Predicting Off-target activity for CrisprigRNA gene editing

Presented by Ceasar



Motivation

Introduction

Method

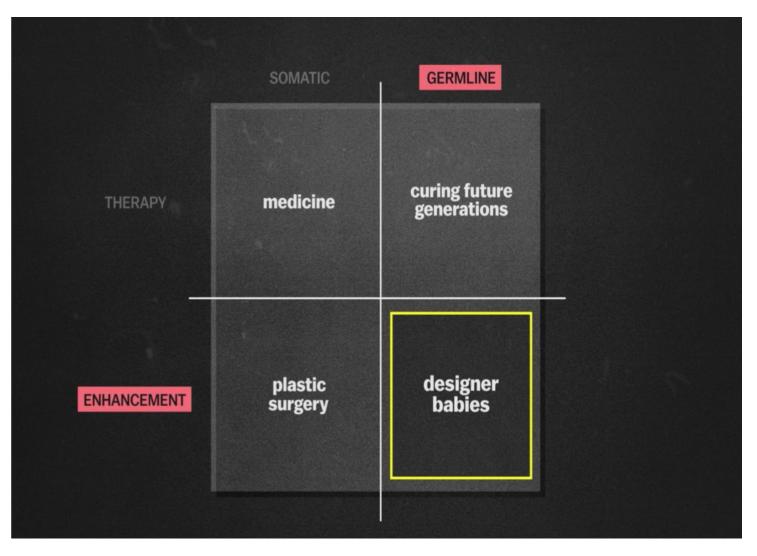
Results

Credits

Motivation

- Invention and Influence of Crispr technology
- Hope for an inclusive bioInformatics future, specially with gene editing

Gene Editing Explained



From Netflix's Explained docuseries, Designer DNA episode

Introduction

Research Question:

Given a DNA target-gRNA pair, can we predict the risk of off-target activity(CFD score) if the Crispr technology were to be deployed for gene editing using that specific pair. (where the

CFD score is a probability likelihood of off-target activity as scored by the cutting frequency determination method)

Literature Review

Microsoft's Elevation models¹:

- A score model that takes a target-gRNA pair and returns an off-activity score. (most similar)
- An aggregating model that takes a single target with many potential off targets and returns a single score from the many off target scores
 MIT's Crispor model²:
 - A model that compares scores from 8 state of the art off target evaluation models and produces the best gRNA based on the results

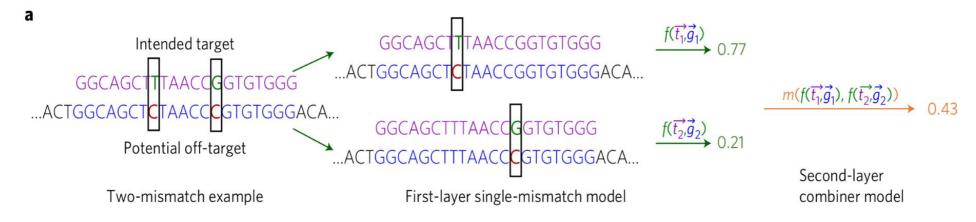


Figure extracted from reference 1(Elevation model paper).



Data:

- 26052 target gRNA pairs with corresponding
 CFD scores from the elevation paper¹
- The pairs are then encoded using a one-off encoder

	target	gRNA	cfdScore
0	GACCCCCTCCACCCCGCCTCCGG	GACCCCCACCCCCCGCCCCTGA	0.006568
1	TGGATGGAGGAATGAGGAGTTGG	AGGAAGGATGACTGAGGAGTGAG	0.018194
2	GGTGAGTGAGTGTGTGCGTGTGG	CGTGTGTGCGTGTGTGG	0.148423
3	GGTGAGTGAGTGTGTGCGTGTGG	TGTGTATGAGTGTGTGGGTGTAG	0.005546
4	GCCTCCCCAAAGCCTGGCCAGGG	GCTTCCCCAGTGCCTGGACATGG	0.063281



	0	1	2	3	4	5	6	7	8	9	•••	13	14	15	16	17	18	19	20	21	22
0	0	0	0	0	0	0	0	1	1	0		0	0	0	0	0	1	0	1	0	1
1	1	0	0	0	1	0	0	0	1	0		0	0	0	0	0	0	0	1	1	0
2	1	0	0	0	1	0	0	0	1	0		0	0	0	0	0	0	0	0	0	0
3	1	0	0	0	1	1	0	0	0	0		0	0	0	1	0	0	0	0	1	0
4	0	0	1	0	0	0	0	0	0	1		0	0	0	0	1	0	0	1	0	0

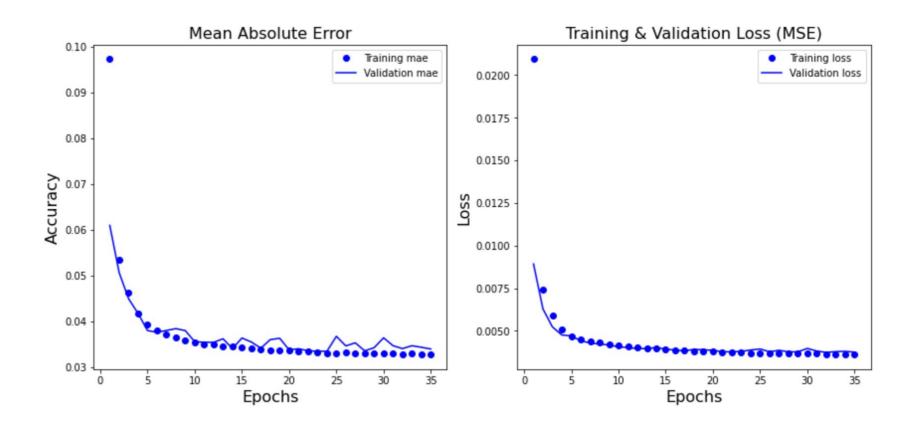
Models

- A simple 3 layer feed forward neural network (16,8,1)
- A simple 3 layer multi-Layer perceptron (16,8,1)
- A simple 3 layer recurrent neural network (16,8,1)
- All three models use the same
 hyperparameters(learning rate, # of hidden layers, epochs, batch size, and activation function)
- All models produce a single risk score for each target-gRNA pair

Evaluation

- 80/20 training/testing data split
- Mean Absolute Error
- Mean Square Error

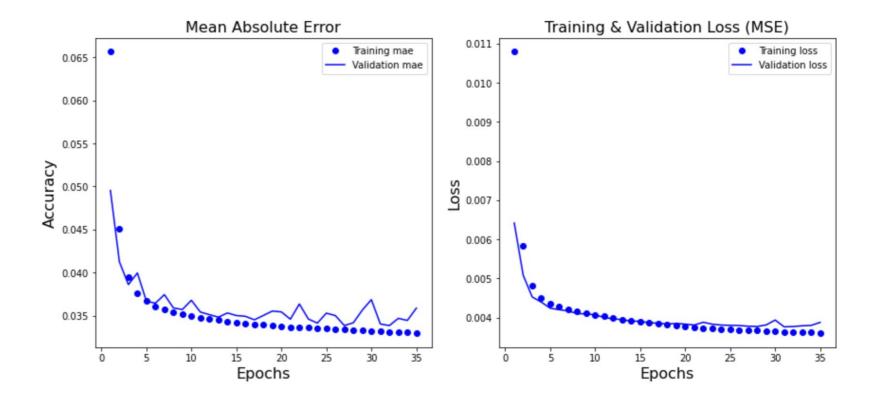
Results:Feed ForwardNetwork



Observation:

- The feed forward network produced the most stable performance on both MAE and MSE
- This is consistent with elevation findings

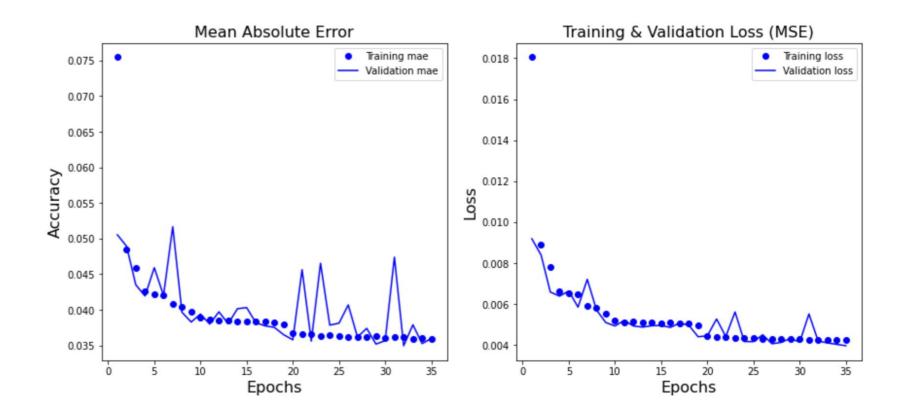
Results:Multi-layerperceptron



Observation:

- The MLP performed quite competitively with the feed forward network but not quite as smooth on the MAE metric
- The FFN is marginally better on both metrics

Results: Recurrent Neural Network



Observation:

- The RNN was the worst performer but the performance was not completely poor. Perhaps with different parameters, the RNN could perform better on the testing data.

Conclusion

Limitations:

- The project does not use the spearman weight and therefore can not properly be compared to state-of-the-art models in this area
- The project uses the CFD score as true label. This caps the level of accuracy to that achieved by the CFD method which is not the current state-of-the-art
- The project uses fixed hyper-parameters for all three models which could be limiting for certain models

Possible Extensions:

- Compare models to current state-of-the-art using the spearman weight

Credits

References

- Jennifer Listgarten*, Michael Weinstein*, Benjamin P. Kleinstiver, Alexander A. Sousa, J. Keith Joung, Jake Crawford, Kevin Gao, Luong Hoang, Melih Elibol, John G. Doench*, Nicolo Fusi*. Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs. Nature Biomedical Engineering Jan 2018, doi:10.1038/s41551-017-0178-6. (*equal contributions, corresponding author)
- Jean-Paul Concordet, Maximilian Haeussler, CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens, *Nucleic Acids Research*, Volume 46, Issue W1, 2 July 2018, Pages W242–W245, https://doi.org/10.1093/nar/gky354
- Joss Fong, Claire Gordon. Netflix Explained: Designer DNA. Netflix Season 1 Episode 2, 23 May 2018. Vox Media. https://www.imdb.com/title/tt8453098/

