Abstract

A session-based recommender system (SBRS) suggests the next item by modeling the dependencies between items in a session. Most of existing SBRSs assume the items inside a session are associated with one (implicit) purpose. However, this may not always be true in reality, and a session may often consist of multiple subsets of items for different purposes (e.g., breakfast and decoration). Specifically, items (e.g., bread and milk) in a subset have strong purpose-specific dependencies whereas items (e.g., bread and vase) from different subsets have much weaker or even no dependencies due to the difference of purposes. Therefore, we propose a mixture-channel model to accommodate the multi-purpose item subsets for more precisely representing a session. To address the shortcomings in existing SBRSs, this model recommends more diverse items to satisfy different purposes. Accordingly, we design effective mixture-channel purpose routing networks (MCPRNs) with a purpose routing network to detect the purposes of each item and assign them into the corresponding channels. Moreover, a purpose-specific recurrent network is devised to model the dependencies between items within each channel for a specific purpose. The experimental results show the superiority of MCPRN over the state-of-the-art methods in terms of both recommendation accuracy and diversity.

1 Introduction

In real world, user requirements keep changing when the corresponding contexts evolve. This context-sensitive demand nature inspires the recent interest on session-based recommender systems (SBRSs) [Wang et al., 2019], which tackle the gaps in the conventional static recommender systems (RSs) that only recommend homogeneous items without considering the demand dynamics across sessions. Accordingly, an SBRS recommends the next item for a session (as a context) to a user who may be interested in by modeling the dependencies between items in the session [Hidasi et al., 2018].

Most of existing SBRSs (e.g., [Twardowski, 2016; Wang et al., 2017; 2018]) make recommendations by forming a session with a single purpose or goal (e.g., shopping for food and beverage products in basket analysis). This often violates the reality that a session may involve multiple types of items where each type corresponds to one purpose. Taking the session illustrated by Row 1 in Figure 1 as an example, Janet first placed bacon into a cart for breakfast while she then was attracted by a piece of lovely rose for decoration and added it into the cart. She further chose eggs and bread for breakfast and a vase for fitting the rose. Finally, she ended this shopping session by picking up a bottle of milk. For this example, existing SBRSs would take a single-channel modeling approach to implicitly associate the items purchased in the session as a homogeneous item sequence, as shown in Row 2 in Figure 1. Obviously, such an approach fails to differentiate the purpose-specific item dependencies (e.g., items for breakfast vs. items for decoration) when the items in a session serve multiple purposes.

The above example reveals significant gaps in two representative types of state-of-the-art SBRSs, namely the single-channel recurrent neural networks (RNN) based SBRSs and attention-based SBRSs. RNN-based models [Hidasi et al., 2018; Quadrana et al., 2017] assume a rigid sequential dependency over any successive items within a session and hence:
(Gap 1) it is easy to generate false dependencies as not all items depend on each other in a multi-purpose session. In addition, due to the memory decay along with the time steps, those contextual items far from the target item will be overwhelmed by the near ones, as a result, (Gap 2) an RNN-based SBRS tends to recommend items biased to the most recent purpose indicated by the nearest items. To reduce the interference of irrelevant items, an attention mechanism has been recently incorporated into shallow networks [Wang et al., 2018], RNN [Li and et al., 2017] or memory networks [Chen et al., 2018] to build SBRSSs. The attention model tends to assign salient weights on a very few significant items [Hu et al., 2018], which makes the purpose of a session dominated by these few items. Consequently, (Gap 3) an attention-based SBRS tends to recommend items to satisfy the dominant purpose while disadvantages others.

This paper addresses the above three gaps by proposing mixture-channel purpose routing networks (MCPRN). MCPRN first automatically detects the possible multiple purposes in a session with a purpose routing network (PRN), and then models a session with a mixture-channel recurrent networks (MCRN) where each channel (i.e., a purpose-specific recurrent network (PSRN) built on purpose-specific recurrent units (PSRUs) models the dependencies between items for a specific purpose to build a purpose-specific embedding. Finally, a multi-purpose context (session) embedding is built by integrating all the channels to rank the candidate items for diverse purpose-sensitive recommendations.

Thanks to the PRN and MCRN, the different purposes in a session are detected and then their item dependencies are respectively modeled in different channels. As a result, the item dependencies are modeled over specific purposes instead of a coarse one, which enables our model to more concentratedly capture purpose-driven user behavior. Consequently, MCPRN leverages the current single-purpose SBRSSs with a more effective and robust mechanism to represent multi-purpose sessions. By retaining and modeling the multiple purposes in a session, MCPRN can recommend diverse items satisfying different purposes, while existing methods tend to recommend items for a single purpose only, as shown in Row 3 and Row 2 in Figure 1 respectively. The main contributions of this work are summarized below:

- We propose multi-purpose modeling to capture user behavior in a session in a more reasonable way to fit real cases. Accordingly, mixture-channel purpose routing networks (MCPRN) are proposed to achieve this.
- In the MCRN, the purpose routing network (PRN) is devised to infer the purposes of each item and route them into specific channels. Moreover, the purpose-specific recurrent unit (PSRU) is the key component to serve as the basic cells of the mixture-channel recurrent networks (MCRNs) w.r.t. each channel.

## 2 Related Work

A variety of SBRSSs including rule/pattern-based RSSs [Yap et al., 2012], neighborhood-based RSSs [Jannach and Ludewig, 2017], Markov chain-based RSSs [Feng and et al., 2015] and factorization machine-based RSSs [Rendle et al., 2010] have been developed. We briefly review two representative state-of-the-art SBRSSs: (1) RNN-based SBRSSs, and (2) attention-based SBRSSs, which are the most relevant ones to our work.

Due to its strength in handling sequential data, RNN is an intuitive choice to capture the complex intra-session dependency in SBRSSs. The first RNN-based SBRSS, GRU4Rec [Hidasi et al., 2016], employed gated recurrent unit (GRU) to capture the long-term dependency within sessions. Later, the performance of GRU4Rec was improved significantly by introducing novel ranking loss functions tailored to RNNs in the recommendation setting [Hidasi et al., 2018]. RNN is easy to generate false dependencies due to the employed rigid order assumption which assumes any adjacent items in a session are highly sequentially dependent. However, this may not be true in most real cases. In addition, RNN usually biases to recent items while missing much information of previous items in a session because of memory decay.

The attention mechanism [Vaswani et al., 2017] has been employed to build more robust SBRSSs by intensifying those relevant and important items in a session context. [Li and et al., 2017] proposed the Neural Attentive Recommendation Machine (NARM), in which a hybrid encoder with an attention mechanism is employed to model the user’s sequential behavior and the user’s main purpose in the current session by differentiating more and less important items. [Wang et al., 2018] designed an attention-based transaction embedding model (ATEM) to build an attentive context embedding over the session context without order assumption. Further, the attention mechanism was used in a memory-augmented neural network for selectively reading out the external memory matrix for next item recommendations [Chen et al., 2018]. However, the attention mechanism attempts to assign larger weights on few significant items while downplaying others, leading to bias to the main purpose indicated by those few items. As a result, attention-based SBRSSs often only cater for the main purpose while ignoring others in a session.

In summary, all the aforementioned SBRSSs are single channel-based, which are effective for one-purpose sessions in which all items within a session are dependent on each other. However, they cannot handle sessions with different types of items for multiple purposes well. Inspired by the great potential of mixture models in handling multiple kinds of relations [Shazeer et al., 2017; Kang et al., 2018], we devise a mixture-channel model for multi-purpose sessions.

## 3 Problem Statement

Given a session $S = \{s_1, ..., s_{|S|}\}$, each session $s = \{v_1, ..., v_{|s|}\} (s \in S)$ consists of a sequence of items that are interacted (e.g., clicked or purchased) sequentially by an anonymous user in one transaction event. Here $|S|$ denotes the number of sessions in $S$ and the subscripts in $s$ indicate the order of item occurrence. All the items occurring in all sessions constitute the universal item set $V = \{v_1, ..., v_{|V|}\}$. For a target item $v_t \in s$, all the items that occurred prior to $v_t$ in $s$ together form the session context (called context for short) of $v_t$ over $s$, represented as $C_{v_t} = \{v_1, ..., v_{t-1}\}$. Each item in $C_{v_t}$ is called a contextual item.

Given a context $C$ with precedent $t - 1$ items, an SBRS
Figure 2: (a) The MCRN model consists of two main modules: Purpose Router and Mixture-Channel Recurrent Networks; (b) The PSRU cell introduces a concentration gate (see the blue dash line square) to model purpose-specific transitions.

is built to recommend the \( t^{th} \) item. Accordingly, MCRN is trained as a probabilistic classifier model that learns to predict a conditional probability distribution \( P(v_t|C) \). Once all the parameters of the model are learned, the MCRN is able to recommend the next item by ranking all the candidate items in terms of their conditional probability over the given context.

4 Mixture-Channel Purpose Routing Networks

The architecture of MCRN is illustrated in Figure 2 (a). MCRN mainly consists of two modules: (1) Purpose Router, and (2) Mixture-Channel Recurrent Networks (MCRNs). The purpose router is employed to route each item to a channel of a specific purpose. Specially, we adopt a soft routing strategy to assign each item embedding to all channels with different purpose weights, so-called mixture channel networks. The soft routing strategy enables to conduct any gradient-based optimization for easing learning. In MCRNs, each channel is equipped with a PSRN built on PSRUs to model the dependencies between items for a specific purpose. The final hidden states \( h_{t-1} \) from different channels are selectively integrated as a multi-purpose context embedding \( v_C \). The embedding of target item \( v_t \) conditional on \( v_C \) is used to calculate the probability of selection as the next item. Next, we present more technical details of these components.

4.1 Purpose Router

Given all the items in a session context, a purpose routing network (PRN) is employed in Purpose Router to extract the purposes of selecting an item \( v_i \) without any human prior knowledge. First, we map each item \( v \) into a \( K \)-dimensional embedding vector \( v \in \mathbb{R}^K \), where \( \mathbb{R}^K \) is the embedding matrix of all items. Then, we input the embedding \( v_i \) of item \( v_i \) into the PRN to identify the purposes of selecting \( v_i \). \( W_p \in \mathbb{R}^{K \times m} \) denotes the purpose filtering parameter of all purposes, where \( m \) is the number of possible purposes (i.e., the number of channels) and it can be determined by cross-validation. Then, the concentration score \( a_{i,j} \) of item \( v_i \) w.r.t. the \( j^{th} \) purpose \( p_j \) is computed as follows:

\[
a_{i,j} = v_i^T W_p[i, j], \quad j \in \{1, \ldots, m\}
\]

where \( W_p[i, j] \) denotes the \( j^{th} \) column of \( W_p \). Further, the normalized concentration weight w.r.t. \( p_j \) can be obtained in terms of the following softmax function:

\[
g_{i,j} = \frac{\exp(a_{i,j}/\tau)}{\sum_{h=1}^{m} \exp(a_{i,h}/\tau)}
\]

where \( \tau \) is a temperature parameter to tune. For high temperatures (\( \tau \to \infty \)), all purposes have nearly the same probability. For a low temperature (\( \tau \to 0^+ \)), it tends to concentrate on a single purpose (nearly a hard routing). In experiments, we use \( \tau = 0.1 \) which can produce the best performance.

4.2 Mixture-Channel Recurrent Networks

Then, the item embedding and its purpose concentration weight are input into each channel of MCRN as shown by the green and purple arrow lines respectively in Figure 2 (a). Each channel is modeled by a PSRN composed of \( t-1 \) PSRU cells to model the purpose-specific sequential dependency over a sequence of \( t-1 \) items. The working mechanism of each channel is similar to a normal RNN whereas the cells are equipped with our designed PSRU cells in PSRNs. The PSRU cells in one channel are identical while cells in different channels share an identical structure but different parameters. We present PSRU in detail in the following subsection.

Purpose-Specific Recurrent Units

RNN cells like long short-term memory (LSTM) or GRU do not consider the degree of membership of items in a sequence. In MCRNs, each channel contains all items in a session but their purpose concentration weights, i.e. the degree of membership, on this purpose are different. Hence, LSTM and GRU are not ready for modeling such purpose-specific dependencies in a channel given the concentration weights.

As a result, we designed the purpose-specific recurrent unit (PSRU) to serve as the cell for each PSRN. Different from GRU or its variants, which usually compute a gate value by only using the current input and the last hidden state without...
any additional information, PSRU introduces a concentration gate to selectively integrate the item information in the transitions w.r.t. a specific purpose.

The structure of a PSRU cell is given in Figure 2 (b). Compared to the traditional GRU cell, an additional gate, namely concentration gate (see the blue dash line square), is added to decide to what extent the current state should be involved into the purpose-specific transitions according to the purpose concentration weight $g_{i,j}$. The PSRU cell at the $i$th time step in channel $c_j$ takes the last hidden state $h_{i-1}$, the embedding of the current item $v_i$ and the concentration weight $g_{i,j}$ (here subscript $j$ is omitted from $g_i$ when the channel $c_j$ is given) of $v_i$ as the input, and outputs the candidate hidden state $\tilde{h}_i$ of current time step. More specifically, the candidate hidden state $\tilde{h}_i$ can be computed as follows, with the reset gate vector $r_i$ and the update gate vector $z_i$:

$$r_i = \sigma_s(W_r[h_{i-1}, v_i])$$
$$z_i = \sigma_s(W_z[h_{i-1}, v_i])$$
$$\tilde{h}_i = \sigma_t(W_h[r_i * h_{i-1}, v_i])$$

where $\sigma_s$ and $\sigma_t$ are activation functions and are specified as sigmoid and tanh respectively. Further, we obtain the concentration gate vector $u_i$ as below:

$$u_i = \sigma_e(g_{i,j} * z_i)$$

where the Delta function $\delta = 1$ if $g_i \geq 0$ and 0 otherwise. $e$ is a threshold parameter to eliminate noisy items in modeling the transitions of a session. That is, it will bypass the item with a small concentration weight less than $e$. We empirically set $e = 0.01$ in our experiment.

As a result, the hidden state of current step, $h_i$, can be determined by $u_i$ in terms of previous state $h_{i-1}$ and current candidate state $\tilde{h}_i$:

$$h_i = (1 - u_i) * h_{i-1} + u_i * \tilde{h}_i$$

For example, if item $v_i$ majorly concentrate on a single purpose $p_1$, i.e., $g_{i,1} \approx 1$ while $g_{i,j} \approx 0$ ($j \neq 1$). Due to $g_{i,j} < e$ for $j \neq 1$, item $v_i$ will be ignored in the transitions for all other channels except Channel 1, i.e. $h_i = h_{i-1}$.

4.3 The Probability of Selection

Once the session context $C$ has been input into MCRPN, the final hidden states $h_{i-1}$ of all $m$ channels are integrated together to build a multi-purpose context embedding $v_C$:

$$v_C = \sum_{j=1}^{m} \hat{g}_{i,j} * h_{i-1}$$

where $\hat{g}_{i,j}$ is the concentration weight of target item $v_i$ w.r.t. purpose $p_j$, $\hat{g}_{i,j}$ weights target purpose to construct the context embedding $v_C$ that probably output the target item $v_i$.

Then, we feed the context embedding $v_C$ together with the embedding of the candidate item into the output layer for the target item prediction. Specifically, a score that quantifies the relevance of the target item $v_i$ w.r.t. the given context $C$ is computed using the inner product to capture the interaction between them.

$$\delta_i(C) = \hat{v}_i^T v_C$$

### Algorithm 1: MCRPN parameter learning procedure

1. $B \leftarrow$ Get mini-batch from all context-target item pairs
2. $N \leftarrow$ Sampling $n$ negative items, $S^-$, for each target item $v_t \in B$
3. Compute mini-batch loss using Eq. 11:

$$L_B \leftarrow \frac{1}{|B|} \sum_{i} \log \frac{1}{q(v_i | C)} + \sum_{s 

Finally, the conditional probability $q(\hat{v}_i | C; \Theta)$ to select the item $\hat{v}_i$ is obtained in terms of the score $\delta_i(C)$:

$$q(\hat{v}_i | C; \Theta) = \frac{1}{1 + e^{-\delta_i(C)}}$$

where $\Theta$ is the set of model parameters that need to be learned over all sessions.

4.4 Optimization and Training

We train our model by using a mini-batch gradient descent on the cross-entropy loss. Given the conditional probability $q$ of selection, the loss function is:

$$L(s^+, S^-) = -[\log(q_{s^+}) + \sum_{s \in S^-} \log(1 - q_s)]$$

Given a session context $C$, we build a contrastive pair by taking the true next item $v_t$ as the positive sample $s^+$ and then randomly sample $n$ items from the item set $I \setminus v_t$ as the negative sample set $S^-$. The loss from the positive sample and each negative sample are $\log(q_{s^+})$ and $\log(1 - q_s)$ respectively. They together form the loss of a contrastive pair $(s^+, S^-)$. The model parameters $\Theta$ are learned by minimizing $L(s^+, S^-)$. Specifically, negative sampling [Goldberg and Levy, 2014] is employed for efficient training.

Our model is implemented using Tensorflow. Due to the limited space, only a brief scheme of the learning procedure on a mini-batch is listed in Algorithm 1, where $\Gamma_{adam}$ denotes Adam [Kingma and Ba, 2015]-based gradient descent optimizer and the batch size is set to 50.

5 Experiments and Evaluation

5.1 Data Preparation

Three real-world transaction datasets are used for experiments: (1) Yoochoose-buys\footnote{https://2015.recsyschallenge.com/challenge.html} released by RecSys Challenge 2015, which records the purchased items in each session on Yoochoose.com; (2) Tmall\footnote{https://tianchi.aliyun.com/dataset/dataDetail?dataId=42} released by IJCAI-15 competition, which records the purchased items in each transaction on Tmall online shopping platform; and (3) TaFeng\footnote{https://www.kaggle.com/chiranjivdas09/ta-feng-grocery-dataset} released on Kaggle, which contains the transaction data of a Chinese grocery store.

First, a set of sessions is extracted from each original dataset by putting all items in one transaction together to form a session. Those sessions containing less than three items are removed since at least two items should be used as the context...
to form a multi-purpose session context and one as the target item. Second, we do three training-test splits on the session set by randomly selecting 20%, 30%, 40% of the sessions from the last 30 days respectively for testing. Our method achieved similar performance on all the three splits and stably outperformed all the baselines and thus only the results from the 30% split is shown. Finally, to build the training-test instances of format \((C, v_t)\), for one session \(s\), each time one item from the third to the last item in \(s\) is picked up as the target item \(v_t\) and all those items before \(v_t\) constitute its corresponding context \(C\). Subsequently, for a session \(s\) containing \(|s| (|s| \geq 3)\) items, \(|s| - 2\) instances are built in total. The characteristics of experimental datasets are shown in Table 2.

### 5.2 Experimental Settings

Three commonly used metrics: Mean Reciprocal Rank (MRR), normalized Discounted Cumulative Gain (nDCG) and Recall, are applied to evaluate the recommendation accuracy [Chen et al., 2018; Wang et al., 2018], together with a newly-designed diversity measure to evaluate the diversity.

#### Comparison Methods

To demonstrate the efficacy of mixture channels for multi-purpose modeling, we implemented two versions of our method: (1) full model of MCPRN proposed in this work; and (2) single channel network (MCPRN-S) with only one channel retained from MCPRN. The following representative state-of-the-art SBRSs built on various frameworks including RNN, attention model, memory network, convolutional neural network (CNN) and shallow network, are selected to be the baselines:

- **iGRU4Rec-BPR**: An improved version of a typical RNN-based SBRS which models a session using RNNs built on GRU. It takes Bayesian Personalized Ranking (BPR) as the loss function [Hidasi et al., 2018].

- **iGRU4Rec-CE**: Similar to iGRU4Rec-BPR except that it replaces the loss function BPR with cross entropy [Hidasi et al., 2018].

- **NARM**: A hybrid encoder with an attention mechanism to users’ sequential behavior and capture users’ main purpose in a session [Li et al., 2017].

- **MANN**: A memory-augmented neural network which employs an attention model to read out historical information explicitly stored in the external memory matrix [Chen et al., 2018].

- **Caser**: A convolutional sequence embedding model which embeds a sequence of items into an “image” and then learns sequential patterns as local features using convolutional filters [Tang and Wang, 2018].

- **ATEM**: A shallow and wide network with the attention mechanism incorporated to learn an attentive embedding of a session context [Wang et al., 2018].

#### Parameter Settings

We initialize all the baseline models with the parameter settings in the corresponding papers and then tune them on our datasets for best performance for a fair comparison. The sizes of item embeddings and hidden states in PSRU are set to 128. The number of channels \(m\) is set to 3 by tuning on the validation data. The initial learning rate for Adam is set to 0.001.

### 5.3 Recommendation Accuracy Evaluation

Extensive experiments are conducted to answer the following questions:

- **Q1**: How does our approach compare to the state-of-the-art SBRSs in terms of recommendation accuracy?

- **Q2**: How do the mixture-channel networks compare to the single channel network?

#### Result 1: Comparison with Baselines w.r.t. Accuracy

To answer question Q1, we compare the recommendation accuracy of our method MCPRN and those of the six state-of-the-art SBRSs. Table 1 reports the MRR and nDCG. The first two methods are built on GRU-based RNN, which can easily generate false dependencies due to the overly strong sequentially dependent-based assumption and bias to the most recent purpose. Therefore, they do not perform well on multipurpose session data. NARM and MANN achieve better performance by incorporating the attention mechanism into an RNN-based and memory network-based SBRS respectively to emphasize the relevant information and reduce noise. But they are easy to bias to the main purpose in a session due to the attention weighting mechanism. Caser does not perform

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Yoochoose-buy</th>
<th>Tmall</th>
<th>Tafeng</th>
</tr>
</thead>
<tbody>
<tr>
<td>#Sessions</td>
<td>83,928</td>
<td>144,936</td>
<td>19,538</td>
</tr>
<tr>
<td>#Items</td>
<td>7,428</td>
<td>27,863</td>
<td>5,263</td>
</tr>
<tr>
<td>Avg. session length</td>
<td>5.13</td>
<td>4.09</td>
<td>7.41</td>
</tr>
<tr>
<td>#Item category</td>
<td>n.a</td>
<td>786</td>
<td>793</td>
</tr>
</tbody>
</table>

Table 2: Statistics of experimental datasets

**Table 1: Recommendation accuracy on three datasets**

<table>
<thead>
<tr>
<th></th>
<th>Yoochoose-buy</th>
<th>Tmall</th>
<th>Tafeng</th>
</tr>
</thead>
<tbody>
<tr>
<td>iGRU4Rec-BPR</td>
<td>0.1150</td>
<td>0.1271</td>
<td>0.1314</td>
</tr>
<tr>
<td>iGRU4Rec-CE</td>
<td>0.1538</td>
<td>0.1715</td>
<td>0.1776</td>
</tr>
<tr>
<td>NARM</td>
<td>0.1514</td>
<td>0.1734</td>
<td>0.1768</td>
</tr>
<tr>
<td>MANN</td>
<td>0.1653</td>
<td>0.1893</td>
<td>0.1986</td>
</tr>
<tr>
<td>Caser</td>
<td>0.1604</td>
<td>0.1825</td>
<td>0.1918</td>
</tr>
<tr>
<td>ATEM</td>
<td>0.1212</td>
<td>0.1324</td>
<td>0.1768</td>
</tr>
<tr>
<td>MCPRN-S</td>
<td>0.1680</td>
<td>0.1861</td>
<td>0.1859</td>
</tr>
<tr>
<td>MCPRN</td>
<td>0.2191</td>
<td>0.2332</td>
<td>0.2547</td>
</tr>
</tbody>
</table>

| Improvement (%) | 32.49 | 21.46 | 26.08 | 1.67 |

Table 1: Recommendation accuracy on three datasets
Recall on Yoochoose-buy
Recall@1 Recall@5
0
0.1
0.2
0.3
0.4
0.5
iGRU4Rec-BPR
iGRU4Rec-CE
NARM
MANN
Caser
ATEM
MCPRN-S
MCPRN

Figure 3: Recalls of MCPRN and other compared methods

well either as it is hard for the pooling operation in CNN to capture the long-range dependencies. ATEM usually biases to the main purpose as other attention-based SBRSs do. By modeling each purpose independently with a recurrent channel, our MCPRN not only treats each purpose equally but also keeps the sequential dependencies within each purpose. As a result, MCPRN achieves the best performance on all datasets. Particularly, MCPRN demonstrates over 20% improvement over the best existing method in terms of MRR@5, MRR@20 and nDCG@5 on the last two datasets (cf. Table 1). The recall (cf. Figure 3) on two datasets (due to the limited space) shows MCPRN leads the baselines with a clear margin.

Result 2: Mixture Channels vs. Single Channel
To demonstrate the efficacy of mixture-channel structure, we compared MCPRN with MCPRN-S. It is clear that MCPRN achieves much higher accuracy than MCPRN-S as shown in Table 1 and Figure 3. The MRR, nDCG and Recall@5 of MCPRN are at least 20% higher than that of MCPRN-S, proving the superiority of the mixture-channel architecture.

5.4 Recommendation Diversity Evaluation
Diversity evaluation has been introduced to make up the limitation of accuracy [Hu et al., 2017; Yao et al., 2019b]. Experiments are conducted to answer the following question:

• Q3: How does our approach compare to the state-of-the-art SBRSs in terms of recommendation diversity?

Intuitively, the more disperse distribution over the item categories of the recommendations means the larger diversity can be produced by the model. As a result, the category distribution of recommendations can be measured by the entropy to quantify the diversity, according to the information theory [Zhang et al., 2016; Yao et al., 2019a].

\[
Diversity@k = - \sum_{i=1}^{k} Pr_i \log_2 Pr_i
\]  

where \( k \) is the number of top-rank items in the recommendation list while \( Pr_i \) is the probability of the category to which item \( v_i \) belongs and is estimated by the frequency.

Result 3: Comparison with Baselines w.r.t. Diversity
Figure 4 shows that MCPRN achieves higher diversity than compared methods, especially on the Tafeng dataset where MCPRN achieves up to 13.01% and 11.61% improvement w.r.t. Diversity@5 and Diversity@10 respectively compared to the best baseline method. The reasons are that the baseline methods usually recommend items that only satisfy either the recent purpose (e.g., RNN-based SBRSs) or the main purpose (e.g., attention-based SBRSs) due to their single-channel-based design. In contrast, MCPRN easily diversifies the recommendation list by modeling multiple purposes in a parallel way. Diversity evaluation is not applicable to Yoochoose-buy as its category information is unavailable (cf. Table 2).

5.5 Purpose Concentration Visualization
To get a deep insight into how the items for different purposes in a session are detected and accordingly routed into different channels in MCPRN, we visualize the concentration weights (cf. Eq. 2) of two sessions randomly sampled from the test set of Yoochoose-buy dataset in Figure 5.

Two observations can be made from Figure 5: (1) Items in one session are usually for multiple purposes and thus are put into different channels, as illustrated by the various distributions of different items on all the three channels; (2) The true target item \( v_t \) may have purposes different from either the most recent purpose or the main purpose in a session.

6 Conclusions
To effectively detect and model the multiple purposes embedded within a session and thus to recommend corresponding diverse items to satisfy those different purposes, which cannot be addressed by existing session-based recommender systems, in this paper, we have proposed mixture-channel purpose routing networks (MCPRN). MCPRN harness a purpose routing network to detect the possible purposes of each item in a session and multiple channels to model the item dependencies within different purposes independently. Empirical evaluations on the real-world transaction data show the superiority of MCPRN in addressing the gaps in the state-of-the-art methods.

Acknowledgements
This work was partially supported by Australian Research Council Discovery Project DP180102378.
References


