Yi HAN, Linbo QIAO, Jianming ZHENG, Hefeng WU, Dongsheng LI, Xiangke LIAO, 2021. A survey of script learning. *Frontiers of Information Technology & Electronic Engineering,* 22(3):341-373. https://doi.org/10.1631/FITEE.2000347

A survey of script learning

Key words: Script learning; Natural language processing; Commonsense knowledge modeling; Event reasoning

Corresponding author: Linbo QIAO
E-mail: qiao.linbo@nudt.edu.cn
ORCID: https://orcid.org/0000-0002-8285-2738

Motivation

1. To understand natural language, machines need to comprehend not only literal meaning but also commonsense knowledge about the real world. Script is the structured knowledge representation of prototypical real-life event sequences, which is critical for humans to remember and comprehend different scenarios. Machines may also gain benefits by learning such knowledge.

2. Script learning is a promising research direction, in which a trained script learning system can process narrative texts to capture the commonsense knowledge inside it and draw inferences.

3. However, there are currently no survey articles on script learning, so we are providing this comprehensive survey to deeply investigate the standard framework and the major research topics on script learning.

Contributions

1. We are the first to provide a comprehensive survey that deeply investigates the standard framework and major research topics in script learning.

2. We thoroughly summarize the development process and the technique taxonomy of the script learning system.

3. We carefully analyze and compare the most representative works on script learning and discuss some promising research directions.

Definition of script

A **script** is a structural knowledge representation that captures the relationships between prototypical event sequences and their participants in a given scenario (Schank and Abelson, 1977).

In other words, the script is a sequence of prototypical events organized in temporal order. These events contain stereotypical human activities, such as eating in a restaurant, cooking dinner, and making coffee.

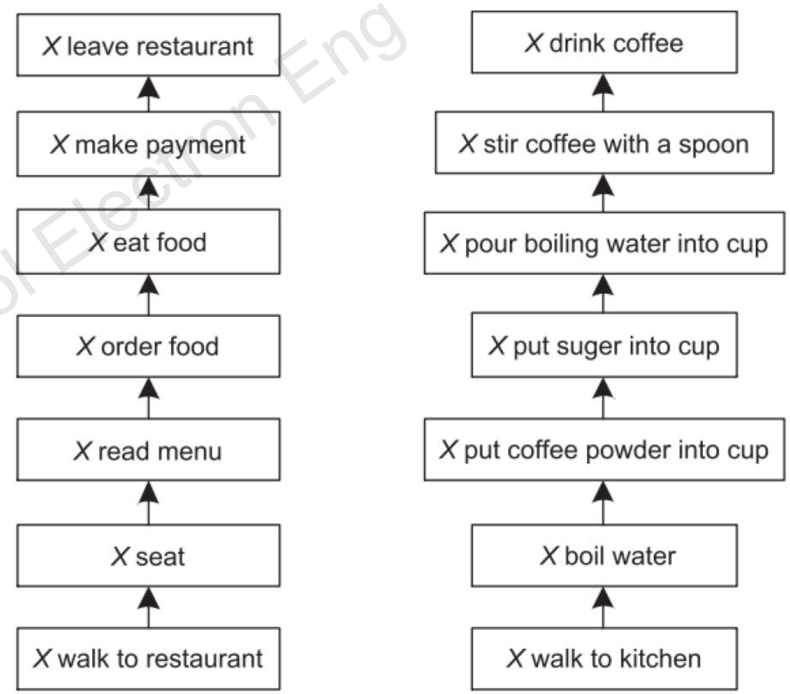


Fig. 1 Scripts for visiting a restaurant and making coffee. Each script includes a sequence of events for a prototypical scenario

Objective of script learning

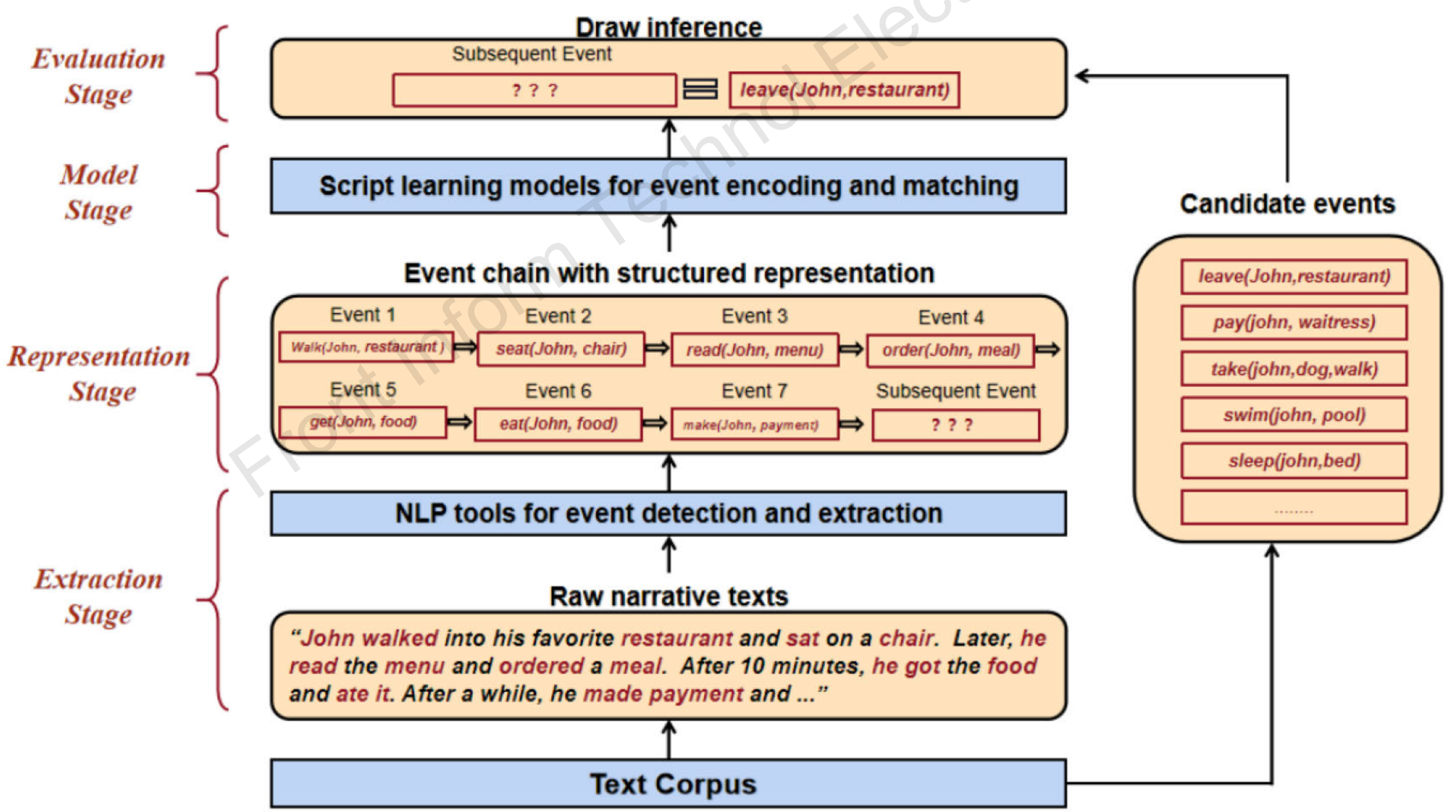
The objective of script learning is encoding the commonsense knowledge contained in prototypical events (also called "**script knowledge**") in a machine readable format.

Then the machine can use the format to draw event-related inferences, such as inferring missing events, predicting subsequent events, determining the order of the event chain, and distinguishing similar events.

Script knowledge also plays a crucial role in a wide range of AI tasks, especially many natural language processing (NLP) tasks, such as event prediction, event extraction, understanding ambiguity resolution discourse, intention recognition, question answering, and coreference resolution.

Framework of script learning

According to the standard script learning settings, its framework can be roughly divided into four stages: **extraction stage**, **representation stage**, **model stage**, and **evaluation stage**.



Event representations

The main aim of event representation is to represent the event with an appropriate structure, which can include essential components of the event and capture the implicit commonsense knowledge hiding behind the text description.

Event representation is very important for script learning, because it specifies what we mean by an "event," and because it is the basic operational element of the subsequent script modeling process.

Protagonist representation

Chambers and Jurafsky (2008) first represented script events as

<verb, dependency>,

where **verb** is the **predicate verb** describing the event, **dependency** is the **grammatical dependency relations** between the **verb** and the **protagonist** such as “subject” and “object.”

“Jessie killed a man. She ran way and got arrested by the police in the street.”

Jessie(protagonist)

(kill, subject), (run, subject), (arrest, object)

Relational representation

Balasubramanian et al. (2013) addressed those flaws using the following relation triple:

<Arg1, Relation, Arg2>,

where **Arg1** and **Arg2** are the subject and object in the event, respectively. This representation aids in keeping coherence between the subject and object.

"He cited a new study that was released by UCLA in 2008."

(He, cited, a new study)

(A new study, was released by, UCLA)

(A new study, was released in, 2008)

Multi-argument representation

Pichotta and Mooney (2014) proposed multi-argument representation:

$v(e_s, e_o, e_p)$

"Mary emailed Jim and he responded to her immediately"

mail(Mary, Jim, -) respond(Jim, Mary, -)

Script learning models

The script learning model is the core component of the script learning system, and undertakes the main tasks of encoding and matching events. Generally, models take the representations of the structured events as the input and take event-matching results as the output.

There are various ways to classify the existing script learning models. Based on modeling methods, they can be roughly divided into three categories: **rule-based models**, **count-based models**, and **deep learning based models**.

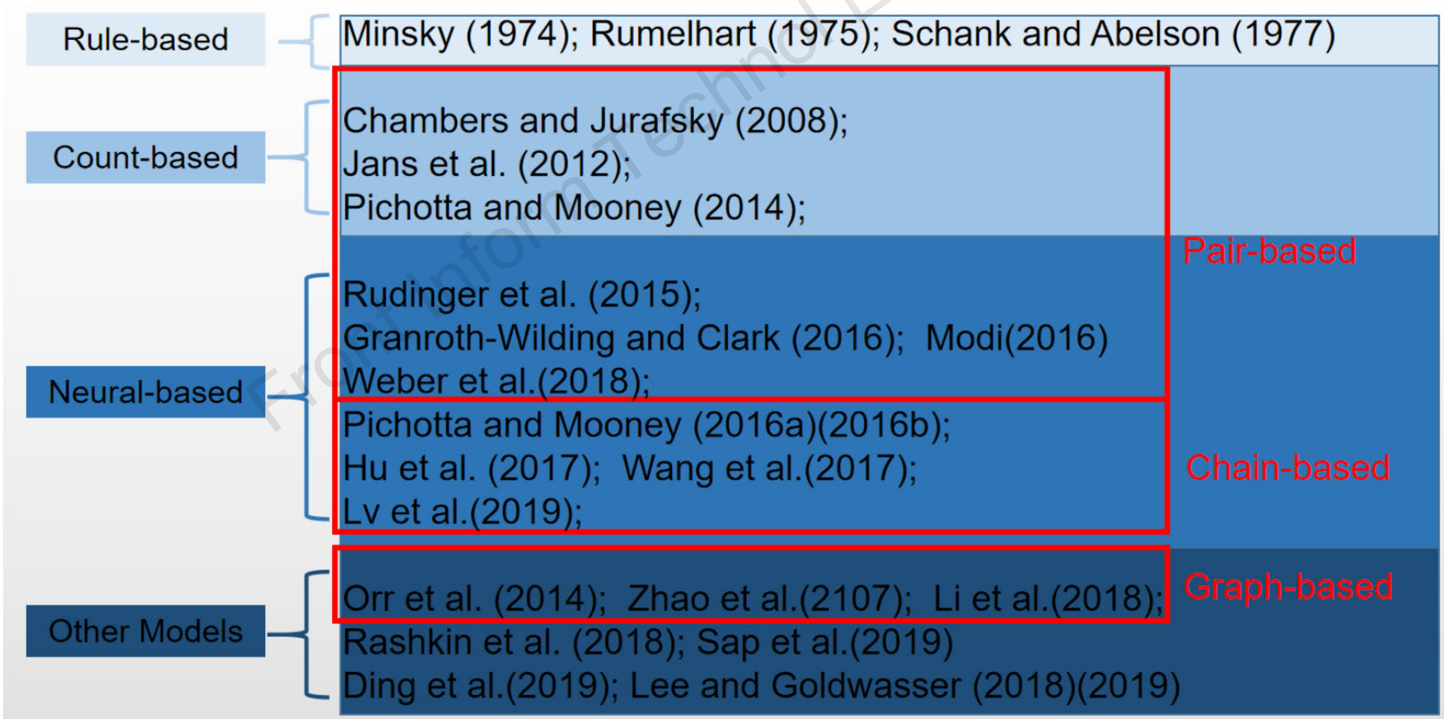
From the perspective of different structures used to handle scripts, they can also be divided into three other categories: **event pair based models, event chain based models,** and **event graph based models.**

Script learning models

- **Rule-based models** use complicated handcrafted rules to model script knowledge of specific domains.
- **Count-based models** use statistical counting approaches to automatically learn broad-domain script knowledge from a large text corpus.
- **Deep learning based models** introduce the neural network and embedding to capture richer script knowledge and overcome the sparsity issues.
- **Pair-based models** focus on calculating associations between pairs of events by count- or vector-based methods.

Script learning models

- **Chain-based models** capture the order information and long-term context information of the full event chain by RNN-based approaches.
- **Graph-based models** extract script knowledge by constructing graph structures that can express denser and broader connections among events.



Evaluation tasks

Essentially, the aim of the script learning system is to encode script knowledge and use it to draw reasonable inferences, so in the final process of script learning, some evaluation approaches are needed to test whether the model can effectively encode the script knowledge and exploit it.

The evaluation approaches can be employed to measure the capabilities of the script learning model by experimentally testing its performance on the task.

Evaluation tasks: NC

Chambers and Jurafsky (2008):

Narrative Cloze (NC)

(walk, subject)→(sit, subject)→(?)→(serve, object)

→(eat, subject)→(leave, subject)

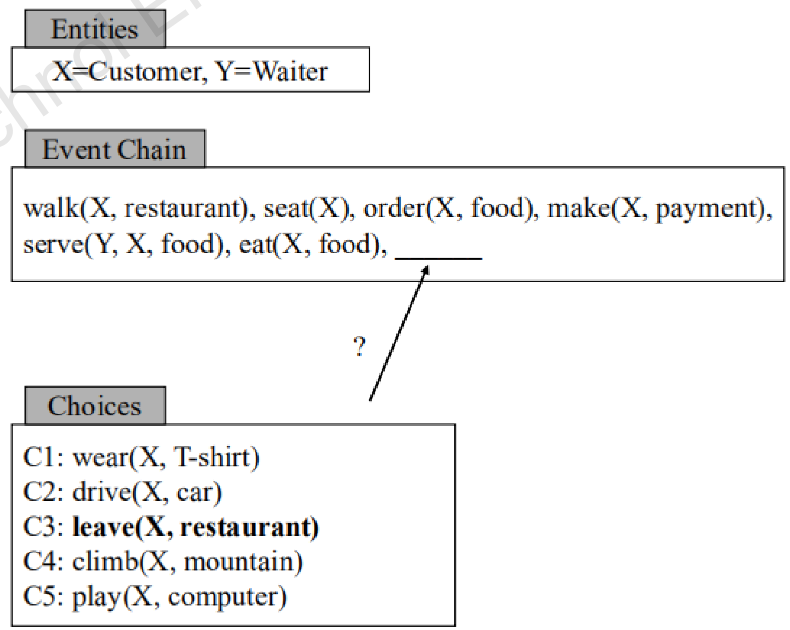
The models need to predict the missing event (?) using the information of the remaining events.

Evaluation tasks: MCNC

Granroth-Wilding and Clark (2016):

Multiple Choice Narrative Cloze (MCNC)

A series of contextual events are given and the model should choose the most likely next event from a set of optional candidates.



Evaluation tasks: MCNC & MCNE

Lee and Goldwasser (2018):

Multiple Choice Narrative Sequences (MCNS)

Multiple Choice Narrative Explanation (MCNE)

MCNS: ***walk(X, restaurant)***→___, →___, →___, →___

MCNE: ***walk(X, restaurant)***→___, →___, →___, → ***leave(X, restaurant)***

Evaluation tasks: ANC & SCT

Modi (2016): Adversarial Narrative Cloze (ANC)

Mostafazadeh et al. (2016): Story Cloze Test (SCT)

ANC:

Correct chain: (walk, subject)→(sit, subject)→(eat, subject)→(leave, subject)

Incorrect chain: (walk, subject)→(sit, subject)→(swim, subject)→(leave, subject)

SCT:

Context	Right Ending	Wrong Ending
Tom and Sheryl have been together for two years. One day, they went to a carnival together. He won her several stuffed bears, and bought her funnel cakes. When they reached the Ferris wheel, he got down on one knee.	Tom asked Sheryl to marry him.	He wiped mud off of his boot.
Karen was assigned a roommate her first year of college. Her roommate asked her to go to a nearby city for a concert. Karen agreed happily. The show was absolutely exhilarating.	Karen became good friends with her roommate.	Karen hated her roommate.
Jim got his first credit card in college. He didn't have a job so he bought everything on his card. After he graduated he amounted a $10,000 debt. Jim realized that he was foolish to spend so much money.	Jim decided to devise a plan for repayment.	Jim decided to open another credit card.

Future directions

- Establishing a standard corpus or evaluation system
- Constructing the event graph
- Using pre-trained language models
- Learning script by reinforcement learning
- Injecting external fine-grained knowledge
- Building an interpretable system

Summary

This survey tried to provide a comprehensive review and introduce some representative script learning works. We first briefly introduced some basic script learning knowledge, including its definition, aim, significance, process, focus, and development timeline. Then we discussed in detail three main research topics: event representations, script learning models, and evaluation approaches. For each one, we introduced some typical and representative works, tried to compare their advantages and disadvantages, and summarized their development process. In the end, we discussed some possible directions for future research.

Yi HAN, first author of this invited paper, received his BS and MS degrees from the National University of Defense Technology (NUDT), Changsha, China in 2016 and 2018, respectively, and is currently a PhD candidate at the College of Computer Science, NUDT. His research interests include natural language processing, event extraction, and few-shot learning.

Linbo QIAO, corresponding author of this invited paper, received his BS, MS, and PhD degrees in computer science and technology from NUDT, Changsha, China in 2010, 2012, and 2017, respectively. Now, he is an assistant research fellow at the National Lab for Parallel and Distributed Processing, NUDT. He worked as a research assistant at the Chinese University of Hong Kong from May to Oct. 2014. His research interests include structured sparse learning, online and distributed optimization, and deep learning for graph and graphical models.

Jianming ZHENG received the BS and MS degrees from NUDT, China in 2016 and 2018, respectively, and is currently a PhD candidate at the School of System Engineering, NUDT. His research interests include semantics representation, few-shot learning and its applications in information retrieval. He has several papers published in SIGIR, WWW, COLING, IPM, Cognitive Computation, etc.

Hefeng WU received the BS degree in Computer Science and Technology and the PhD degree in Computer Application Technology from Sun Yat-sen University, China in 2008 and 2013, respectively. He is currently a full research scientist with the School of Data and Computer Science, Sun Yat-sen University. His research interests include computer vision, multimedia, and machine learning. He has served as a reviewer for many academic journals and conferences, including TIP, TCYB, TSMC, TCSVT, PR, CVPR, ICCV, NeurIPS, and ICML.

Dongsheng LI received his BS degree (with honors) and PhD degree (with honors) in computer science from the College of Computer Science, NUDT, Changsha, China in 1999 and 2005, respectively. He was awarded the prize of National Excellent Doctoral Dissertation by the Ministry of Education of China in 2008, and the National Science Fund for Distinguished Young Scholars in 2020. He is now a full professor at the National Lab for Parallel and Distributed Processing, NUDT. He is a corresponding expert of Frontiers of Information Technology & Electronic Engineering. His research interests include parallel and distributed computing, cloud computing, and large-scale data management.

Xiangke LIAO received his BS degree from the Department of Computer Science and Technology, Tsinghua University, Beijing, China in 1985, and his MS degree from NUDT, Changsha, China in 1988. He is currently a full professor of NUDT, and an academician of the Chinese Academy of Engineering. His research interests include parallel and distributed computing, high-performance computer systems, operating systems, cloud computing, and networked embedded systems.