TraininG towards a society of data-saVvy inforMation prOfessionals to enable open leadership INnovation

MOVING

Multi-Modal Adversarial Autoencoders for Recommendations of Citations and Subject Labels

Lukas Galke Florian Mai Iacopo Vagliano Ansgar Scherp

Kiel University, Germany

UMAP '18: 26th Conference on User Modeling, Adaptation and Personalization, July 8–11, 2018, Singapore

www.moving-project.eu

"Avoid using GANs, if you care for your mental health"

- Alfredo Canziani

Motivation

- Adversarial regularization improves autoencoders on images (Makhzani et al. 2015)

Motivation

- Adversarial regularization improves autoencoders on images (Makhzani et al. 2015)
- Autoencoders enable flexible treatment of multi-modal input (Barbieri et al. 2017)

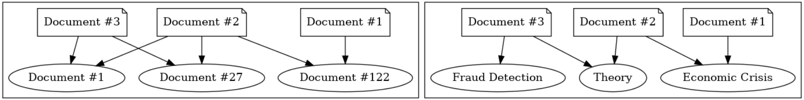
Motivation

- Adversarial regularization improves autoencoders on images (Makhzani et al. 2015)
- Autoencoders enable flexible treatment of multi-modal input (Barbieri et al. 2017)

Research Questions

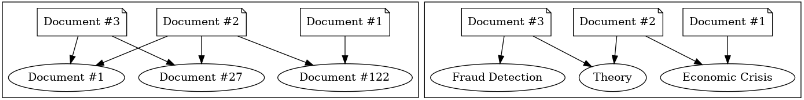
1. Does adversarial regularization improve autoencoders for recommendation?
2. To what extent do preferable input modalities depend on task?
3. What is the effect of sparsity?

Two Different Tasks



Recommendations for citations (left) and subject labels (right)

Two Different Tasks



Recommendations for citations (left) and subject labels (right)

- Two recommendation tasks on scientific documents
- Items are either citations or subject labels
- Assumption: test documents unknown

Example

- A researcher is writing a new paper

Example

- A researcher is writing a new paper
- the draft cites already 10 other publications (**item set**)

Example

- A researcher is writing a new paper
- the draft cites already 10 other publications (**item set**)
- "Am I missing any relevant related work?"

Example

- A researcher is writing a new paper
- the draft cites already 10 other publications (**item set**)
- "Am I missing any relevant related work?"

$\rightarrow$ Recommend citation candidates

Example

- A researcher is writing a new paper
- the draft cites already 10 other publications (**item set**)
- "Am I missing any relevant related work?"

$\rightarrow$ Recommend citation candidates

Important: the draft is unseen by the system (New User)

Exploit Multiple Modalities

Use only citations of draft?

Exploit Multiple Modalities

Use only citations of draft? Hmm, there is more...

Exploit Multiple Modalities

Use only citations of draft? Hmm, there is more...

Use more data of the draft?

Exploit Multiple Modalities

Use only citations of draft? Hmm, there is more...

Use more data of the draft? Yes.

Exploit Multiple Modalities

Use only citations of draft? Hmm, there is more...

Use more data of the draft? Yes.

$\rightarrow$ Start with `title`, but it could be more (**condition**)

Related Work

- Document-level citation recommendation: collaborative filtering (McNee et al. 2002), SVD (Caragea et al. 2013)

Related Work

MOVING

- Document-level citation recommendation: collaborative filtering (McNee et al. 2002), SVD (Caragea et al. 2013)
- Subject Labeling: MLP for Multi-label classification (Galke et al. 2017), but professionals use predictions only as recommendations

Approach

M🌍VING

Multi-Modal Adversarial Autoencoder

- Train autoencoder on **item sets**

Multi-Modal Adversarial Autoencoder

- Train autoencoder on **item sets**
- Supply **condition** to the decoder (multi-modal)

Approach

Multi-Modal Adversarial Autoencoder

- Train autoencoder on **item sets**
- Supply **condition** to the decoder (multi-modal)
- Jointly train encoder to produce code indistinguishable from a sample of indepentend Gaussians (adversarial)

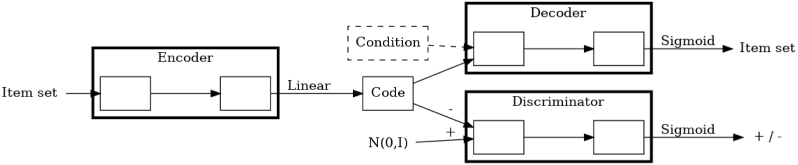
Rationale

- Recommendation tasks are highly sparse
- *Good* representations (Bengio, Courville, and Vincent 2012) might be helpful, e.g. smoothness $a \approx b \rightarrow f(a) \approx f(b)$
- Enforce smoothness on the code (adversarial regularization)

Rationale

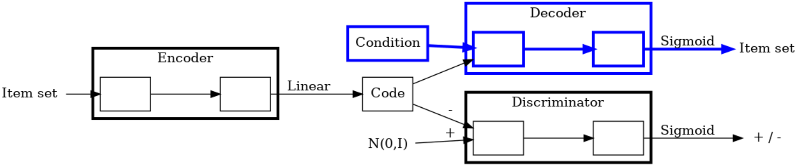
- Recommendation tasks are highly sparse
- *Good* representations (Bengio, Courville, and Vincent 2012) might be helpful, e.g. smoothness $a \approx b \rightarrow f(a) \approx f(b)$
- Enforce smoothness on the code (adversarial regularization)
- Leads to more generalizable reconstruction? $\rightarrow$ RQ 1

Adversarial Autoencoders



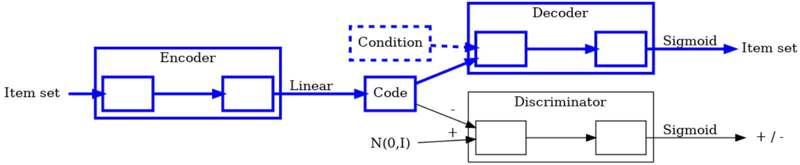
Model Overview

Multi-Layer Perceptron

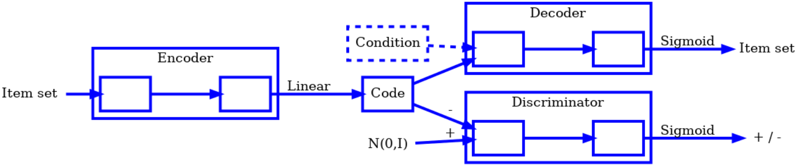


Parts used for the Multi-Layer Perceptron (MLP)

Undercomplete Autoencoders



Parts used for the Autoencoder (AE)



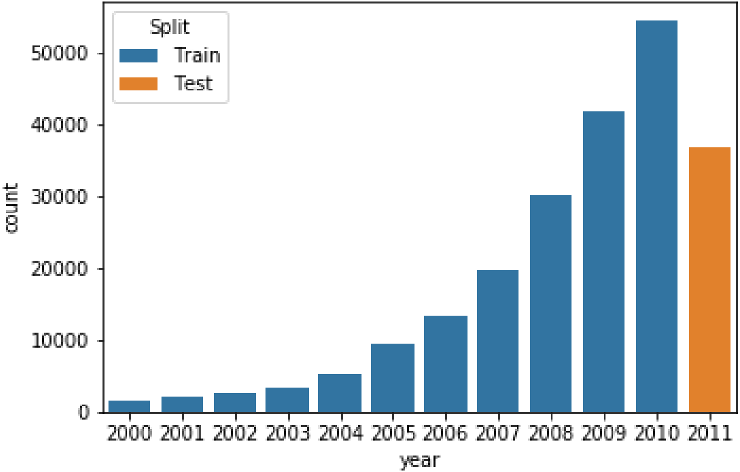
Parts used for the Adversarial Autoencoder (AAE)

Experimental Setup

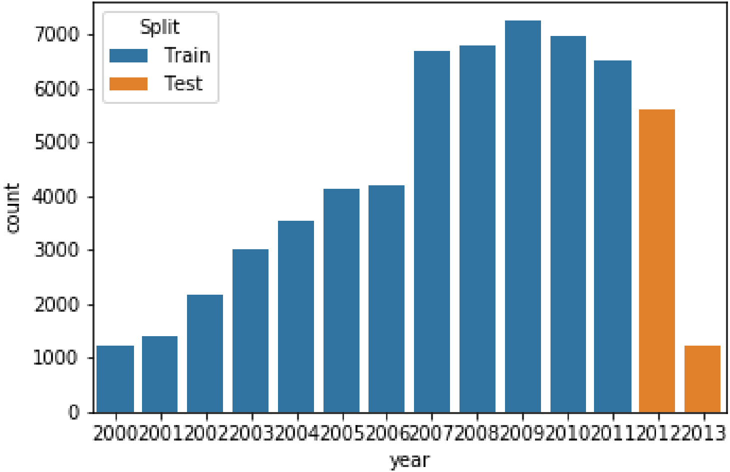
Close to real-world evaluation:

- Train and test split on the time axis $\rightarrow$ **disjoint**
  "only published resources are citable"
- Number of considered items is crucial (Beel et al. 2016) $\rightarrow$ pruning thresholds as controlled variable
- Title as additional input (as condition) vs. only item sets
- Datasets: PubMed for citations, Econ62k for subject labels
- Evaluate mean reciprocal rank (MRR) of **one** dropped item among the predictions.
- Re-run three times $\rightarrow$ 408 experiments.

Time Split: PubMed



Time Split: Economics



Task Definition

Task: Given a partial set of items $x \setminus \{i^*\}$, find the missing item i^*.

x row of binary ratings over documents $\times$ items.

c condition: documents' title

y predicted probabilities for items: $p(y|x, c)$

Goal: Missing item on high rank $i^* = \arg\max y$

Method Summary

MOVING

Only item sets

IC Item Co-occurence (McNee et al. 2002)

Only titles

MLP $y = \text{MLP}_{\text{dec}}(c)$

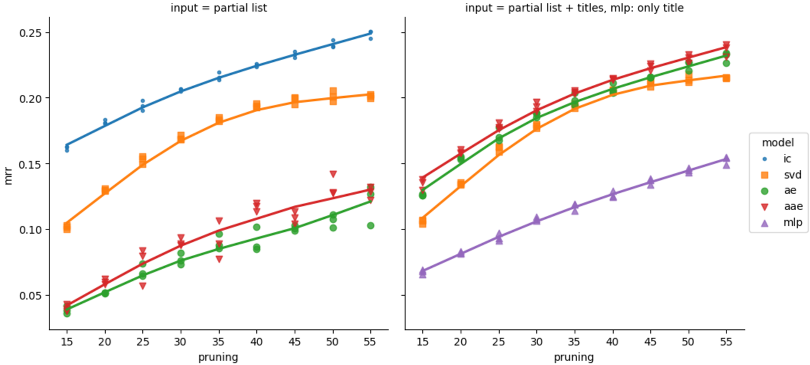
Multi-Modal

SVD Singular value decomposition (Caragea et al. 2013)

AE $y = \text{MLP}_{\text{dec}}(\text{MLP}_{\text{enc}}(x)[, c])$

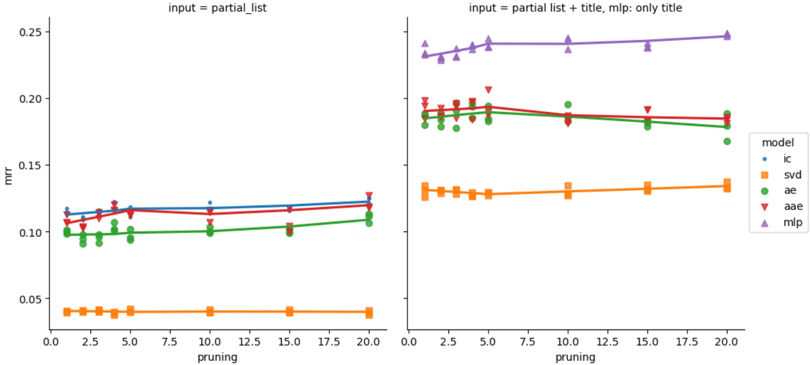
AAE $y = \text{MLP}_{\text{dec}}(\text{MLP}_{\text{enc}}(x)[, c])$. Encoder MLP_{enc} jointly optimized to fool discriminator MLP_{disc}.

Citation Recommendation



PubMed: MRR of methods by pruning threshold on minimum item count

Subject Label Recommendation



Economics: MRR of methods by pruning threshold on minimum item count

Results

- AAE yields consistently higher scores than AE
- Multiple input modalities improve both AE and AAE
- **Surprising**: MLP using only title data yields higher scores than AAE on subject labels but lower scores than AAE on citations

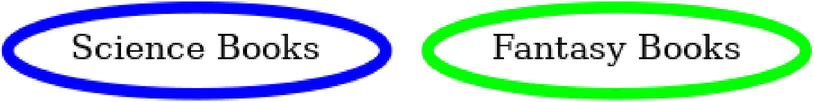
What does it mean if two items co-occur in a document?

- Citation co-occurrence $\approx$ relatedness (Small 1973)
  $\rightarrow$ partial item set helpful
- Subject label co-occurrence $\approx$ diversity (guidelines)
  $\rightarrow$ partial item set **not** helpful, rather distracting

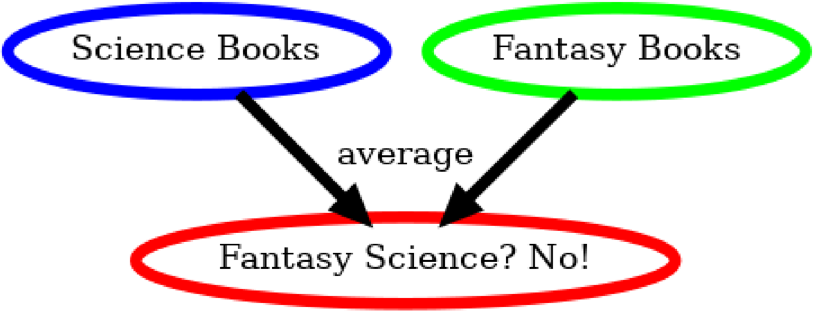
What does it mean if two items co-occur in a document?

- Citation co-occurrence $\approx$ relatedness (Small 1973)
  $\rightarrow$ partial item set helpful
- Subject label co-occurrence $\approx$ diversity (guidelines)
  $\rightarrow$ partial item set **not** helpful, rather distracting

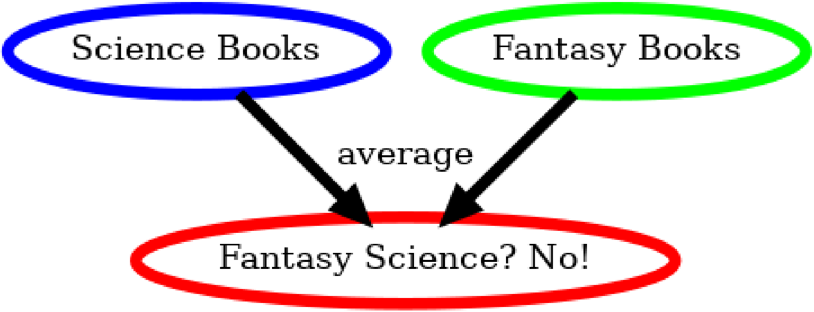
Our two tasks are prototypical for each case. What is inbetween?


Science Books
Fantasy Books

Some Intuition

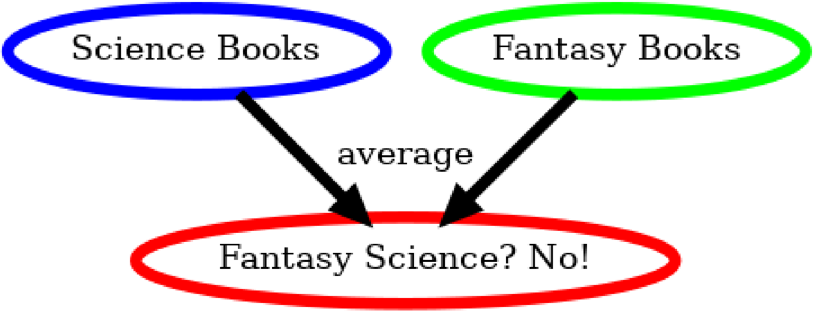


Some Intuition

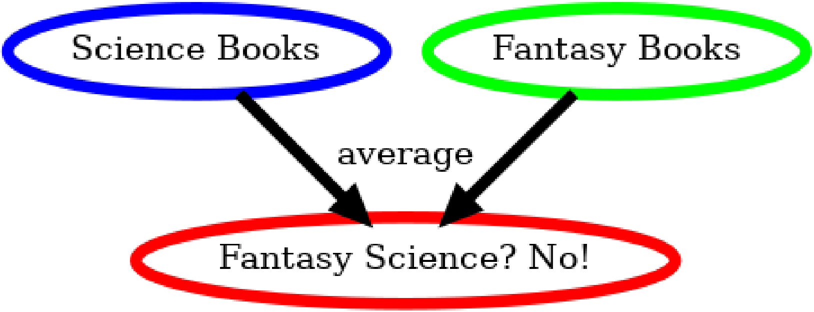


1. Manifold Learning

Some Intuition

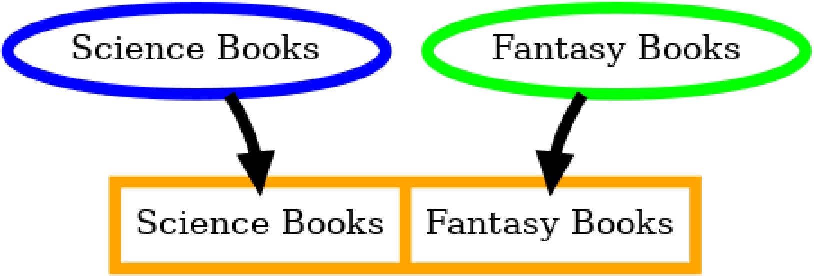


1. Manifold Learning
2. Linear interpolations on the code



1. Manifold Learning
2. Linear interpolations on the code
3. Mixing well between modes

Some Intuition



1. Manifold Learning
2. Linear interpolations on the code
3. Mixing well between modes

Conclusion

Adversarial Autoencoders:

- consistent improvement over undercomplete autoencoders
- capable of exploiting different input modalities
- robust to sparsity as other approaches

Take-home

Consider the semantics of item co-occurrence for the choice of an appropriate model.

Code available at `github.com/lgalke/aae-recommender`
Contact me via `http://lpag.de`

References I

- Barbieri, Julio et al. (2017). "Autoencoders and recommender systems: COFILS approach". In: *Expert Syst. Appl.* 89, pp. 81–90.
- Beel, Joeran et al. (2016). "paper recommender systems: a literature survey". In: *International Journal on Digital Libraries* 17.4, pp. 305–338.
- Bengio, Yoshua, Aaron C. Courville, and Pascal Vincent (2012). "Unsupervised Feature Learning and Deep Learning: A Review and New Perspectives". In: *CoRR* abs/1206.5538.
- Caragea, Cornelia et al. (2013). "Can't see the forest for the trees?: a citation recommendation system". In: *JCDL*. ACM, pp. 111–114.
- Galke, Lukas et al. (2017). "Using Titles vs. Full-text as Source for Automated Semantic Document Annotation". In: *K-CAP*. ACM, 20:1–20:4.

References II

- Makhzani, Alireza et al. (2015). "Adversarial Autoencoders". In: *CoRR* abs/1511.05644.
- McNee, Sean M. et al. (2002). "On the recommending of citations for research papers". In: *CSCW*. ACM, pp. 116–125.
- Small, Henry (1973). "Co-citation in the scientific literature: A new measure of the relationship between two documents". In: *JASIS* 24.4, pp. 265–269.

Hyperparameters

Gridsearch on PubMed$_{\geq 50}$:

- Hidden layer sizes between 50 and 1,000: **100**
- Code sizes between 10 and 500: **50**
- Drop probabilities between .1 and .5: **.2**
- Stochastic Gradient Descent or Adam: **Adam**
- Initial learning rates between 0.01 and 0.00005: **0.001**
- Activation functions Tanh, ReLU, SELU: **ReLU**
- Prior distribution: Gaussian, Bernoulli, Multinomial: **Gaussian**
- SVD truncated at first **1,000** singular values

Dataset: PubMed

pruning	cited documents	citations	documents	density
15	35,664	1,173,568	136,911	0.000240
20	20,270	878,359	121,374	0.000357
25	12,881	692,037	105,170	0.000511
30	8,906	568,563	96,980	0.000658
35	6,469	478,693	87,498	0.000846
40	4,939	413,746	79,830	0.001049
45	3,904	363,870	73,200	0.001273
50	3,185	324,693	67,703	0.001506
55	2,643	292,791	62,647	0.001768

Dataset: Economics

pruning	labels	assigned labels	documents	density
1	4,568	323,670	61,104	0.001160
2	4,103	323,060	61,090	0.001289
3	3,760	322,199	61,060	0.001403
4	3,497	321,213	61,039	0.001505
5	3,259	320,048	60,983	0.001610
10	2,597	314,738	60,778	0.001994
15	2,192	309,101	60,524	0.002330
20	1,924	303,693	60,272	0.002619