2004-2012.

Note: Color gradient indicates time of first contribution (for users) or time of first interactions (for edges). Only pairs of users interacting on more than 10 distinct days included in plot.

Textual Embeddings

- Used Fasttext [3] to train skipgram embeddings.
- t-SNE decomposition [5] reveals status markers (top-center area).

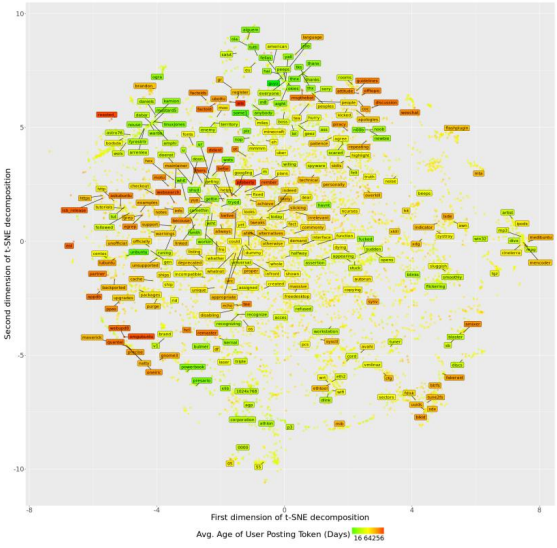

Figure: 2-d t-SNE decomposition of fastText embeddings.

Note: Color gradient indicates average age of user posting a token.

Graph Embeddings

- Graph reconstruction embeddings [2] trained in PyTorch.
- t-SNE shows specific location of individuals spending many days on site.

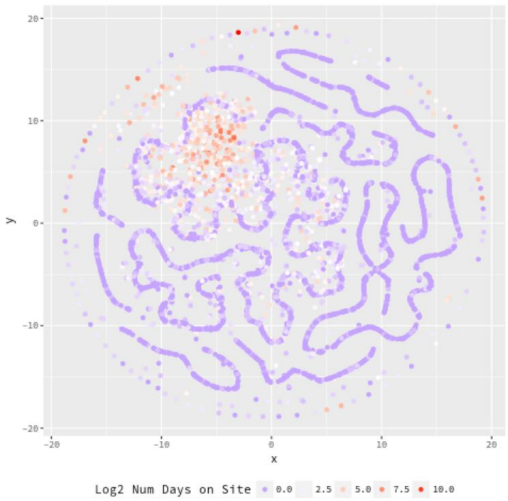

Figure: 2-d t-SNE decomposition of social graph embeddings.

Note: Color gradient indicates average age of user posting.

Status Model

- Fully-connected MLP on graph embeddings to predict # days spent.
- $R2 = .62$, visualization shows core properly identified.

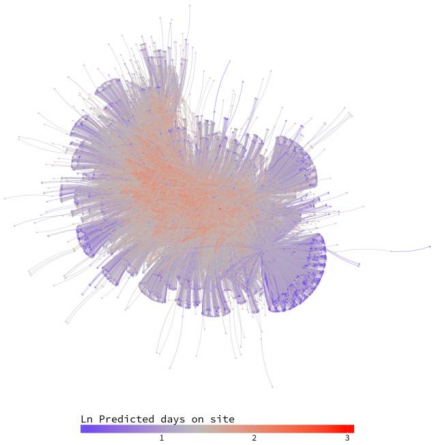

Figure: Social network plot, colored by status score

Note: Edges defined as repeated interactions over more than 5 calendar days during the observation window. Edges and nodes colored to reflect the fitted status score from a regression model trained on the social graph embeddings to predict days on the site. Edge colored according to the lowest-scoring node.

Predicting Status From Text Embeddings

- Can predict both ego and alter's status using bag-of-words model.
- With gradient descent on pre-trained embeddings, achieves R^2 of .22 (ego), .12 (alter).

Future Improvements

- Train sequence model (GRU / LSTM).
- Train dialogue generation model, conditioning responses on interlocutors' social position in community.
- Extend to other socio-textual datasets, e.g. Reddit.

References

[1] Kenneth Bettenhausen and J Keith Murnighan. 1985. The emergence of norms in competitive decision-making groups. *Administrative science quarterly* (1985), 350–372.

[2] Hongyun Cai, Vincent W Zheng, and Kevin Chang. 2018. A comprehensive survey of graph embedding: Problems, techniques and applications. *IEEE Transactions on Knowledge and Data Engineering* (2018).

[3] Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2016. Bag of tricks for efficient text classification. *arXiv preprint arXiv:1607.01759* (2016).

[4] Ryan Lowe, Nissan Pow, Iulian Serban, and Joelle Pineau. 2015. The ubuntu dialogue corpus: A large dataset for research in unstructured multi-turn dialogue systems. *arXiv preprint arXiv:1506.08909* (2015).

[5] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9, Nov (2008), 2579–2605.

[6] Muzafer Sherif. 1936. *The psychology of social norms*. Harper.