



Goal-driven Command Recommendations for Analysts

Samarth Aggarwal | IIT Delhi
Rohin Garg | IIT Kanpur
Abhilasha Sancheti | Adobe Research
Bhanu Prakash Reddy Guda* | **Adobe Research**
Iftikhar Ahamath Burhanuddin | Adobe Research

#AdobeRemix
Vasjen Katro / Baugasm

There has been a tremendous growth in the domain of data analysis as the volume of data has increased



Data from various sources



Decision Making
Reporting and Analysis



Insights from data



Demand for systems to query, analyse and draw inferences with a low lat



Research Question

How to guide the analyst with his/her workflow
based on input goal information?



Recommender System

- Act as a **primary filter** for options that are **completely irrelevant**.
- A **Guide** for a novice user.

Guide towards the insights that the **analyst needs**.



Improve performance of recommender system with help of goal information



Recommendations should be **relevant** to the analyst's goal



Steer the recommendations whenever the analyst is deviating from the goal



Related Work

- The concept of goals in process mining, web mining, education, and HCI.
- Definition of a goal - A set of target tasks or subtasks at the focus of a user's attention^[1]
- Liu, Xumin ^[2] - probabilistic suffix trees (PST) for modelling workflows; topic modelling to determine the workflows or tasks.
- Nambhi, Aadhavan M., et al. ^[3] proposed neural network architecture to predict next command in the sequence.

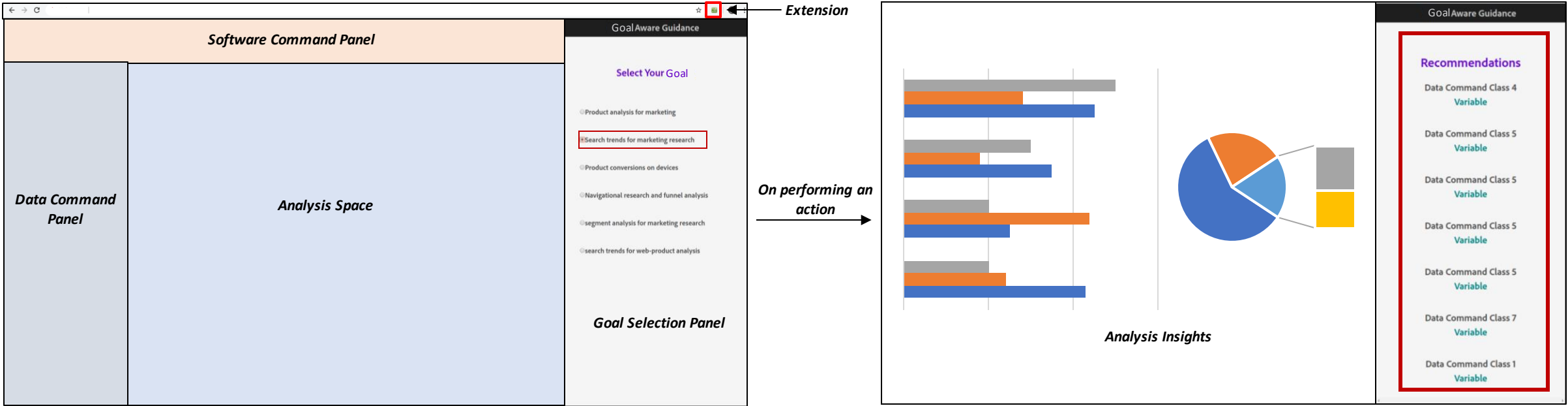
[1] Horvitz, Eric J., et al. "The Lumiere project: Bayesian user modeling for inferring the goals and needs of software users." *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, July 1998.

[2] Liu, Xumin. "Unraveling and learning workflow models from interleaved event logs." *2014 IEEE International Conference on Web Services*. IEEE, 2014.

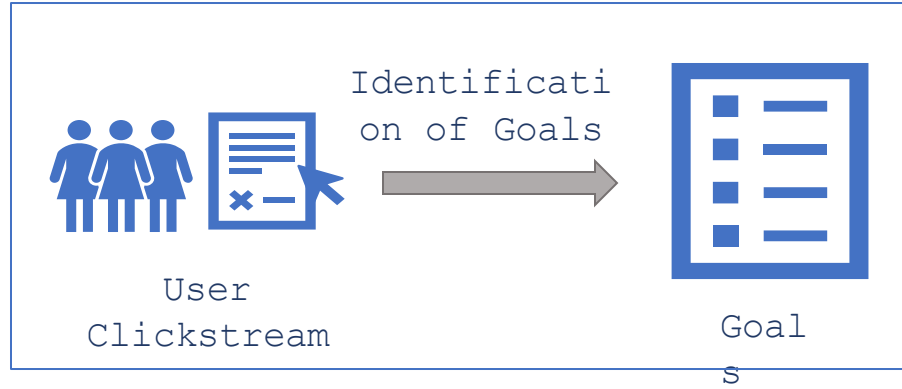
[3] Nambhi, Aadhavan M., et al. "Stuck? No worries! Task-aware Command Recommendation and Proactive Help for Analysts." *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*. 2019.



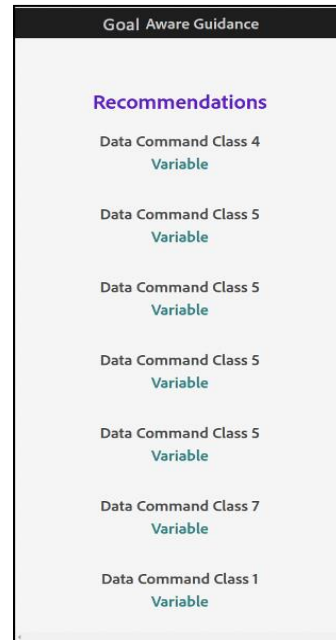
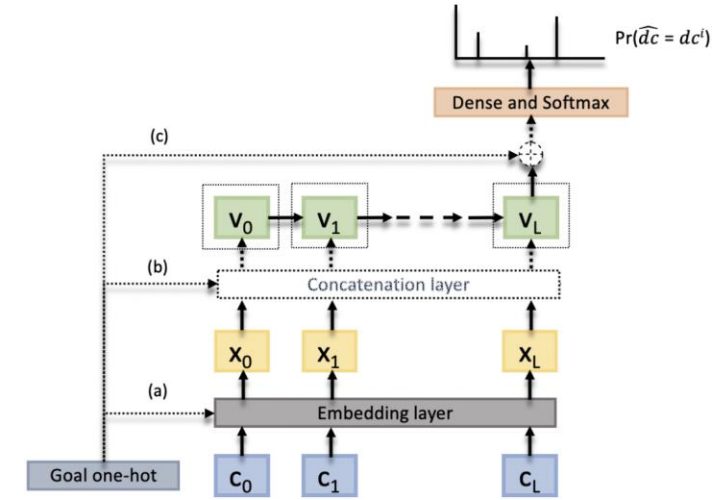
Workflow of our approach



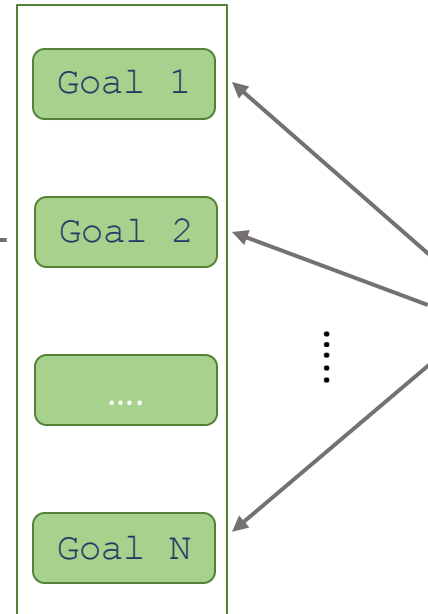
Pre-processed information



Generalised Model Training with Goal Information



Goal specific recommendations



Goal specific models

Goal Specific Fine Tuning

Goal Oriented Loss Function



Dataset

- A web-analytics system used to track, report, analyze, and visualize web traffic.
- User interface clicks classified into two categories
 - Software Commands - Ex. Opening, loading or saving a spreadsheet, changing the color of a cell.
 - Data Commands - Ex. sorting a data column in an increasing or decreasing order.

`SC, DC, DC, SC, ..., SC, SC, SC, DC, DC, SC, ..`

Logs from: April 2018 to June 2018.

Average session length: 30.

Window size to be 30.

Total sequences: 2.7M

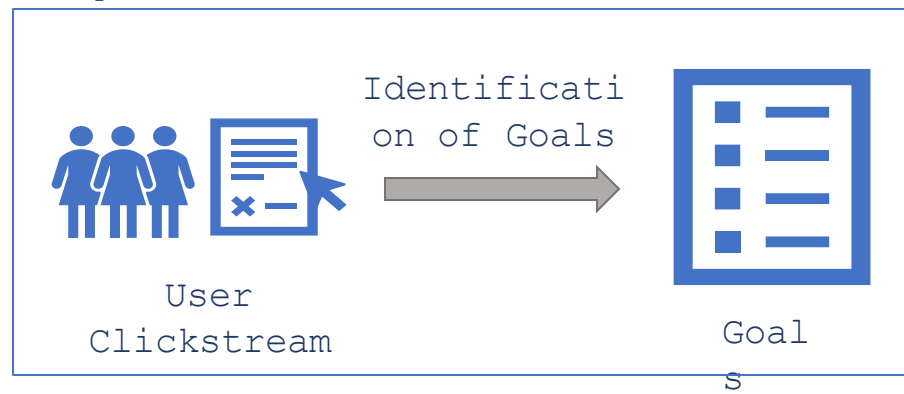
train : valid : test - 75 :

12.5 : 12.5

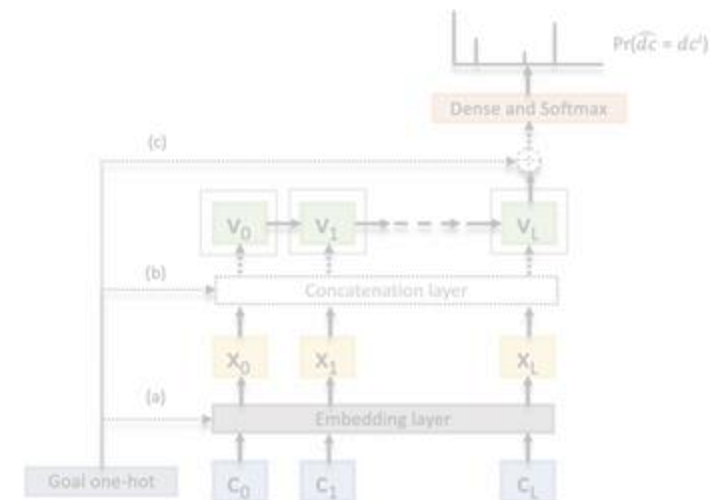
A command or an action is a click that gets registered in the log data when a user interacts with the interface of the system.



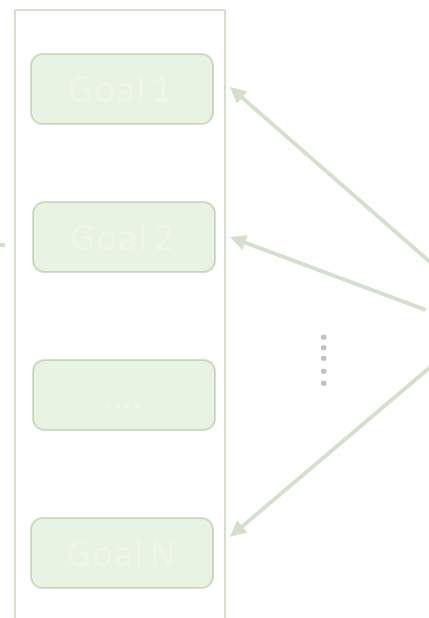
Pre-processed information



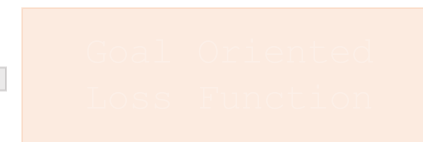
Generalised Model Training with Goal Information



Goal specific recommendations



Goal specific models



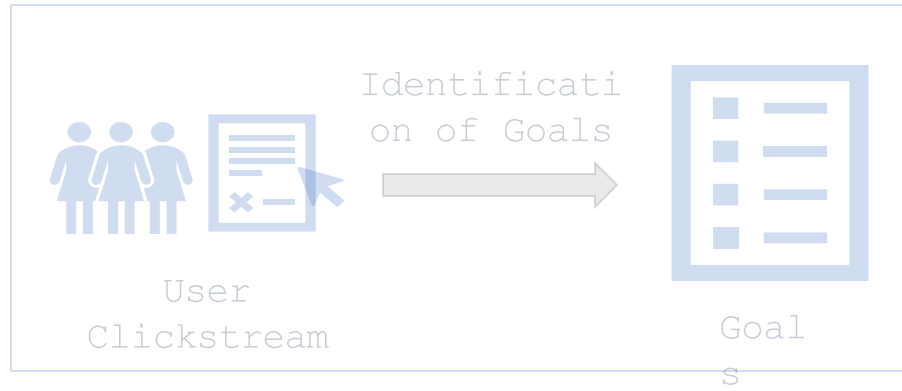
Identification of Goals

- Analogy: Identification of Identification of Goals
 Topics
 Documents Sessions
 Topics Goals
 Words Commands
- Bi-term Topic Modeling (BTM) ^[1]
- Fixing the number of goals
 - Goal Coherence
 - UCI
 - UMass
 - Human Evaluation

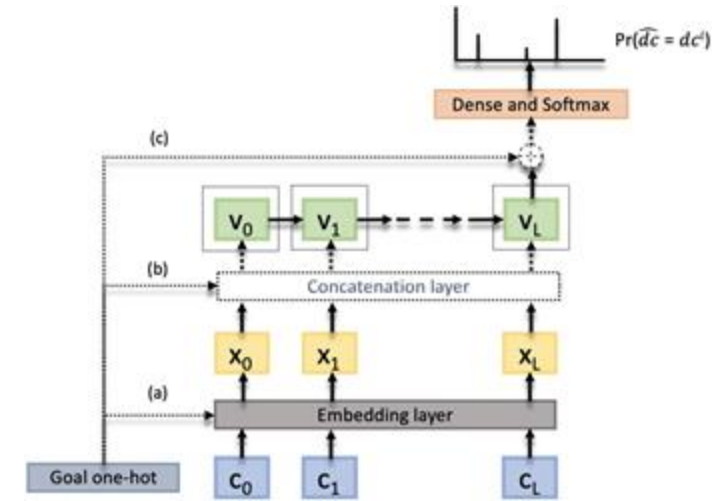
[1] Yan, Xiaohui, et al. "A biterm topic model for short texts." *Proceedings of the 22nd international conference on World Wide Web*. 2013.



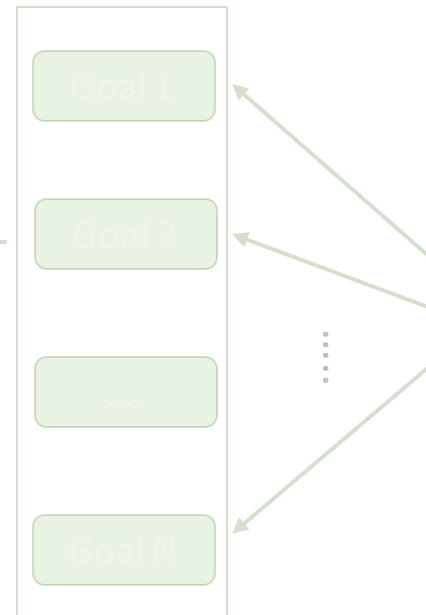
Pre-processed information



Generalised Model
Training with Goal
Information



Goal specific
recommendations



Goal specific
models

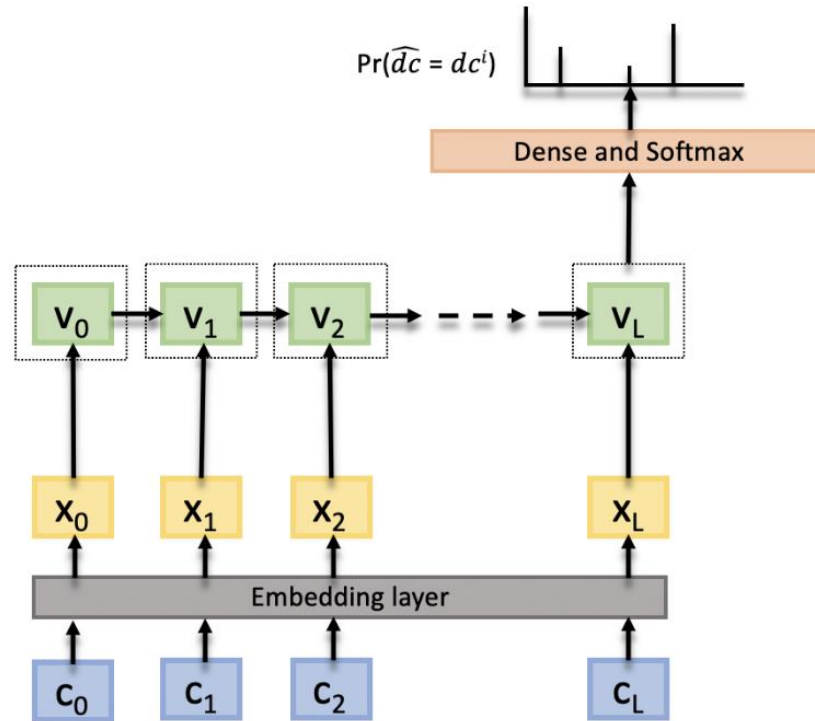
Goal Specific
Fine Tuning

Goal Oriented
Loss Function



Goal-Driven Recommender System

- Ensemble models - data segregation
- Goal informed models



One model per Goal

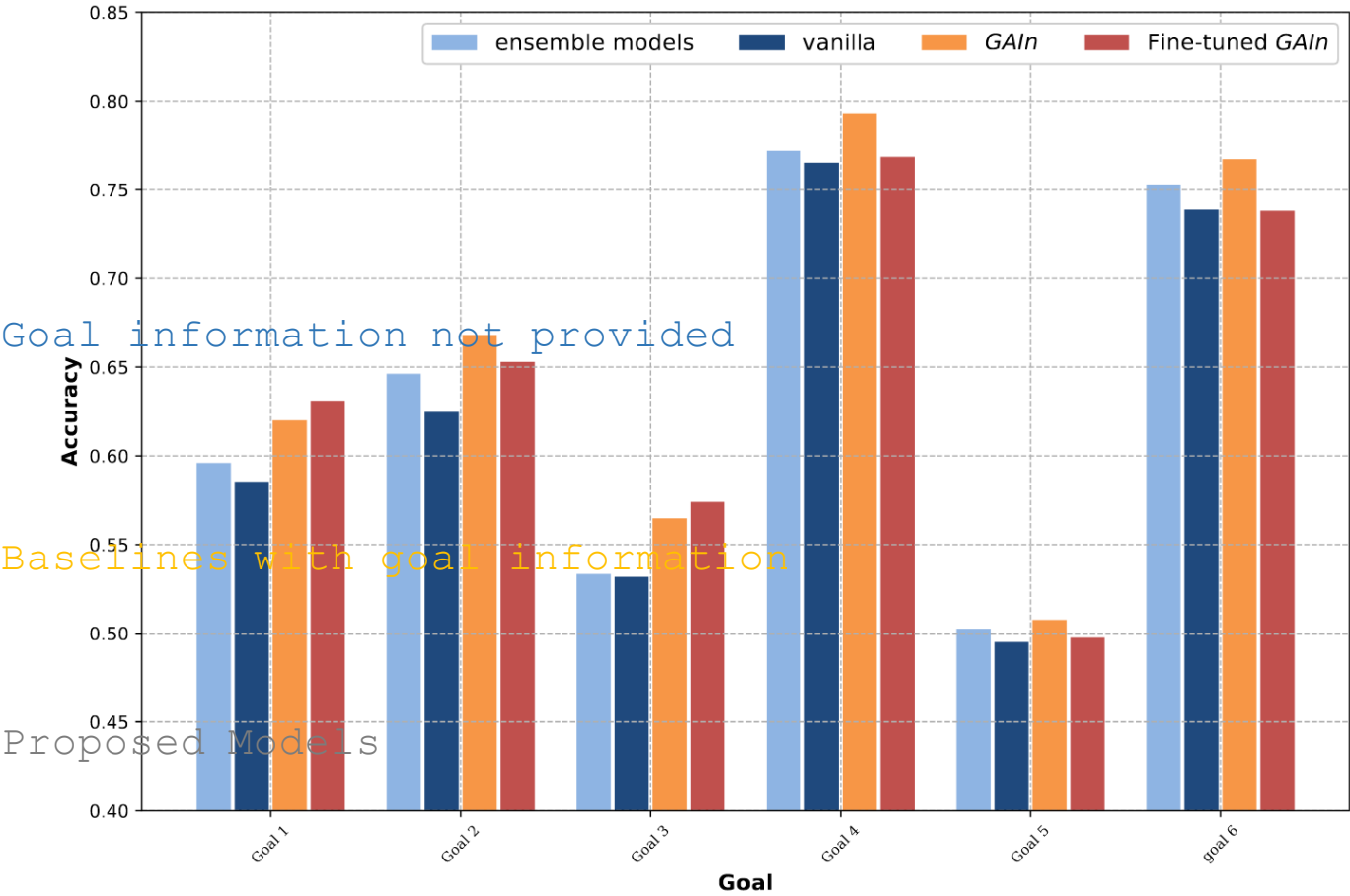
Single model



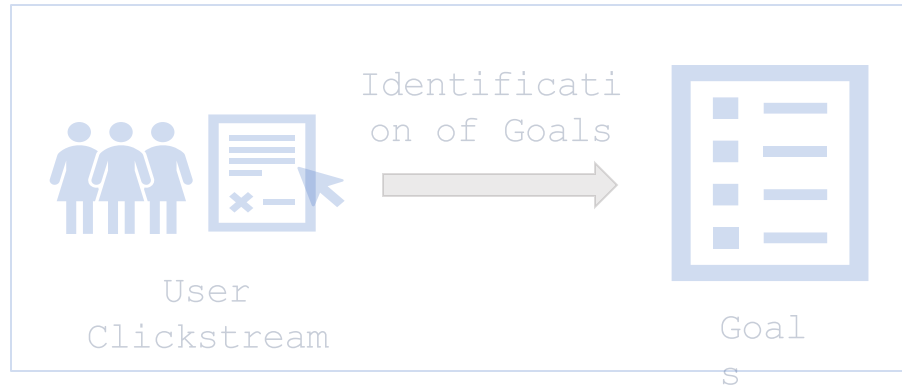


Improve performance of recommender system with help of goal information

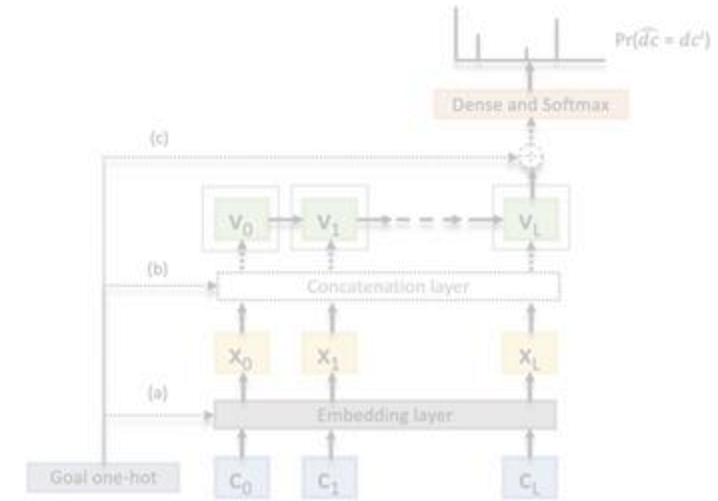
MODEL	ACCURACY
-------	----------



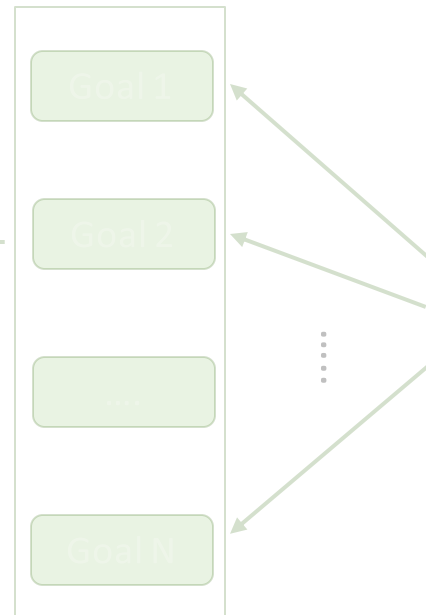
Pre-processed information



Generalised Model Training with Goal Information



Goal specific recommendations



Goal specific models



Goal-Oriented Loss Function

- Standard cross entropy loss
 - Recommendations aligned with input sequence.
 - Limitation - Goal Orientation is not considered.
- Information about goal
 - Probability distribution obtained from BTM output $P(dc \mid goal)$.
 - Deviation of predicted command distribution with the goal's command distribution - KL Divergence

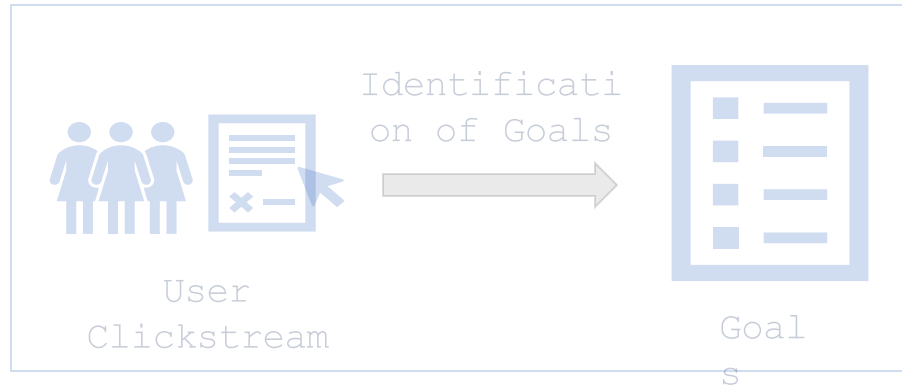
$$\mathcal{L}_{KL}(\theta) = \mathcal{D}_{KL}(P||Q)$$

- Balanced loss

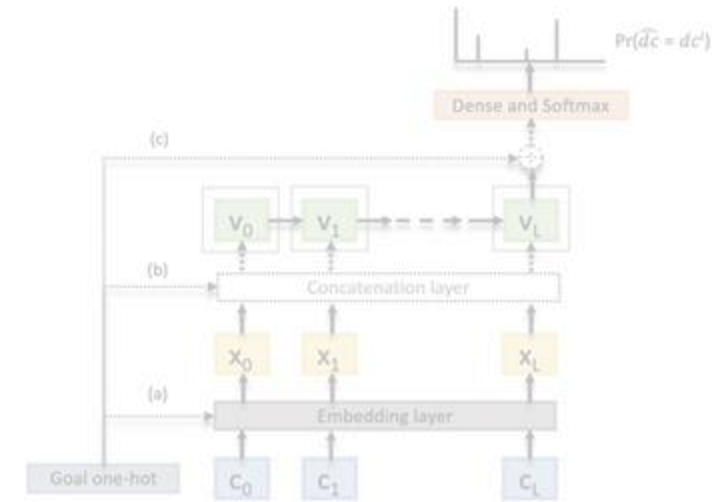
$$\mathcal{L}(\theta) = \alpha \mathcal{L}_{CE}(\theta) + (1 - \alpha) \mathcal{L}_{KL}(\theta)$$



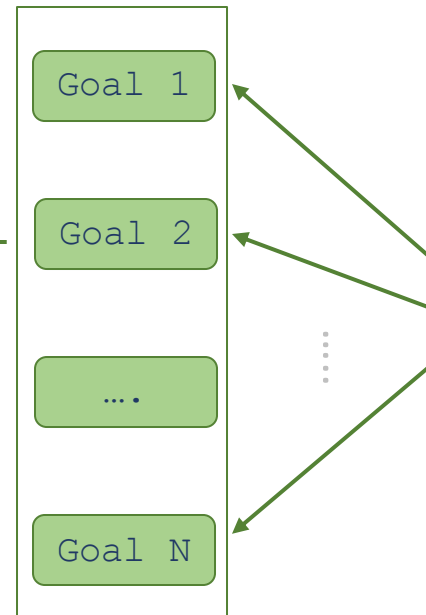
Pre-processed information



Generalised Model Training with Goal Information



Goal specific recommendations



Goal specific models

Goal Specific Fine Tuning

Loss function



Goal-Specific Fine-Tuning

- Component Q (goal-command distribution) of the loss function is different for different goals.

$$\mathcal{L}_{\text{KL}}(\theta) = \mathcal{D}_{\text{KL}}(P||Q)$$

- Pre-train (cross-entropy) + Fine-tune (modified loss function).
- Accurate and goal relevant data command recommendations.
- Superior performance for low resource goals.

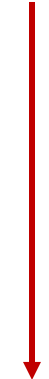


Evaluation

- Standard evaluation metric – test accuracy
 - Goal information influences and improves (13% margin).
 - Recommendations aligned with the selected goal ?
- Goal Orientation Measure (GO-Measure)
 - Goal information in the evaluation.
 - Balanced performance in terms of both accuracy and goal orientation

$$GO_1(goal) = 2 \cdot \frac{accuracy \cdot \bar{P}(dc|goal)}{(accuracy) + \bar{P}(dc|goal)}$$

MODEL	ACCURACY
Top 50 Frequency	0.1633
First-order MM	0.2621
Second-order MM	0.3210
CPT+	0.3444
vanilla (LSTM4REC)	0.5875
ensemble First-order MM	0.3043
ensemble Second-order MM	0.3429
ensemble CPT+	0.4154
ensemble Vanilla (LSTM4REC)	0.6894
GCoRe	0.6839
GComm	0.6970
GAI _n	0.7189

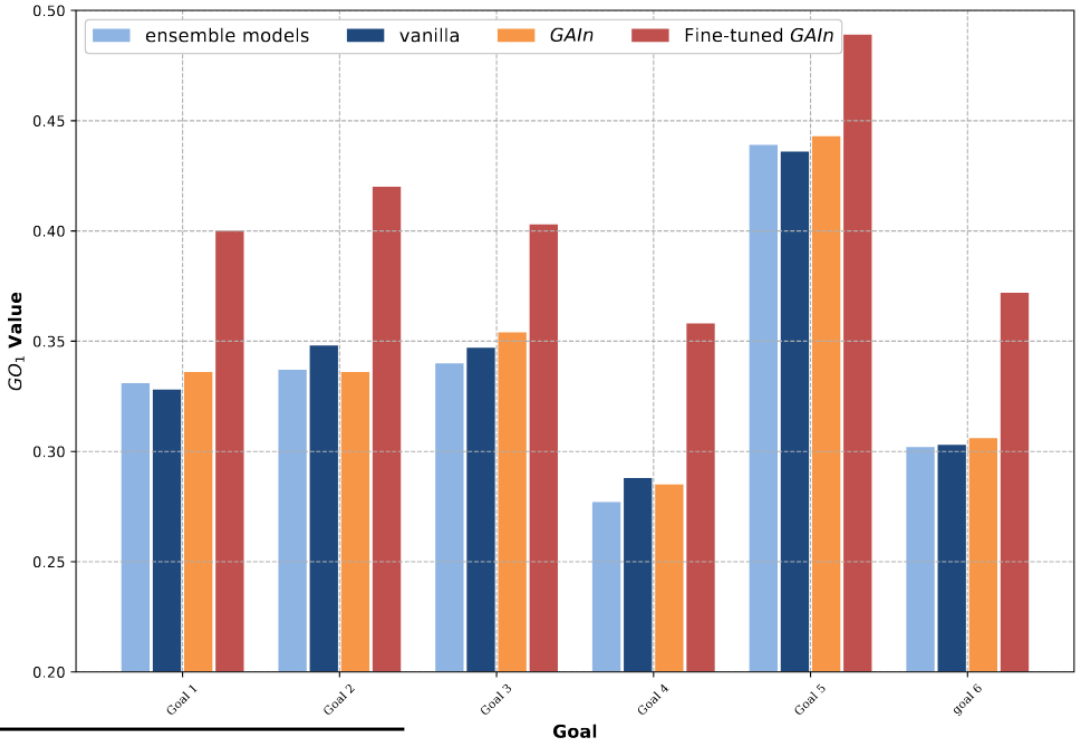


Evaluation



Recommendations should be relevant to the analyst's goal

MODEL	ACCURACY	GO ₁ SCORE	MODE OF GOAL INPUT SPECIFICATION
Top 50 Frequency	0.1633	0.1997	Not Specified
First-order MM	0.2621	0.2251	
Second-order MM	0.3210	0.2322	
CPT+	0.3444	0.2542	
vanilla (LSTM4REC)	0.5875	0.3176	
ensemble First-order MM	0.3043	0.2571	Implicitly through data partition
ensemble Second-order MM	0.3429	0.2679	
ensemble CPT+	0.4154	0.2723	
ensemble Vanilla (LSTM4REC)	0.6894	0.3211	
GCoRe	0.6839	0.3984	Encoded Representation
GComm	0.6970	0.3959	
GAIIn	0.7189	0.3933	



MODEL	ACCURACY	GO ₁ SCORE
Vanilla	0.5563	0.3852
Fine-tuned GCoRe	0.6647	0.4322
Fine-tuned GComm	0.6820	0.4268
Fine-tuned GAIIn	0.7010	0.4811

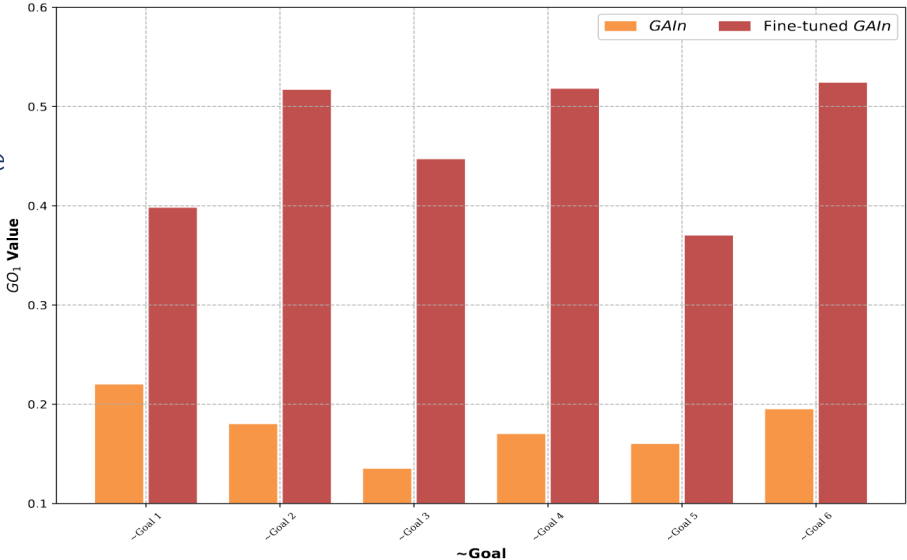


Adversarial Testing

 **Steer the recommendations** whenever the analyst is deviating from the goal

- User might deviate from the specified goal while progressing the session.
- Simulation
 - A model is provided inputs from data distributions different from what it was trained on.
 - A decrease in the accuracy of the models; the

MODEL	ACCURACY	GO ₁ SCORE
ensemble vanilla	0.1525	0.2566
<i>GAI</i> n	0.4919	0.1966
Fine-tuned <i>GAI</i> n	0.2795	0.4823



Concluding...

1. Notion of a goal in data analytics software applications.
2. Incorporating goal information improves recommending data commands to the user.
3. Custom loss function, fine-tuning approach.
4. Novel evaluation method – GO_1 measure.



In Future...

1. Attention mechanisms and transfer learning.
2. Predict the user's goal in real time based on the progress of the session.
3. Handle the problem of a novice user mis-specifying goal - better user experience.



Thank You

Any Questions?



