

Using Reviews to Create a Recommender System That Works

Originally published at <https://www.unite.ai/using-reviews-to-create-a-recommender-system-that-works/>



If you have ever bought a product online and marveled at the inanity and non-applicability of the ‘related items’ that haunt the buying and after-sales process, you already understand that popular and mainstream [recommender systems](#) tend to fall short in terms of understanding the relationships between prospective purchases.

If you buy a unlikely and infrequent item, such as an oven, recommendations for other ovens are likely to be superfluous, though the worst recommender systems [fail to acknowledge this](#). In the 2000s, for example, TiVo’s recommender system created an early controversy in this sector by [reassigning the perceived sexuality](#) of a user, who subsequently sought to ‘re-masculinize’ his user profile by selecting war movies – a crude approach to algorithm revision.

Worse yet, you don’t need to actually buy anything at (for instance) Amazon, or actually begin watching a movie whose description you’re browsing at any major streaming platform, in order for [information-starved](#) recommender algorithms to start merrily down the wrong path; searches, dwells and clicks into the ‘details’ pages are enough, and this scant (and probably incorrect) information is likely to be perpetuated across future browsing sessions at the platform.

Trying to make a Recommender System Forget

Sometimes it’s possible to intervene: Netflix provides a ‘thumbs up/down’ system which should in theory help its machine learning algorithms remove certain embedded concepts and words from your recommendations profile (though its efficacy has [been questioned](#), and it remains much easier to evolve a personalized recommender algorithm from scratch [than it is to remove undesired ontologies](#)), while Amazon lets you [remove titles](#) from your customer history, which should downgrade any unwelcome domains that infiltrated your recommendations.

Hulu has a [similar feature](#), while HBO Max has [partially retreated](#) from algorithm-only recommender systems, in the face of their current shortcomings.

None of these strictly consumer-level experiences even touch on the widespread and growing criticism of ‘passive’ advertising platform recommender systems (where [notable change is coming](#) due to public ire), or the incendiary topic of social media AI recommendations, where sites such as [YouTube](#), [Twitter](#) and [Facebook](#) continue to endure criticism for non-relevant or even damaging recommendations.

The machine doesn’t seem to know what we want, unless we want the *adjacent item* that came up in our search – even if that item is essentially a duplicate or alternate to the primary item that we may have just bought, rather than a potential complementary or ancillary purchase.

Accurate Recommendations with Review Data

A new research collaboration from China and Australia offers a novel method to address such non-apposite recommendations, by using external user-reviews to gain a better understanding of the real relationships between items in a shopping session. In tests, the architecture outperformed all current state-of-the-art methods, offering hope for recommender systems that have a better internal map of the dependencies of items:

Table 1: Experimental Results on Sessions with More than 1 Item.												
Dataset	Pet Supplies						Movies and TV					
	P@10	P@20	MRR@10	MRR@20	P@10	P@20	MRR@10	MRR@20	P@10	P@20	MRR@10	MRR@20
S-PQI	6.52	6.52	5.25	5.26	1.83	1.84	1.59	1.59				
S-KNN	21.86	24.30	12.22	12.39	20.33	23.46	8.40	8.62				
GRU4Rec	8.79	9.81	6.26	6.34	8.32	8.37	5.61	5.72				
NARM	24.48	27.68	19.48	19.70	22.40	25.40	17.05	17.26				
STAMP	21.50	24.66	16.41	16.60	20.59	23.96	14.69	14.92				
BERT4Rec	24.18	27.40	19.29	19.51	23.21	26.46	17.75	17.97				
SR-GNN	24.64	27.81	19.53	19.75	23.69	26.74	18.24	18.45				
GCE-GNN	24.73	28.28	18.16	18.41	24.25	27.64	17.02	17.25				
DHCCN	24.23	27.45	17.80	18.02	23.33	26.49	15.94	16.16				
RI-GNN	25.43	28.88	19.93	20.15	24.69	27.94	18.93	19.14				
Improv. (%)	2.83	2.12	2.05	2.02	1.81	1.09	3.78	3.74				

Table 2: Experimental Results on Sessions with More than 5 Items.												
Dataset	Pet Supplies						Movies and TV					
	P@10	P@20	MRR@10	MRR@20	P@10	P@20	MRR@10	MRR@20	P@10	P@20	MRR@10	MRR@20
NARM	26.07	30.34	19.33	19.62	20.98	24.73	14.27	14.53				
STAMP	24.44	28.06	18.28	18.53	21.27	25.40	14.49	14.77				
BERT4Rec	25.85	30.04	19.49	19.77	22.24	26.43	15.87	16.15				
SR-GNN	26.29	30.49	19.31	19.60	22.73	26.57	16.11	16.37				
GCE-GNN	25.95	30.36	17.51	17.81	23.18	27.51	14.39	14.71				
DHCCN	24.96	29.20	17.25	17.54	21.83	25.71	13.01	13.27				
RI-GNN	27.72	32.11	20.26*	20.57*	24.15*	28.30*	16.99*	17.27*				
Improv. (%)	5.44	5.31	3.95	4.05	4.19	2.87	5.46	5.50				

RI-GNN outperforms major competitors in terms of accuracy of relationships between items, performing best on sessions with more than five items. The system was tested on *Pet Supplies* and *Movies and TV* datasets from [Amazon Review Data](#) (2018). Source: <https://arxiv.org/pdf/2201.12532.pdf>

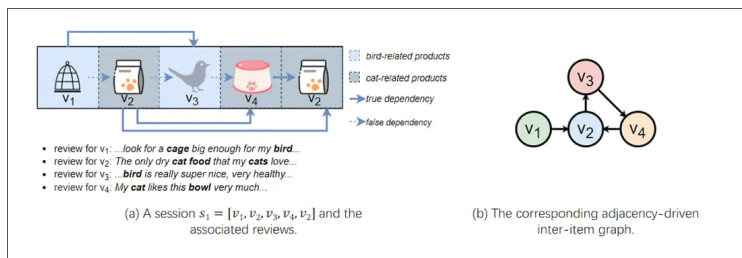
To boot, the project addresses the notable challenge of creating recommendations even in anonymous sessions, where the recommender system has no access to user-contributed details, such as purchase history, or the user's own online reviews of prior purchases.

The new [paper](#) is called *Rethinking Adjacent Dependency in Session-based Recommendations*, and comes from researchers at the Qilu University of Technology and the Beijing Institute of Technology in China, RMIT University at Melbourne, and the Australian Artificial Intelligence Institute at the University of Technology Sydney.

What's Next?

The core task of session-based recommendations (SBR) is to determine the 'next' item along from the current item, based on its calculated relationship to the current item. In practical terms, this could manifest as a list of 'Related items' in an item page for a bird-cage at an ecommerce web site.

If you're buying a bird cage, what else are you likely to need? Well, at the very least, you're going to need a bird to put in it – that's a *true dependency*. However, the bird-cage is featured in the ontology *pet goods*, where birds are not sold. Perversely, *cat food* sits in the same ontology, though appending a cat-feeding bowl as an associated recommendation for a bird cage product is a *false dependency* – a mistaken and misguided association.

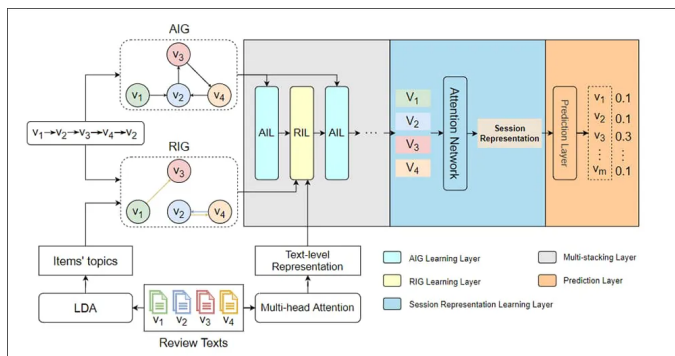


From the paper: true and false relationships between several items, visualized on the right as an inter-item graph.

As is so often the case in machine learning architectures, it is a challenge to persuade a recommender system that a 'distant' entity (*bird* does not feature at all in *pet products*) may have an intrinsic and important relationship to an item, whereas items that are in the same category, and very close in function and central concept (such as *cat feeding bowl*), may be orthogonal or directly opposed to the purchase being considered.

The only way to create these mappings between 'non-adjacent' entities is to crowdsource the problem, since the relationships in question are a facet of human experience, can't be guessed programmatically, and are probably beyond the affordable scope of conventional approaches to dataset labeling, such as [Amazon Mechanical Turk](#).

Therefore the researchers have employed Natural Language Processing (NLP) mechanisms to extract salient words from reviews for a product, and have used frequencies from these analyses to create embeddings capable of 'matching' apparently distant items.



The architecture for Review-refined Inter-item Graph Neural Network (RI-GNN).

Architecture and Data

As the new paper notes, prior works of a similar nature have exploited a logged-in user's own review history to provide rudimentary mappings. [DeepCONN](#) and [RNS](#) both used this approach. However, this discounts the fact that a user may not have written any reviews, or any reviews pertinent to a particular item that is 'out of range' of their usual buying habits. Additionally, this is something of a 'white box' approach, since it assumes that the user has already engaged sufficiently with the outlet to create an account and log in.

The extended Graph Neural Network (GNN) proposed by the researchers takes a more oracle-driven approach, deriving true dependencies *a priori*, so that, presumably, the anonymous and logged-out user can experience more relevant recommendations with minimal input required.

The review-augmented system is titled *Review-refined Inter-item Graph Neural Network* (RI-GNN). The researchers have tested it against two [datasets from Amazon](#), *Pet Supplies* and *Movies and TV*. Though this solves the problem of review availability rather neatly, an in-the-wild implementation would need to locate and scrape an appropriate reviews database. Such a dataset source could, in theory, be anything from posts on a social network to answers on Quora.

High-level relationship mappings of this nature would, additionally, be valuable to a range of machine learning applications beyond recommender systems. Many current projects are hamstrung by lack of inter and intra-domain mapping due to limited funds and scope, whereas the commercial impetus of a truly knowledgeable and crowdsourced ecommerce recommender system could potentially fill that gap.

Metrics and Testing

The authors tested RI-GNN against two versions of each dataset, each of which is comprised of a user's purchase history and general reviews of the product. Items appearing less than five times were removed, and the user history split into units of a week. The first dataset version featured all sessions with more than one item, and the second all sessions with over five items.

The project used P@K (Precision) and MRR@K (Mean Reciprocal Rank) for its evaluation metrics. Rival architectures tested were: [S-KNN](#); [GRU4Rec](#); [S-POP](#); [STAMP](#); [BERT4Rec](#); [DHCN](#); [GCE-GNN](#); [SR-GNN](#); and [NARM](#).

The framework was trained in batches of 100 on [Adam](#) at a learning rate of 0.001, with the number of topics set to 24 and 20, respectively, for *Pet Supplies* and *Movies and TV*.

First published 1st February 2022.

Writer on machine learning, domain specialist in human image synthesis. Former head of research content at Metaphysic.ai, until its dissolution into DNEG's Brahma.ai.

Portfolio site: martinanderson.ai

Contact: martin@martinanderson.ai



Discover More