

Large-Scale Assessment of Deep Relational Machines

(Presentation at the 28th ILP 2018)

Tirtharaj Dash¹, Ashwin Srinivasan¹, Lovekesh Vig²,
Oghenejokpeme I. Orhobor³, Ross D. King³

¹BITS Pilani, Goa Campus, Goa

²TCS Innovation Labs, New Delhi

³University of Manchester, Manchester

The Goals

Main:

- (1) To establish a good, simple baseline for comparison of neuro-symbolic models using large amount of relational data and background knowledge
- (2) To compare the performance of the baseline against state-of-the-art.

Additional:

- (1) Use of symbolic domain-knowledge by the simple baseline
- (2) The limitations of the baseline

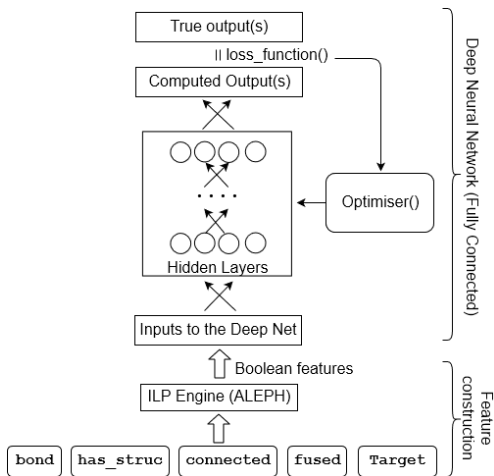
This Talk

- ▶ **What** is a “Deep Relational Machine (DRM)”?
- ▶ **How** is it different from a “Deep Neural Network”?
- ▶ **Why** is it useful to study DRMs?

Deep Relational Machine (DRM)

The term DRM is introduced in Lodhi¹.

- ▶ ILP (Data, BK) + Deep Network
- ▶ Input: First-order boolean functions (E.g. Function F_1 is TRUE if the instance x is a molecule containing 7-membered ring connected to a lactone ring)

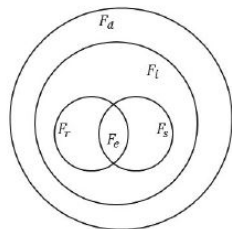


¹Lodhi, H.: Deep relational machines. In Proc: ICONIP 2013.

Feature classes

A comprehensive study was conducted in Saha et al.². The Venn diagram shows the relationships among various classes of features.

- ▶ F_d : unrestricted definite clauses, F_i : body contains only one independent component, F_s : body contains only one sink literal.
- ▶ Every feature in F_d, F_i, F_r, F_e can be constructed from simple features F_s .
- ▶ Every feature in F_d can be constructed from independent features F_i .
- ▶ Our present work uses features from F_d .



²Saha, A., Srinivasan, A., & Ramakrishnan, G.: What Kinds of Relational Features Are Useful for Statistical Learning?. In Proc: ILP 2012

Inputs to the DRM (Propositionalisation)

The following procedure is used in Vig et al.³

Repeat:	Comments
1. Randomly draw an example e	Ln 1. with replacement
2. Construct the bottom clause for e (most specific clause)	Ln 2. use a depth-limited mode language
3. Draw a clause (from a feature class) that subsumes the bottom clause	Ln 3. Max. literals in body: 3 Ln 4. to avoid redundancy of features
4. Check subsumption equivalence with already-selected features	Ln 5. add it to the feature set
5. Construct the feature	

³Vig, L., Srinivasan, A., Bain, M., & Verma, A. (2017, September). An Investigation into the Role of Domain-Knowledge on the Use of Embeddings. In Proc. ILP 2017.

Deep Network

- ▶ Dense Multi-layered Perceptron (MLP) with various depths
- ▶ Minimum 1 and maximum 4 hidden layers of neurons
- ▶ Number of hidden neurons $\in \{5, 10\}$
- ▶ Number of deep networks evaluated: $(2+4+8+16=30)$

Problems (Data)

So far, DRMs have been tested on very small amounts of data (7 datasets, few 1000s of instances).

In this work, DRM is evaluated on: (1) 73 anti-cancer datasets – classification, (2) 50 QSAR datasets – regression

Task	Size		Features		Target Distributions
	Datasets	Examples	<i>AB</i>	<i>ABFR</i>	
Classification	73	≈ 220,000	≈ 3000	≈ 4000	0.4 – 0.9 (% positives)
Regression	50	≈ 18,000	≈ 900	≈ 2200	1.5 – 11.0 (predicted values)

AB: Just the bond description of a molecule (does not use domain knowledge)

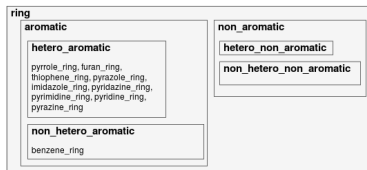
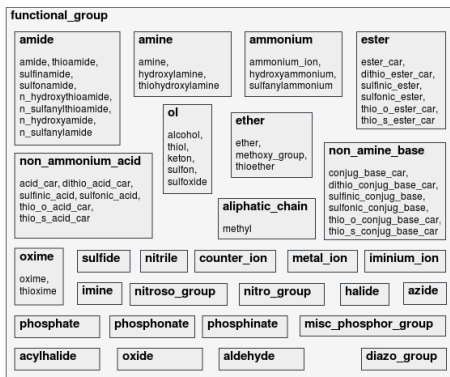
ABFR: bond description along with functional groups and rings (uses domain knowledge)

Classification: National Cancer Institute (NCI) (www.cancer.gov)

Regression: ChEMBL database (www.ebi.ac.uk/chembl)

Background Knowledge

We use the same background knowledge as DMax⁴ with minor modification for tractable computation.



⁴ <https://dtai.cs.kuleuven.be/software/dmax/>

Results I (Comparable to State-of-the-art)

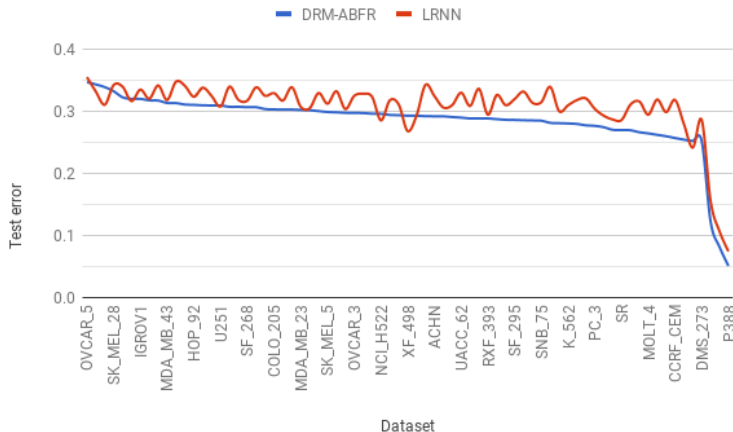


Figure: Classification (all 73) (Comparing with LRNN⁵)

⁵Sourek, G., Aschenbrenner, V., Zelezny, F., & Kuzelka, O. (2015). Lifted relational neural networks. arXiv preprint arXiv:1508.05128.

Results II (Comparable to State-of-the-art)

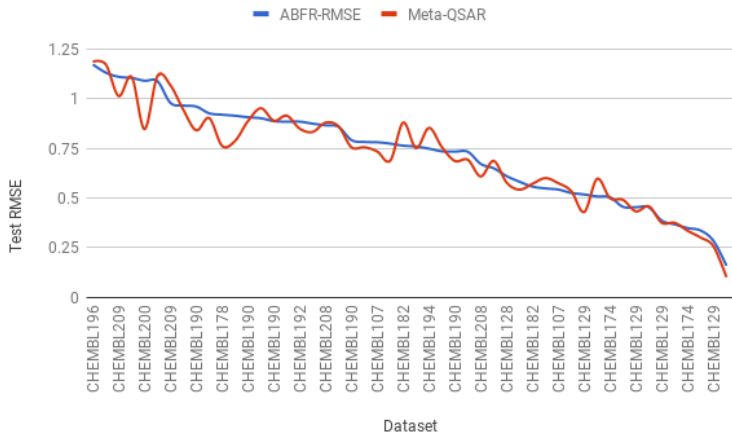


Figure: Regression (all 50) (Comparing with Meta-QSAR⁶)

⁶Olier, I. et al.: Meta-QSAR: a large-scale application of meta-learning to drug design and discovery. Machine Learning, 2018

Results III (Gets better with domain knowledge)

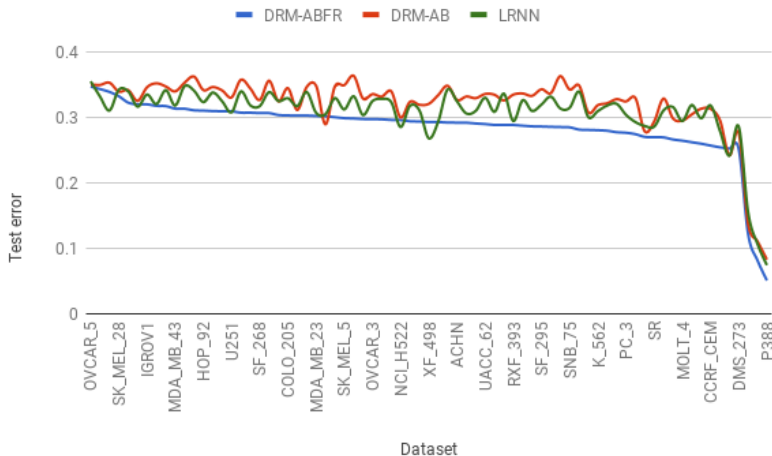


Figure: DRM performance when high-level background knowledge is not used

Results IV (Increasing features helps)

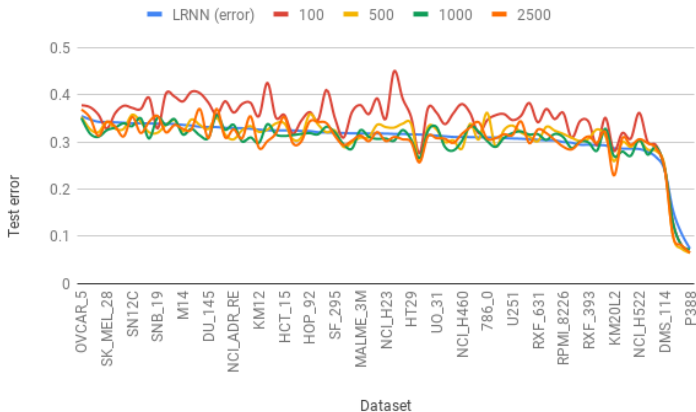


Figure: Effect of increasing features (comparing against state-of-the-art)

Deep Net building time (approx): 50: 15s, 100: 20s, 250: 25s, 500: 35s, 1000: 45s, 2500: 60s, 5000: 90s

Results V (But, features need to be expressive)

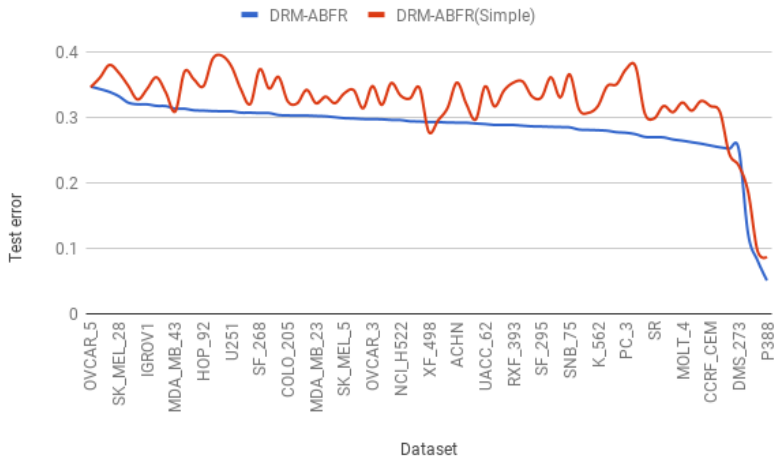


Figure: Unrestricted features versus simple features

Regression can be improved

- ▶ DRMs are better in 50% and worse in 50%
- ▶ The features used in the Meta-QSAR study are pharmacophore features (FCFP4 fingerprint representation): contains more detailed information of structural and chemical properties of molecules
- ▶ Enriching our feature set by augmenting it with FCFP4 features

Results VI (Feature enrichment helps regression)

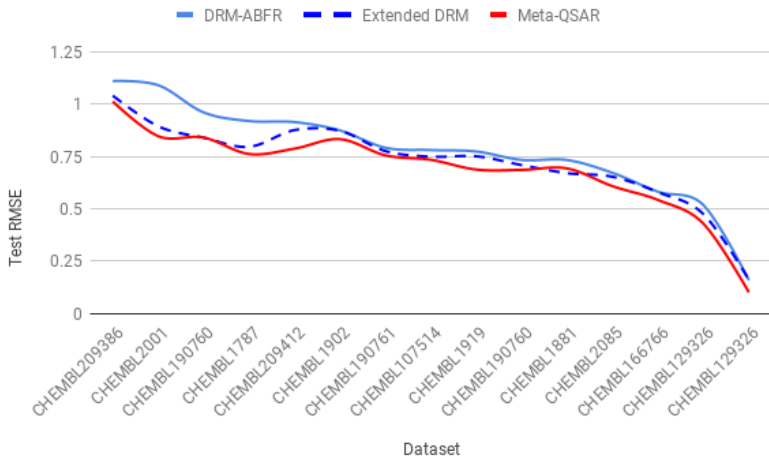


Figure: Extended DRM: Effect of feature enrichment in DRM for bottom 30% of datasets

Summary of results

- (1) The idea of propositionalisation have been around for a long time starting with LINUS. It is a simple way to introduce background knowledge into feature based learning.
- (2) Results of deep neural networks and propositionalisation (which we call 'DRM') are surprisingly good even with randomly selected features
- (3) The datasets and results here provide a good baseline to compare neuro-symbolic models on relational data.
- (4) DRMs may be more scalable than more elaborate methods like ∂ -ILP⁷.
 - ▶ But, see limitations (next)

⁷Evans, R., & Grefenstette, E. (2018). Learning explanatory rules from noisy data. *Journal of Artificial Intelligence Research*, 61, 1-64.

Limitations

- ▶ Performance of DRM depends on the expressive power of features used as input (This work: unrestricted class of definite clauses)
- ▶ Intractable to provide all features with sufficient expressive power
- ▶ Deep Network can not do relational join. Example: A neuron taking two features:

$F1 : \forall x(East(x) \leftarrow \exists y(HasCar(x, y), Short(y)))$, and

$F2 : \forall x(East(x) \leftarrow \exists y(HasCar(x, y), Closed(y)))$.

can not produce

$F : \forall x(East(x) \leftarrow \exists y(HasCar(x, y), Short(y), Closed(y)))$

but, will produce an approximation to

$F' : \forall x(East(x) \leftarrow$

$\exists y, z(HasCar(x, y), HasCar(x, z), Short(y), Closed(z)))$

i.e.

$F'' : \phi(w_1 F1 + w_2 F2 + w_0)$

Acknowledgments

- ▶ DST-SERB grant EMR/2016/002766
- ▶ EurAI for Travel Grant to participate in ACAI-2018
- ▶ Department of CSIS, BITS Pilani, Goa for financial assistance
- ▶ Researchers at the DTAI, University of Leuven
- ▶ Ing. Gustav Sourek (Czech Technical University, Prague)
- ▶ Dr. Ivan Olier Caparroso (Liverpool John Moores University, UK)

Thank you.