Learning Recommendations from User Actions in the Item-poor Insurance Domain

November 2022

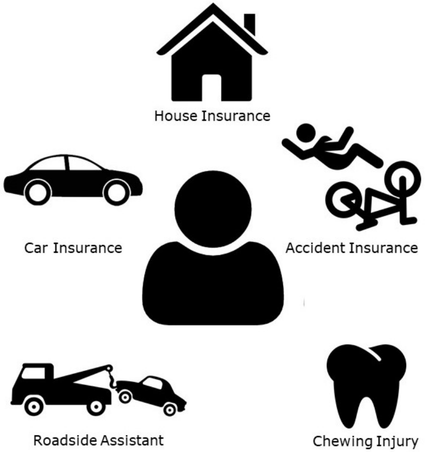
Simone Borg Bruun

UNIVERSITY OF COPENHAGEN

Introduction to the Insurance Domain

Items: Insurance products for individuals and additional coverages of insurances.

Objective: Personalised recommendations can help customers continuously adjust their insurances to suit their needs.



Characteristics of the Insurance Domain

- Customers have **dynamic needs** of insurance products as they are closely connected to life events.
- The insurance domain is **data-sparse** because of few different items and low purchase frequency.
- Many users navigate through the insurance website, but still prefer to purchase over the **phone**.

Characteristics of the Insurance Domain

- Customers have **dynamic needs** of insurance products as they are closely connected to life events.

  $\Rightarrow$ Session-based approach.

- The insurance domain is **data-sparse** because of few different items and low purchase frequency.

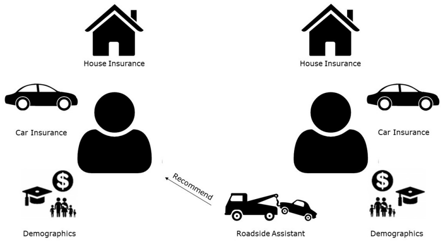
  $\Rightarrow$ Several types of user actions.

- Many users navigate through the insurance website, but still prefer to purchase over the **phone**.

  $\Rightarrow$ Relationships between sessions and purchases after the sessions.

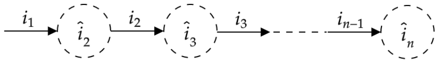
Related Work

Insurance Domain



Categorise the users based on demographic characteristics.

Sessions-based Recommender System



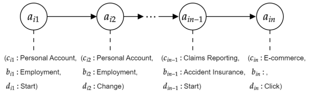
For every step in a user's session, $\{i_1, i_2, i_3, \ldots\}$, the task is to predict the next item, the user is going to interact with.

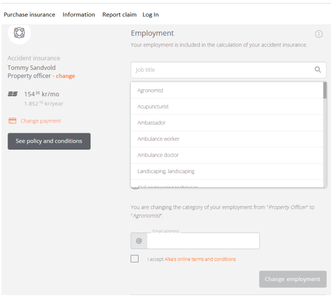
Representation of Sessions

User session: $s_i = \{a_{i1}, a_{i2}, a_{i3}, \dots, a_{in}\}$

Action representation: $a_{ij} = (c_{ij}, b_{ij}, d_{ij})$

- c_{ij}: action section
- b_{ij}: action object
- d_{ij}: action type





Problem Formalization

Task: $f(s_1, s_2, s_3, ..., s_m) = (\hat{p}_1, \hat{p}_2, \hat{p}_3, ..., \hat{p}_K)$

Recent sessions: $\{s_1, s_2, s_3, ..., s_m\}$

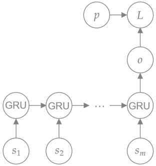
Session threshold: s_1 and s_2 belong to the same task if $\texttt{start\_time}(s_2) - \texttt{start\_time}(s_1) \leq t$

User session: $s_i = \{a_{i1}, a_{i2}, a_{i3}, ..., a_{in}\}$

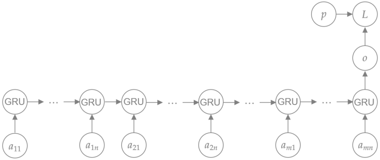
Approach

Cross-sessions Encode:

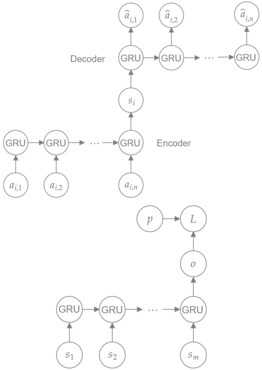
$$s_i = \max_{element}(a_{i1}, a_{i2}, a_{i3}, \ldots, a_{in})$$



Cross-sessions Concat:



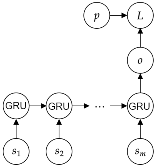
Cross-sessions Auto:



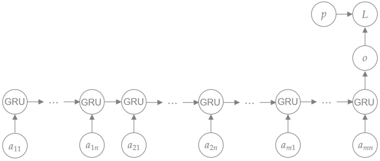
Approach

Cross-sessions Encode:

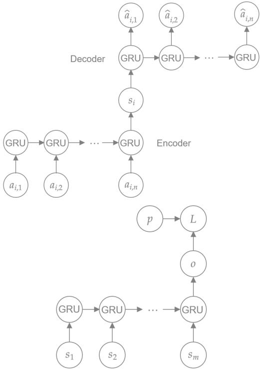
$$s_i = \max_{element}(a_{i1}, a_{i2}, a_{i3}, \ldots, a_{in})$$



Cross-sessions Concat:



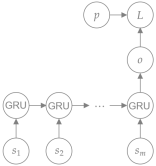
Cross-sessions Auto:



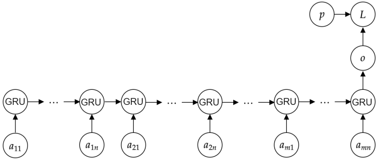
Approach

Cross-sessions Encode:

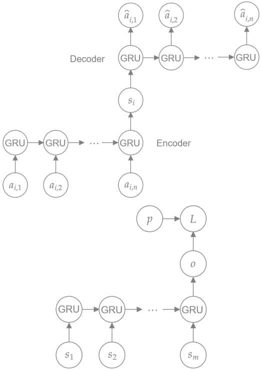
$$s_i = \max_{element}(a_{i1}, a_{i2}, a_{i3}, \ldots, a_{in})$$



Cross-sessions Concat:



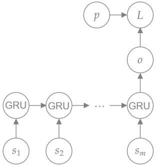
Cross-sessions Auto:



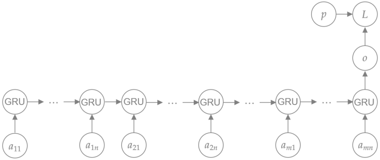
Approach

Cross-sessions Encode:

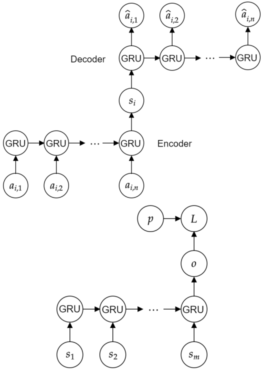
$s_i = \max_{element}(a_{i1}, a_{i2}, a_{i3}, \ldots, a_{in})$



Cross-sessions Concat:



Cross-sessions Auto:



Experiment

Real-world dataset spanning a 2 year period:

Users	44,434
Items	16
Purchases	53,757
Sessions	117,163

Evaluation measures for ranking:

HR@3, Precision@3, Recall@3, MRR@3 and MAP@3.

Insurance baselines and session-based baselines:

$$\left.\begin{array}{l}\text{SVD}\\ \text{Demographic}\end{array}\right\}\text{ insurance portfolios and demographic data}$$

$$\left.\begin{array}{l}\text{SKNN\_E}\\ \text{SKNN\_EB}\\ \text{GRU4REC}\\ \text{GRU4REC Concat}\end{array}\right\}\text{ user session data}$$

Experiment

Results:

Model	HR@3	Precision@3	Recall@3	MRR@3	MAP@3
Popular	0.6217*	0.2145*	0.5855*	0.4764*	0.4540*
SVD	0.6646*	0.2372*	0.6327*	0.4997*	0.4829*
Demographic	0.7392*	0.2649*	0.7095*	0.5620*	0.5446*
GRU4REC	0.6479*	0.2313*	0.6208*	0.5443*	0.5264*
GRU4REC Concat	0.6616*	0.2365*	0.6362*	0.5620*	0.5453*
SKNN_E	0.8106*	0.2914*	0.7848*	0.6740*	0.6567*
SKNN_EB	0.8132*	0.2922*	0.7872*	0.6785*	0.6610*
Cross-sessions Encode	**0.8380 (3.04%)**	**0.3030 (3.67%)**	**0.8145 (3.46%)**	**0.7093 (4.53%)**	**0.6923 (4.73%)**
Cross-sessions Concat	0.8265 (1.62%)	0.2984 (2.12%)	0.8019 (1.87%)	0.7051 (3.92%)	0.6876 (4.02%)
Cross-sessions Auto	0.8356 (2.74%)	0.3024 (3.48%)	0.8128 (3.24%)	0.7085 (4.41%)	0.692 (4.69%)

All results marked with * are significantly different from cross-sessions encode. The best score for each measure is in bold. Percentages in brackets denote the difference of our models from the strongest baseline (SKNN_EB).

Experiment

Results:

Model	HR@3	Precision@3	Recall@3	MRR@3	MAP@3
Popular	0.6217*	0.2145*	0.5855*	0.4764*	0.4540*
SVD	0.6646*	0.2372*	0.6327*	0.4997*	0.4829*
Demographic	0.7392*	0.2649*	0.7095*	0.5620*	0.5446*
GRU4REC	0.6479*	0.2313*	0.6208*	0.5443*	0.5264*
GRU4REC Concat	0.6616*	0.2365*	0.6362*	0.5620*	0.5453*
SKNN_E	0.8106*	0.2914*	0.7848*	0.6740*	0.6567*
SKNN_EB	0.8132*	0.2922*	0.7872*	0.6785*	0.6610*
Cross-sessions Encode	**0.8380 (3.04%)**	**0.3030 (3.67%)**	**0.8145 (3.46%)**	**0.7093 (4.53%)**	**0.6923 (4.73%)**
Cross-sessions Concat	0.8265 (1.62%)	0.2984 (2.12%)	0.8019 (1.87%)	0.7051 (3.92%)	0.6876 (4.02%)
Cross-sessions Auto	0.8356 (2.74%)	0.3024 (3.48%)	0.8128 (3.24%)	0.7085 (4.41%)	0.692 (4.69%)

SKNN and cross-sessions outperform the non-session-based methods.

This is not the case for GRU4REC.

All results marked with * are significantly different from cross-sessions encode. The best score for each measure is in bold. Percentages in brackets denote the difference of our models from the strongest baseline (SKNN_EB).

Experiment

Results:

Model	HR@3	Precision@3	Recall@3	MRR@3	MAP@3
Popular	0.6217*	0.2145*	0.5855*	0.4764*	0.4540*
SVD	0.6646*	0.2372*	0.6327*	0.4997*	0.4829*
Demographic	0.7392*	0.2649*	0.7095*	0.5620*	0.5446*
GRU4REC	0.6479*	0.2313*	0.6208*	0.5443*	0.5264*
GRU4REC Concat	0.6616*	0.2365*	0.6362*	0.5620*	0.5453*
SKNN_E	0.8106*	0.2914*	0.7848*	0.6740*	0.6567*
SKNN_EB	0.8132*	0.2922*	0.7872*	0.6785*	0.6610*
Cross-sessions Encode	**0.8380 (3.04%)**	**0.3030 (3.67%)**	**0.8145 (3.46%)**	**0.7093 (4.53%)**	**0.6923 (4.73%)**
Cross-sessions Concat	0.8265 (1.62%)	0.2984 (2.12%)	0.8019 (1.87%)	0.7051 (3.92%)	0.6876 (4.02%)
Cross-sessions Auto	0.8356 (2.74%)	0.3024 (3.48%)	0.8128 (3.24%)	0.7085 (4.41%)	0.692 (4.69%)

The cross-sessions methods outperform SKNN.

All results marked with * are significantly different from cross-sessions encode. The best score for each measure is in bold. Percentages in brackets denote the difference of our models from the strongest baseline (SKNN_EB).

Experiment

Results:

Model	HR@3	Precision@3	Recall@3	MRR@3	MAP@3
Popular	0.6217*	0.2145*	0.5855*	0.4764*	0.4540*
SVD	0.6646*	0.2372*	0.6327*	0.4997*	0.4829*
Demographic	0.7392*	0.2649*	0.7095*	0.5620*	0.5446*
GRU4REC	0.6479*	0.2313*	0.6208*	0.5443*	0.5264*
GRU4REC Concat	0.6616*	0.2365*	0.6362*	0.5620*	0.5453*
SKNN_E	0.8106*	0.2914*	0.7848*	0.6740*	0.6567*
SKNN_EB	0.8132*	0.2922*	0.7872*	0.6785*	0.6610*
Cross-sessions Encode	**0.8380 (3.04%)**	**0.3030 (3.67%)**	**0.8145 (3.46%)**	**0.7093 (4.53%)**	**0.6923 (4.73%)**
Cross-sessions Concat	0.8265 (1.62%)	0.2984 (2.12%)	0.8019 (1.87%)	0.7051 (3.92%)	0.6876 (4.02%)
Cross-sessions Auto	0.8356 (2.74%)	0.3024 (3.48%)	0.8128 (3.24%)	0.7085 (4.41%)	0.692 (4.69%)

Encoding of sessions is better than concatenating sessions.

All results marked with * are significantly different from cross-sessions encode. The best score for each measure is in bold. Percentages in brackets denote the difference of our models from the strongest baseline (SKNN_EB).

Experiment

Results:

Model	HR@3	Precision@3	Recall@3	MRR@3	MAP@3
Popular	0.6217*	0.2145*	0.5855*	0.4764*	0.4540*
SVD	0.6646*	0.2372*	0.6327*	0.4997*	0.4829*
Demographic	0.7392*	0.2649*	0.7095*	0.5620*	0.5446*
GRU4REC	0.6479*	0.2313*	0.6208*	0.5443*	0.5264*
GRU4REC Concat	0.6616*	0.2365*	0.6362*	0.5620*	0.5453*
SKNN_E	0.8106*	0.2914*	0.7848*	0.6740*	0.6567*
SKNN_EB	0.8132*	0.2922*	0.7872*	0.6785*	0.6610*
Cross-sessions Encode	**0.8380 (3.04%)**	**0.3030 (3.67%)**	**0.8145 (3.46%)**	**0.7093 (4.53%)**	**0.6923 (4.73%)**
Cross-sessions Concat	0.8265 (1.62%)	0.2984 (2.12%)	0.8019 (1.87%)	0.7051 (3.92%)	0.6876 (4.02%)
Cross-sessions Auto	0.8356 (2.74%)	0.3024 (3.48%)	0.8128 (3.24%)	0.7085 (4.41%)	0.692 (4.69%)

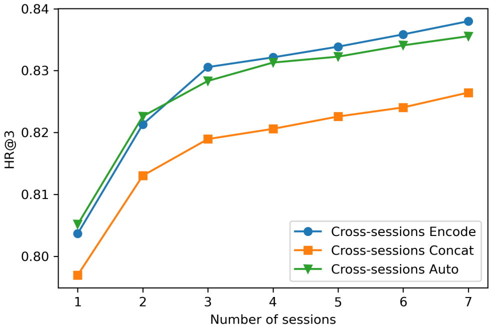
Model	HR@3	Precision@3	Recall@3	MRR@3	MAP@3
Cross-sessions Encode with Demographic	**0.8542* (5.03%)**	**0.3103* (6.17%)**	**0.8313* (5.6%)**	0.7268* (7.11%)	0.7099* (7.41%)
Cross-sessions Concat with Demographic	0.8497* (4.48%)	0.3087* (5.64%)	0.8269* (5.04%)	**0.7298* (7.55%)**	**0.7131* (7.88%)**
Cross-sessions Auto with Demographic	0.8460 (4.03%)	0.3072 (5.13%)	0.8228 (4.52%)	0.7223 (6.45%)	0.7050 (6.66%)

Hybrid of cross-sessions and demographic model.

All results marked with * are significantly different from cross-sessions encode. The best score for each measure is in bold. Percentages in brackets denote the difference of our models from the strongest baseline (SKNN_EB).

Analysis

Performance broken down by number of session, starting with only the most recent session, up to including all the available sessions.



There is additional contribution in using all sessions of each user rather than just the last one.

Analysis

Performance after randomly shuffling the session order. Relative change in parentheses.

Model		HR@3	Precision@3	Recall@3	MRR@3	MAP@3
Cross-sessions Encode	original session order	0.8380	0.3030	0.8145	0.7093	0.6923
	shuffled session order	0.8345 (-0.41%)	0.3008 (-0.7%)	0.8096 (-0.60%)	0.7058 (-0.49%)	0.688 (-0.61%)
Cross-sessions Concat	original session order	0.8265	0.2984	0.8019	0.7051	0.6876
	shuffled session order	0.8209 (-0.20%)	0.2935 (-0.52%)	0.7925 (-0.37%)	0.6978 (-0.28%)	0.6759 (-0.49%)
Cross-sessions Auto	original session order	0.8356	0.3024	0.8128	0.7085	0.6920
	shuffled session order	0.8305 (-0.61%)	0.3003 (-0.70%)	0.8069 (-0.72%)	0.704 (-0.64%)	0.6867 (-0.76%)

The decrease in performance is limited to less than 1%.

Analysis

Ablation study to analyse the influence of different actions. Relative change in parentheses. (We observe similar results for cross-sessions concat and cross-sessions auto.)

Model		HR@3	Precision@3	Recall@3	MRR@3	MAP@3
Cross-sessions Encode	all actions	0.8380	0.3030	0.8145	0.7093	0.6923
	without E-commerce	0.7526 (-10.19%)	0.2698 (-10.95%)	0.7249 (-11.00%)	0.5951 (-16.09%)	0.5764 (-16.74%)
	without Claims reporting	0.8250 (-1.55%)	0.2979 (-1.68%)	0.8012 (-1.64%)	0.7006 (-1.22%)	0.6829 (-1.35%)
	without Information	0.8317 (-0.75%)	0.3000 (-0.98%)	0.8072 (-0.89%)	0.7045 (-0.68%)	0.6863 (-0.87%)
	without Personal account	0.8067 (-3.73%)	0.2894 (-4.48%)	0.7803 (-4.19%)	0.6604 (-6.89%)	0.6438 (-7.00%)
	without Items	0.7379 (-11.94%)	0.2652 (-12.46%)	0.7116 (-12.63%)	0.5720 (-19.36%)	0.5548 (-19.86%)
	without Services	0.8032 (-4.15%)	0.2880 (-4.93%)	0.7765 (-4.66%)	0.6639 (-6.40%)	0.6465 (-6.61%)
	without Start	0.8162 (-2.60%)	0.2935 (-3.11%)	0.7906 (-2.94%)	0.6771 (-4.55%)	0.6592 (-4.77%)
	without Act	0.8318 (-0.73%)	0.3005 (-0.80%)	0.8082 (-0.77%)	0.7035 (-0.82%)	0.6864 (-0.85%)
	without Complete	0.8317 (-0.75%)	0.3005 (-0.82%)	0.8078 (-0.82%)	0.7036 (-0.80%)	0.6861 (-0.89%)

All considered action types are beneficial for the model.

Conclusion and Future Work

We have

- tackled insurance recommendations with a session-based approach,
- used an RNN framework to model relationships between multiple actions and purchases.

We plan to

- run A/B-test to evaluate our models with online users,
- study explainability of the models.

Thank you!

Questions?

simoneborgbruun@di.ku.dk