Neural Outfit Recommendation

DAPA Workshop @ WSDM 2019

Maarten de Rijke

February 15, 2019

University of Amsterdam

derijke@uva.nl

Based on joint work with Jun Ma, Pengjie Ren, Yujie Lin, Zhaochun Ren, and Zhumin Chen

- Background
- Outfit recommendation
- Fashion recommendation machine
- Some results
- Conclusion

Neural IR

Big uptake and injection of energy in the field

- Learning to match
- Learning to rank
- Content understanding – text, image, video, ...
- Behavior understanding
- ...

The need to take stock, repeatedly

Quickly building up a rich body of knowledge

- Li and Xu (2013) – Semantic matching in search
- Onal et al. (2018) – Neural information retrieval: At the end of the early years
- Mitra and Craswell (2019) – An introduction to neural information retrieval
- Li et al. (20XX) – ...

Rough edges

Lin (2018) – The Neural Hype and Comparisons Against Weak Baselines

- Everyone is trying to **win**
- “demonstrating that a new method beats previous methods on a given task or benchmark”
- Often, our baselines are weak

Rough edges

How to improve ourselves

- **Compare apples to apples**
- **Work on insights – reasons for success, reasons for failure**
- **Use reference baselines**

Rough edges

How to improve ourselves

- **Compare apples to apples**
- **Work on insights – reasons for success, reasons for failure**
- **Use reference baselines**
- **Share everything**
- **Use reference implementations**
- **Engage with product owners for additional eyes and checks**
- **Win in different ways – task, constraints, metrics, . . .**

Twin-Win Model: A human-centered approach to research success

Ben Shneiderman[a,1]

[a]Department of Computer Science & Human-Computer Interaction Lab, University of Maryland, College Park, MD 20814

Edited by William Rouse, Stevens Institute of Technology, Hoboken, NJ, and accepted by Editorial Board Member Pablo G. Debenedetti May 4, 2018 (received for review February 15, 2018)

A 70-year-old simmering debate has erupted into vigorous battles over the most effective ways to conduct research. Well-established beliefs are being forcefully challenged by advocates of new research models. While there can be no final resolution to this battle, this paper offers the Twin-Win Model to guide teams of researchers, academic leaders, business managers, and government funding policymakers. The Twin-Win Model favors a problem-oriented approach to research, which encourages formation of teams to pursue the dual goals of breakthrough theories in published papers and validated solutions that are ready for widespread dissemination. The raised expectations of simultaneously pursuing foundational discoveries and powerful innovations are a step beyond traditional approaches that advocate basic research first. Evidence from citation analysis and researcher interviews suggests that simultaneous pursuit of both goals raises the chance of twin-win success.

research model | basic research | applied research | Twin-Win Model | human-centered research

The usually quiet world of academic research is being awakened by explosive battles over how to do research (1–4). The traditional linear model of research argued for curiosity-driven basic research in laboratories to acquire new knowledge. This may have been productive in the knowledge-poor early days of discovery, but now, in our knowledge-rich, information-overloaded world, new models are needed. Since collecting new knowledge has become so easy, researchers need to consider which forms of new knowledge would be most beneficial. Collecting the length of every rat's tail or the number of characters in every tweet would add to the store of knowledge. However, it seems clear that collecting the location of every rat to understand the spread of disease or the time stamp of every tweet to understand sleep patterns in different cities would be more helpful in raising further questions and useful in recommending constructive actions.

In short, some knowledge is likely to be more useful than others, because the knowledge relates to meaningful problems and may suggest constructive actions. Knowledge is tied to meaningful problems by way of a causal theory that permits intervention so as to contribute to improvements in human life or environmental preservation. Therefore, my claim is that research can become more productive if the pursuit of new knowledge is tied to actionable insights that can lead to societal benefits and sustainable conservation.

Leading organizations have identified key challenges, such as the papers and validated solutions that are ready for widespread dissemination.

Background

The idea of bringing academic researchers in closer contact with professionals who face authentic problems has long been discussed as a way to achieve higher societal benefits. The famed American poet and philosopher Ralph Waldo Emerson spoke in 1837 about academics working more closely with farmers, business people, and government. Emerson called for academics to engage in the real world: "Action … is essential … Without it, thought can never ripen into truth." That encouragement remains valid today.

More than a century later, Vannevar Bush's (5) 1945 manifesto *Science: The Endless Frontier, a Report to the President on a Program for Postwar Scientific Research* sought to separate academic work from practical problems. He argued for a linear model, suggesting that basic research came first, which led to applied research and then commercial development. The linear model was vigorously opposed by Tom Allen (6) in the 1970s, Deborah Shapley and Rustum Roy (7) in the 1980s, and many others. An important contribution was Donald Stokes' (8) 1997 book *Pasteur's Quadrant: Basic Science and Technological Innovation*, which proposed a fresh strategy: "use-inspired basic research." His reference to Pasteur reminded readers about Pasteur's work on the problems of vintners and dairy farmers, which produced the twin-win of solutions to their problems and the germ theory of disease. Lewis Branscomb's (9) 2007 essay supported the idea that creativity and utility (basic and applied) research were happy partners. Steven Chu, Nobel Prize winner in physics and US Secretary of Energy, reinforced the need for a shift in research: "We seek solutions. We don't seek—dare I say this—just scientific papers anymore."

In the past few years, *The New ABCs of Research: Achieving Breakthrough Collaborations* (10), which advocated for "applied and basic combined (ABC)," was joined by Narayanamurti and Odumosu's (2) book on *Cycles of Invention and Discovery: Rethinking the Endless Frontier*. Dan Sarewitz (3) wrote a powerful essay on "Saving science," pushing for reform of science to increase its impact, while reducing the prevalence of results that could not be replicated. Sarewitz (3) stressed that "scientists must come out of the lab and into the real world." A similar call for emphasizing applications as the path to discoveries came

Background

Outfit recommendation

Fashion recommendation machine

Some results

Conclusion

Outfit recommendation

A different task, with a twist

Fashion recommendation – increased attention

Outfit recommendation – given a top (i.e., upper garment), recommend a list of bottoms (e.g., trousers or skirts) from a large collection that best match the top, and vice versa

- Allow users to provide some descriptions as conditions that the recommended items should accord with as much as possible

Unpacking the task

Two main challenges

- **visual understanding** – aims to extract effective visual features
- **visual matching** – aims to model a human notion of compatibility to compute a match between fashion items

Unpacking the task

Two main challenges

- **visual understanding** – aims to extract effective visual features
- **visual matching** – aims to model a human notion of compatibility to compute a match between fashion items

Typically, visual understanding and matching conducted based on recommendation loss alone

- Supervision signal is just whether two given items are matched or not and no supervision is available to directly connect the visual signals of the fashion items
- Can we come up with a sense of **esthetics**?

Background

Outfit recommendation

Fashion recommendation machine

Some results

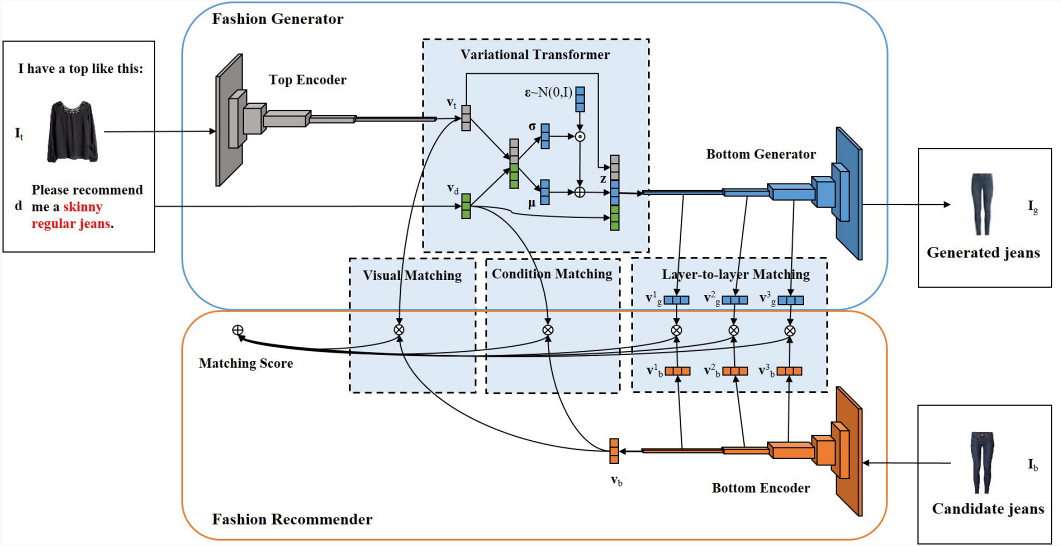
Conclusion

Fashion recommendation machine

Lin et al. (2019) – Improving Outfit Recommendation with Co-supervision of Fashion Generation

1. Neural co-supervision learning framework, FARM, for outfit recommendation that simultaneously yields **recommendation** and **generation**
2. Layer-to-layer matching mechanism as a bridge between generation and recommendation – improves recommendation by leveraging generation features

FARM architecture



FARM architecture

For the **fashion generator**

- Use CNN as top encoder to extract visual features from top image $\mathbf{I}_t$
- Learn semantic representation for bag-of-words vector $\mathbf{d}$ of bottom description
- Use variational transformer to learn mapping from bottom distribution to Gaussian distribution based on visual features of $\mathbf{I}_t$ and semantic representation of $\mathbf{d}$
- Sample a random vector from Gaussian distribution and input it to a DCNN (as bottom generator) to generate bottom image $\mathbf{I}_g$ that matches $\mathbf{I}_t$ and $\mathbf{d}$
- Explicitly forces top encoder to encode more aesthetic matching information into visual features

FARM architecture

For the **fashion recommender**

- Also employs CNN as bottom encoder to extract visual features from candidate bottom image $\mathbf{I}_b$
- Evaluate matching score between $\mathbf{I}_b$ and $(\mathbf{I}_t, \mathbf{d})$ pair from three angles
  1. Visual matching between $\mathbf{I}_b$ and $\mathbf{I}_t$
  2. Description matching between $\mathbf{I}_b$ and $\mathbf{d}$
  3. Layer-to-layer matching between $\mathbf{I}_b$ and $\mathbf{I}_g$, which leverages generation information to improve recommendation

FARM architecture

FARM jointly trains the fashion generator and fashion recommender

Three types of loss

1. Generation loss (visual + textual)
2. Loss based on ELBO
3. Recommendation loss (like BPR)

Background

Outfit recommendation

Fashion recommendation machine

Some results

Conclusion

A sample of results

FashionVC and ExpFashion datasets sampled from Polyvore online community

4-tuples (top, top description, bottom, bottom description)

Bake-off

Table 2: Recommendation results on FashionVC and ExpFashion datasets (%).

	FashionVC			
	Top		Bottom	
Method	AUC	MRR	AUC	MRR
LR	48.7	4.5	46.4	4.4
IBR_d	52.8	6.1	62.9	10.3
BPR-DAE_d	62.9	8.6	70.2	10.9
$DVBPR_d$	64.6	9.1	76.9	13.0
FARM	**71.2***	**12.6***	**77.8**	**15.3***
	ExpFashion			
	Top		Bottom	
Method	AUC	MRR	AUC	MRR
LR	50.5	5.4	48.4	4.4
IBR_d	56.1	7.1	68.9	12.0
BPR-DAE_d	73.0	12.3	79.9	14.7
$DVBPR_d$	82.4	18.5	83.7	15.4
FARM	**85.2***	**25.1***	**88.4***	**24.3***

The superscript * indicates that FARM significantly outperforms $DVBPR_d$, using a paired t-test with $p < 0.05$.

Co-supervision learning

Table 3: Analysis of co-supervision learning. Recommendation results on the FashionVC and ExpFashion datasets (%).

	FashionVC			
	Top		Bottom	
Method	AUC	MRR	AUC	MRR
FARM-G	54.8	8.4	60.9	9.8
FARM-R	68.0	9.8	77.2	12.8
FARM	**71.2***	**12.6***	**77.8**	**15.3***
	ExpFashion			
	Top		Bottom	
Method	AUC	MRR	AUC	MRR
FARM-G	64.4	14.2	72.4	21.3
FARM-R	82.3	18.9	84.2	15.2
FARM	**85.2***	**25.1***	**88.4***	**24.3***

The superscript * indicates that FARM significantly outperforms FARM-R, using a paired t-test with $p < 0.05$.

Layer-to-layer

Table 4: Analysis of layer-to-layer matching. Recommendation results on the FashionVC and ExpFashion datasets (%).

	FashionVC			
	Top		Bottom	
Method	AUC	MRR	AUC	MRR
FARM-WL	59.8	7.6	67.8	8.2
FARM	**71.2***	**12.6***	**77.8***	**15.3***
	ExpFashion			
	Top		Bottom	
Method	AUC	MRR	AUC	MRR
FARM-WL	68.6	9.9	74.3	10.3
FARM	**85.2***	**25.1***	**88.4***	**24.3***

The superscript * indicates that FARM significantly outperforms FARM-WL, using a paired t-test with $p < 0.05$.

Some samples: Real vs generated

(a) Top generation.

(b) Bottom generation.

Figure 3: Comparison between the real and generated images.

Some samples: Recommendations

(a) Top recommendation.

(b) Bottom recommendation.

Figure 4: Case studies of recommendation. The items highlighted in the red boxes are the positive ones.

Some samples: Real vs generated

Figure 5: Case studies of generation. Each case is in the form: "given description + given item = generated item".

Background

Outfit recommendation

Fashion recommendation machine

Some results

Conclusion

What have we done?

Outfit recommendation

- Visual understanding
- Visual matching

Proposed a co-supervision learning framework, FARM

- For visual understanding, FARM captures more aesthetic characteristics with supervision of generation learning
- For visual matching, FARM incorporates layer-to-layer matching mechanism to evaluate matching score of candidate and generated items at different neural layers

What should we do next?

Effectiveness of generated images to **explain** the recommendations?

Improvement in quality of generated images leads to improvement in recommendations?

How to recommend complete outfits?

Playing the winning game

How to improve ourselves

- Compare apples to apples
- Work on insights – reasons for success, reasons for failure
- Use reference baselines
- Share everything
- Use reference implementations
- Engage with product owners for additional eyes and checks
- Win in different ways – task, constraints, metrics, . . .

References i

H. Li and J. Xu. Semantic matching in search. *Foundations and Trends in Information Retrieval*, 7(5):343–469, 2013.

J. Lin. The neural hype and comparisons against weak baselines. *SIGIR Forum*, 52(2):40–51, 2018.

Y. Lin, P. Ren, Z. Chen, Z. Ren, J. Ma, and M. de Rijke. Improving outfit recommendation with co-supervision of fashion generation. In *The Web Conference 2019*, May 2019.

B. Mitra and N. Craswell. An introduction to neural information retrieval. *Foundations and Trends in Information Retrieval*, 13(1), January 2019.

K. D. Onal, Y. Zhang, I. S. Altingovde, M. M. Rahman, P. Karagoz, A. Braylan, B. Dang, H.-L. Chang, H. Kim, Q. McNamara, A. Angert, E. Banner, V. Khetan, T. McDonnell, A. T. Nguyen, D. Xu, B. C. Wallace, M. de Rijke, and M. Lease. Neural information retrieval: At the end of the early years. *Information Retrieval Journal*, 21(2–3):111–182, June 2018.

Acknowledgments

All content represents the opinion of the author(s), which is not necessarily shared or endorsed by their employers and/or sponsors.