Goal-driven Command Recommendations for Analysts

Samarth Aggarwal
Rohin Garg
Abhilasha Sancheti
Bhanu Prakash Reddy Guda*
Iftikhar Ahamath Burhanuddin

| IIT Delhi |
| IIT Kanpur |
| Adobe Research |

*Adobe Research
There has been a tremendous growth in the domain of data analysis as the volume of data has increased. Demand for systems to query, analyse and draw inferences with a low latency.
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.

---


Workflow of our approach
Pre-processed information

User Clickstream → Identification of Goals → Goal S

Generalised Model Training with Goal Information

Embedding layer → Concatenation layer → Goal one-hot → Goal Specific Fine Tuning

Goal Oriented Loss Function

Goal Aware Guidance

Recommendations
- Data Command Class 4 Variable
- Data Command Class 5 Variable
- Data Command Class 5 Variable
- Data Command Class 5 Variable
- Data Command Class 5 Variable
- Data Command Class 7 Variable
- Data Command Class 1 Variable

Goal specific recommendations

Goal specific models

Goal 1 → Goal 2 → ...

Goal N
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

Identification of Goals

User Clickstream

Goal

Generalised Model Training with Goal Information

Goal specific recommendations

Goal specific models

Goal OrientedUser Functions
Identification of Goals

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

Pre-processed information

Identification of Goals

User Clickstream

Goal Information

Generalised Model

Training with Goal Information

Goal specific recommendations

Goal specific models
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

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
</tr>
</thead>
</table>

Goal information not provided

Baseline with goal information

Proposed Models
Goal Oriented Loss Function

Pre-processed information

User Clickstream

Identification of Goals

Goal Oriented Training with Goal Information

Goal specific recommendations

Goal specific models

Goal Oriented Loss Function

Generalised Model
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 $P(dc | goal)$.
  - Deviation of predicted command distribution with the goal’s command distribution – KL Divergence
    $$L_{KL}(\theta) = D_{KL}(P||Q)$$
- Balanced loss
  $$L(\theta) = \alpha L_{CE}(\theta) + (1 - \alpha) L_{KL}(\theta)$$
Pre-processed information

User Clickstream

Identification of Goals

Goal Specific models

Goal Specific Fine Tuning

Goal-oriented recommendations

Generalised Model Training with Goal Information

Goal 1

Goal 2

... 

Goal N

Goal-specific recommendations
Goal-Specific Fine-Tuning

- Component $Q\text{\{goal-command distribution\}}$ of the loss function is different for different goals.

$$L_{KL}(\theta) = D_{KL}(P||Q)$$

- 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

\[
GO_1(\text{goal}) = 2 \cdot \frac{\text{accuracy} \cdot P(\text{dc}|\text{goal})}{\text{accuracy} + P(\text{dc}|\text{goal})}
\]

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 50 Frequency</td>
<td>0.1633</td>
</tr>
<tr>
<td>First-order MM</td>
<td>0.2621</td>
</tr>
<tr>
<td>Second-order MM</td>
<td>0.3210</td>
</tr>
<tr>
<td>CPT+</td>
<td>0.3444</td>
</tr>
<tr>
<td>vanilla (lstm4rec)</td>
<td>0.5875</td>
</tr>
<tr>
<td>ensemble First-order MM</td>
<td>0.3043</td>
</tr>
<tr>
<td>ensemble Second-order MM</td>
<td>0.3429</td>
</tr>
<tr>
<td>ensemble CPT+</td>
<td>0.4154</td>
</tr>
<tr>
<td>ensemble Vanilla (lstm4rec)</td>
<td>0.6894</td>
</tr>
</tbody>
</table>

\[GCoRe\] 0.6839  
\[GComm\] 0.6970  
\[GAI\]N 0.7189
Recommendations should be **relevant** to the analyst’s goal
Adversarial Testing

- 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

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>$G_{O1}$ Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>ensemble vanilla</td>
<td>0.1525</td>
<td>0.2566</td>
</tr>
<tr>
<td>$GAIn$</td>
<td>0.4919</td>
<td>0.1966</td>
</tr>
<tr>
<td>Fine-tuned $GAIn$</td>
<td>0.2795</td>
<td>0.4823</td>
</tr>
</tbody>
</table>
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.

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?