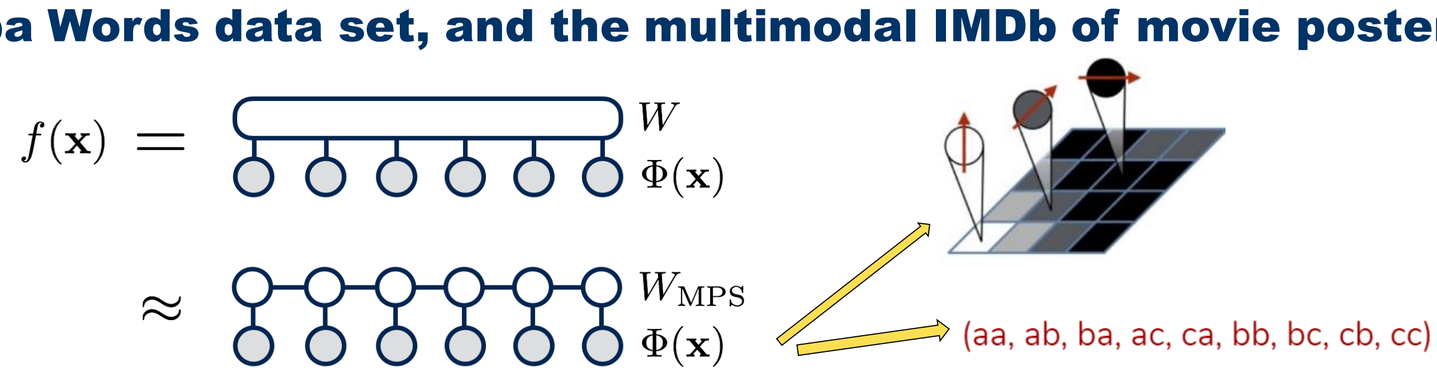
Tensor Networks for Visual and Textual Classification

Sneyder Gantiva, Diego Useche
Director: Prof. Fabio Gonzalez.
Masters in Systems and Computing Engineering,
Universidad Nacional de Colombia.

OVERVIEW

This project consists on applying a recent physics-inspired machine learning method called Tensor Networks for visual and textual classification. We used three data sets; the MNIST data set, an Eng/Spa Words data set, and the multimodal IMDb of movie posters and plots.



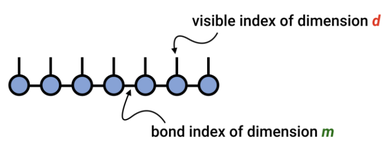
Link Presentación Youtube

MPS Learning Method

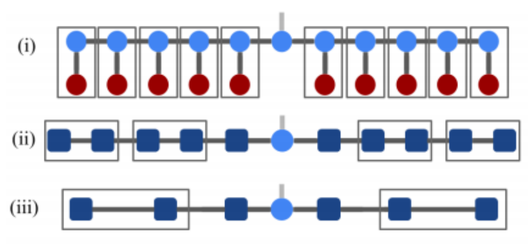
Tensor network method

$$f(\mathbf{x})^l = \sum_{i_1,i_2,\cdots,i_N} W^l_{i_1,i_2,\cdots,i_N}\phi_{i_1,i_2,\cdots,i_N}(\mathbf{x})$$

$$\phi_{i_1,i_2,\cdots,i_N}(\mathbf{x}) = \phi_{i_1}(x_1)\otimes\phi_{i_2}(x_2)\cdots\otimes\phi_{i_N}(x_N)$$



Efthymiou Contraction



Quantum Feature Maps

Visual Binary Pixel

$$|\psi(x)\rangle = \begin{cases} |0\rangle & \text{if} \quad \frac{x}{255} < 0.5 \\ |1\rangle & \text{if} \quad \frac{x}{255} > 0.5 \end{cases}$$

Visual RGB Fusion

$$X = R * 0.21 + G * 0.72 + B * 0.07$$

Visual Binary Feature

$$|\psi(x)\rangle = \begin{cases} |0\rangle & \text{if} \quad \sigma(x) < 0.5 \\ |1\rangle & \text{if} \quad \sigma(x) > 0.5 \end{cases}$$

Textual n-gram

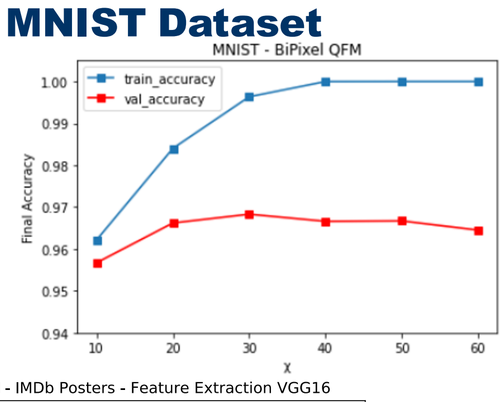
$$|\psi(x)\rangle = \begin{cases} |1\rangle & \text{if} \quad \text{n-gram in text} \\ |0\rangle & \text{otherwise} \end{cases}$$

Visual Exps and Results

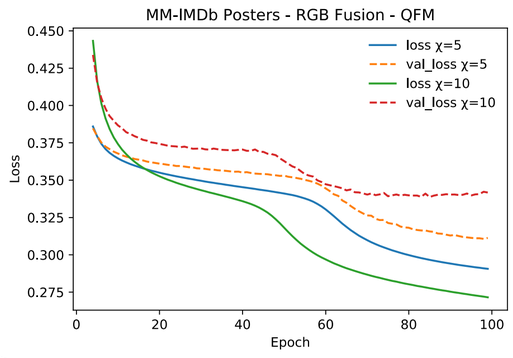
IMDb Posters

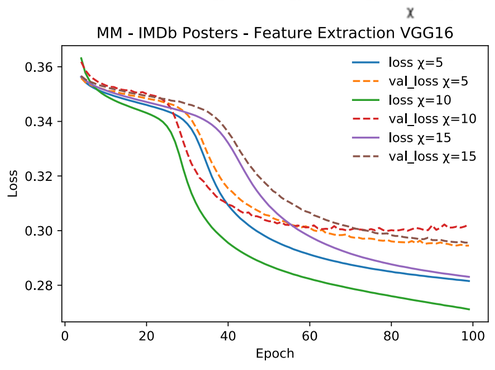
F-Score on Test: MM-IMDb Posters				
Model	**macro**	**micro**	**weighted**	**samples**
MobileNetV2 Fine Tuning	0.0847	0.3385	0.2363	0.3453
RGB Fussion	0.0507	0.2144	0.1629	0.1862
Feature Extraction VGG16	0.1106	0.2547	0.2424	0.2275

MNIST Dataset



IMDb Posters – Learning Process





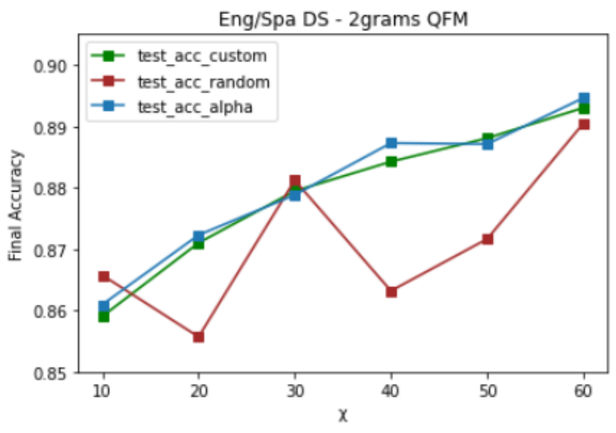
Conclusions

- MNIST: 96.83% accuracy on test data, with a bond dimension of 30.
- IMDb posters: Best f-scores obtained with the Feature Extraction VGG and MPS, still lower scores compared with usual transfer learning methods.
- Eng/Spa Words: The accuracy of test data depends on the bond, but not on the order of the 2-grams. The MPS outperforms SVM with histogram intersection kernel.
- IMDb plots: Acceptable f-scores with MPS and Tfidf QFM, still lower than a usual method based on LSTM and w2v.
- Future work: Explore the MPS implementation and feature map of Stoudenmire [1].

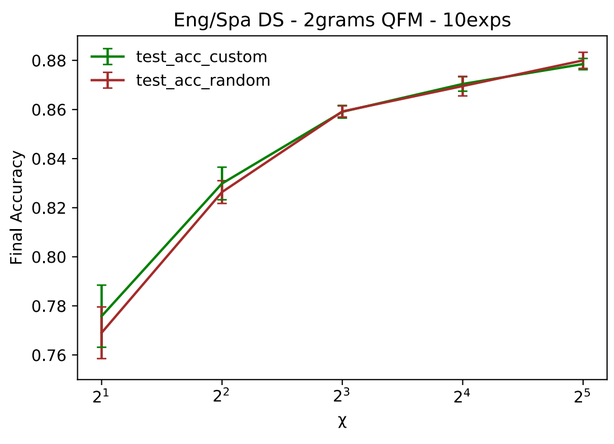
Textual Exps and Results

Eng/Spa Words Data Set

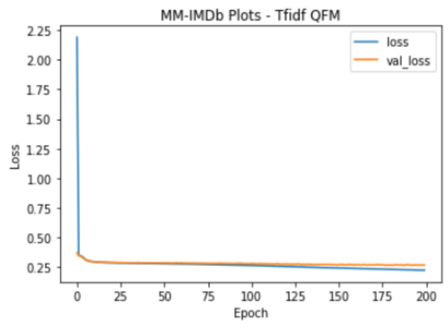
Changing 2-grams orders in MPS



Model	Accuracy on Test
MPS ($\chi = 60$)	89.45%
SVM HistoInter K	87.82%



IMDb Plots



F-Score on Test: MM-IMDb Plots				
Model	macro	micro	weighted	samples
MPS Tfidf	0.170	0.387	0.338	0.344
LSTM w2v	0.266	0.478	0.434	0.471

Main References

[1] E. M. Stoudenmire and D. J. Schwab, “Supervised Learning withQuantum-Inspired Tensor Networks,” 5 2016.
[2] S. Efthymiou, J. Hidary, and S. Leichenauer, “TensorNetwork forMachine Learning,” 6 2019.
[3] J. Arevalo, Solorio, M. Montes-Y-Gómez, and F. A. González, “Work-shop track-ICLR 2017 GATED MULTIMODAL UNITS FOR INFORMATION FU-SION,” tech. rep.

Acknowledgements: We would like to thank the seminar groups Physics and Machine Learning and the MindLab group, both seminars oriented by Prof. Fabio Gonzalez.