
Learnersourcing Subgoal Labeling to Support Learning from How-to Videos

Juho Kim

MIT CSAIL
Cambridge, MA USA
juhokim@mit.edu

Robert C. Miller

MIT CSAIL
Cambridge, MA USA
rcm@mit.edu

Krzysztof Z. Gajos

Harvard School of Engineering
and Applied Sciences
Cambridge, MA USA
kgajos@eecs.harvard.edu

Abstract

Subgoal labeling [1] is a technique known to support learning new knowledge by clustering a group of steps into a higher-level conceptual unit. It has been shown to improve learning by helping learners to form the right mental model. While many learners view video tutorials nowadays, subgoal labels are often not available unless manually provided at production time. This work addresses the challenge of collecting and presenting subgoal labels to a large number of video tutorials. We introduce a mixed-initiative approach to collect subgoal labels in a scalable and efficient manner. The key component of this method is *learnersourcing*, which channels learners' activities using the video interface into useful input to the system. The presented method will contribute to the broader availability of subgoal labels in how-to videos.

Author Keywords

Video tutorials; how-to videos; subgoal labeling; learnersourcing; mixed-initiative; learning support.

ACM Classification Keywords

H.5.1 [Information Interfaces and Presentation]:
Multimedia Information Systems.

Copyright is held by the author/owner(s).

CHI 2013 Extended Abstracts, April 27–May 2, 2013, Paris, France.

ACM 978-1-4503-1952-2/13/04.

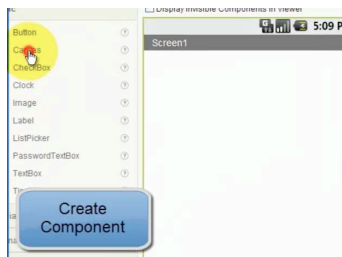


Figure 1. An example of a subgoal label presented as a video callout (“create component”). Image retrieved from [6].

Introduction

Learning how-to skills such as cooking, image manipulation, and makeup requires mastering procedural steps involved in the workflow. Mental models help this process by allowing learners to integrate simple steps into a complex step and to focus on high level functions rather than low level details [9]. The subgoal learning model [1] states that learners can develop better mental models by constructing a hierarchical structure of the steps by clustering them into subgoals. Subgoals enable learners to more flexibly adjust to a new problem when it differs from the original example they encountered before. Subgoals also let learners avoid optimizing to the surface-level instructions but encourage understanding a meaningful hierarchical structure. *Subgoal labels* refer to cues on grouping steps in an example into a meaningful unit, which aids in learning subgoals. An example is shown in Figure 1, where a video callout is added to show that the current subgoal is to “*create component*”, which allows learners to not just focus on the surface-level step itself but also associate that step in terms of a higher-level goal of creating a component.

This research addresses the question of applying the subgoal learning model to how-to videos, a popular medium today's learners turn to for learning how-to skills. Video tutorials are effective because they include a live demonstration, which makes following along easy. They also capture richer context than static text+image tutorials, making them an effective medium for learning procedural skills [3]. However, learners using video struggle in building the right mental model required to solve the problem at hand. First, steps are not often indexed, causing extra cognitive load in constructing higher-level steps. Second, a linear flow

often omits the hierarchical structure of a problem, causing learners to memorize a sequence of steps [7] rather than to build the right mental model.

A recent study has shown that subgoal-labeled video instructions led to better knowledge transfer and longer retention of the material [6]. Students taught in subgoals approached new problems using their developed subgoals. Generalizing this finding to a large set of videos is challenging, however. Previous research on subgoal labeling either takes for granted the availability of the labels or manually generates them. But most how-to videos are not created with explicit subgoal labels. They are either missing or implicit, making it difficult for novices to construct the right mental model. In order for subgoal labels to be widely available to a large number of video tutorials, we need a scalable and efficient way to collect them.

This work-in-progress proposes a subgoal labeling method involving a mixed-initiative approach. We introduce a system where learners and AI feed each other with improved labels. On the learner's side, we use a technique we refer to as “*learnersourcing*”, a method that transforms learners' activities into useful input to the system. The key insight is that while attempting to master new skills from videos, learners are actively analyzing and inferring mental models anyway. So why not channel the natural learning effort into subgoal labels that the system can store and use? On the AI's side, components include the label generator and quality control manager that improve the labels with an input set of possibly noisy labels.

In this work-in-progress we limit our scope to video tutorials of Photoshop, popular graphical design software, due to its high penetration and the availability

Tool name

- What is the tool used here?
- e.g., *Gaussian Blur*

Ways to use a tool

- How to set parameters for this tool? What is the technique required for this tool?
- e.g., Make the color channel line to an S-curve.

Clusters

- What tools go together to accomplish a single goal?
- e.g., *Desaturate* should be applied before *Invert*

Subgoal labels / high-level description

- What is the reason for using this set of tools here?
- e.g., Why use an s-curve? To maximize contrast.

Figure 2. Types of information learners rely on while working with how-to videos.

of large, open repositories of videos on the web. However, we envision the same technique to be applicable to other how-to videos that involve the procedural workflow with distinct steps and actions.

Initial Study

To better understand how learners formulate subgoals, what information they rely on, and how they struggle when learning from video tutorials, we ran an informal study with novice Photoshop users. We invited 12 participants (8 male) to the laboratory and observed them perform design tasks in Photoshop. They were screened by a web survey to ensure they had no existing skills to solve tasks in the study. We alternated between two tasks: applying retro effect and transforming a photo to a sketch. For each task, they were given a source image to work on that none of the video tutorials contain. This design requires learners to not just replicate the steps but apply their knowledge. The only external resource available to them was video tutorials. Learners were asked to think-aloud while performing tasks. Here we summarize key findings drawn from observations during the sessions.

Following along without understanding

A common complaint learners had was that they do not understand the purpose of certain actions. Many of them followed along verbatim the instructions from a video tutorial they watched. A participant said, "*I don't know why I'm using Curves here. But I'm just doing it because the video tells me to.*" This resulted in a poor quality final image because the image from the video tutorial is not identical to the given image, and the task required acquiring and applying higher-level skills. For example, a video tutorial with a dark background applied the RGB value of (100, 76, 79) in a gradient. Unfortunately the task image had a light background

and the same RGB value would not yield a desired gradient effect. Many learners replicated the exact values and never changed them even after seeing poor quality gradient. One explanation might be that they were caught in the details and resorted to memorizing the steps instead of forming the right mental model with higher-level intentions for the steps.

Unable to group steps

It is often the case that learners miss certain steps from a tutorial. Some steps pass by too rapidly, and other steps are difficult to understand in the first attempt. In these cases participants tried to go back to a specific position in a clip, but often picked a wrong spot and ended up re-finding the desired position after watching the wrong segment multiple times. For example, many tutorials on transforming a photo to a sketch perform *desaturate* (desaturate an image to leave only black and white) before applying *invert* (invert colors of an image). Without the *desaturate* step, *invert* will not yield a clean white background with black lines. Many participants did not notice the connection between the two steps. When *invert* did not give a desired result, many participants repeatedly watched the segment applying *invert*. The problem in many cases, however, was that *desaturate* was skipped or applied with an error. As seen in this example, some steps are better understood when grouped together. Participants were not sure which steps are related, mostly because the videos they selected did not explain or contain such information. A meaningful hierarchical structure for a solution could have helped.

In summary, these novice Photoshop users struggled because they did not possess the right mental model to solve the problem they encountered. They have to handle too much information interspersed at different

(a) Fact verification

What do you think is a tool used here?

Curves
 Hue/Saturation
 Gradient Map
 Rectangular Marquee

ToolScape thinks 'motion blur' is used here. Is this correct?

Yes
 No

(b) Subgoal labeling

How would you describe the effect from left to right?

Why do you think 'Desaturate' was used here?

(c) Grouping

Can 'Duplicate Layer' and 'Invert' be combined as a single step?

Yes
 No
Why?

Figure 3. Learner quizzes asking for various types of input.

cognitive levels at once. Reducing the amount of information required in the problem solving process can reduce cognitive load for novices [8].

Based on learner observations we categorize four information types that learners rely on when solving a design task (Figure 2). Our next goal is to find ways to collect and present such information to offload the manual bookkeeping.

Approach

We propose a method for large-scale subgoal labeling. A mixed-initiative system between AI and learners benefits from each other's label-generation activity. Our approach is inspired by a mixed-initiative information extraction system by Hoffmann *et al.* [5] They demonstrated how user contribution increased while the information extraction system achieved higher accuracy with more training examples. In the learning context, we note that learners actively search for and process information while watching tutorials, although questions they had and answers they found are often not explicitly documented. Learners analyze and infer from workflows inside a video tutorial to learn skills. So why not channel the natural learning effort into subgoal labels the system can store and use?

We inject short quizzes while watching videos to encourage learners to explicitly enter information they learn from a tutorial. A design goal in designing learner quizzes is to ensure that the input soliciting method aligns with learners' natural inquiries. Our current design includes three types of questions: fact verification, subgoal labeling, and grouping (Figure 3). Fact verification questions (Figure 3(a)) simply check factual information about a workflow. For example, they ask for a tool used at a given point to remind the

learner with the name. Subgoal labeling questions (Figure 3(b)) ask for free-form textual input on the intention of a step or a set of steps. They address the problem of following along without understanding, as noted in our initial study. Collected input can be used as training data for AI to generate subgoal labels. Grouping questions (Figure 3(c)) ask if tools can go together and why. This type of question addresses the difficulty of grouping steps as observed by our initial study. The clustering information can be used as training data for the clustering algorithm.

On the AI side, we propose that the system generates the best labels it can, using clustering algorithms for grouping and topic modeling and LDA for label generation. Learners' input would be continuously inserted into the system database, which would be verified before being used as training data. A quality control mechanism using majority voting would be added to reduce noise in the learner input. For example, if three learners responded "*maximize contrast*" for a subgoal labeling question while only one said "*minimize contrast*", the system would accept "*maximize contrast*" as the correct label.

A key challenge in activating the mixed-initiative system and closing the feedback loop between learners and AI is to bootstrap the system in the initial phase. The initial stage suffers from the lack of subgoal information and low accuracy in the labels, potentially failing to give learners satisfactory learning experiences. Specifically, we plan to address the following two challenges: first, cope with the lack of information to present for new tasks or early users; second, get the critical mass of learners required to generate enough training data. A solution to populate the initial subgoal set is to mine static HTML tutorials

for subgoal candidates. This technique ensures that early learners will still see reasonable answers. A plan for reaching out to more learners is to use online advertisements. A similar idea successfully invited people to contribute to Wikipedia edits [5]. In our case we will use a Youtube-like video page that learners can visit to watch video tutorials.

Evaluation

We present preliminary, qualitative results from a pilot study with three users. We only gave them subgoal labeling quizzes because they are of our primary interest. More specifically, we asked the “how would you describe the effect from left to right?” question for 6 before-after image pairs.

Overall, many of the answers were similar among the participants with appropriate labeling assigned, which researchers manually judged. For example, for an image pair of a girl, participants replied, “*make it monotonic*” and “*make it more classy by turning into B&W and old tone*”. For an image pair of a house, participants answered, “*all shadings are removed*” and “*removing shadings*”.

Just asking the simple question does not necessarily lead to usable subgoal labels, however. Firstly, many of the labels remained at the surface level. The framing of the question can be revised to ask for more higher-level subgoal intentions. Secondly, even with the three participants we had, their input spanned different abstraction levels. For example, for a bicycle image pair, one participant wrote, “*the floor becomes brighter*”, while another wrote, “*Original bicycle color was a little dark, so it was brightened to make some details of bike stand out more clearly, perhaps contrast was enhanced at the same time. Might be color*”

histogram expanded also but not very sure!” One way to address this individual variance problem is to provide an example response, which will set the right level of expectation for labels entered.

A more systematic evaluation will require answering two questions: 1) Does the system generate good subgoal labels? 2) Do learners with injected quizzes benefit from having them?

To evaluate the quality of subgoal labels generated by our system, we will compare the labels against expert-generated labels. The TAPS method [2] can be used by the experts to guide the labeling process. To evaluate the learning experience of learners using the system, we will run a controlled laboratory user study. The study will have two conditions: a control group given only the video tutorials and an experimental group given the video tutorials with injected quizzes. Similar to our initial study, we will ask participants to perform design tasks in Photoshop using only the given system as their help resource. Then we will compare knowledge transfer, retention of the material, quality of designs produced, and self-efficacy in their ability to design.

Possible Applications

Once a database of subgoal labels is built, we envision a number of novel applications leveraging it. Here we present two possible application ideas.

Video-hopping

One value of learning with a large repository of videos is in leveraging the collective knowledge. There are often multiple alternative approaches to a problem, but many learners rely on a single solution. In the video-hopping application, the interface presents alternative video segments that share the same subgoal while the learner watches a video. For example, if the currently

playing subgoal is “*maximize contrast*”, the interface shows many other video segments that accomplish maximizing contrast.

Subgoal-based search

Many novices suffer from the vocabulary problem, which is caused by not having the domain-specific vocabulary required to perform actions in a domain. This gap between the novice language and the domain language is often a source of failure in locating the right help material because search would fail. In the subgoal-based search interface, learners can simply query using their natural language or subgoals that are abstract and generic, and the interface will translate to the domain-specific terms using the mapping between subgoals and individual steps. While the idea is similar to the query-feature graph [4] in that it translates between the user and system, this application handles hierarchical relationships. For example, if a learner searches for “*maximize contrast*”, the interface returns all videos that contain this subgoal and also suggests relevant low-level tools commonly associated with the subgoal.

Future Work

We will take an iterative design approach to revise the design of learner quizzes. We will explore successful methods in instructional design to better promote learning. A possible idea is to apply scaffolding, so that a learner is provided with more labels and guidance in the beginning, but gradually this support will be removed, turning the learner into a more active contributor from a passive receiver.

Acknowledgements

This work is supported in part by Adobe, and by Quanta Computer as part of the T-Party project. Opinions, findings, conclusions, or recommendations expressed

herein are those of the authors and do not necessarily reflect the views of the sponsors.

References

- [1] Catrambone, R. The subgoal learning model: Creating better examples so that students can solve novel problems. *Journal of Experimental Psychology: General*, 127, (1998), 355-376.
- [2] Catrambone, R. Task analysis by problem solving (TAPS): uncovering expert knowledge to develop high-quality instructional materials and training. 2011 Learning and Technology Symposium.
- [3] Clark, R., and Mayer, R. *E-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning*. Wiley & Sons, 2007.
- [4] Fournery, A., Mann, R., and Terry, M. Query-feature graphs: bridging user vocabulary and system functionality. In *Proc. of UIST 2011*.
- [5] Hoffmann, R., Amershi, S., Patel, K., Wu, F., Fogarty, J., and Weld, D.S. Amplifying community content creation with mixed-initiative information extraction. In *Proc. of CHI 2009*.
- [6] Margulieux, L., Guzdial, M., and Catrambone, R. Subgoal-labeled instructional material improves performance and transfer in learning to develop mobile applications. In *Proc. of ICER 2012*.
- [7] Palmiter, S., Elkerton, J., & Baggett, P. Animated demonstrations versus written instructions for learning procedural tasks: A preliminary investigation. *International Journal of Man-Machine Studies*, 34, (1993), 687-701.
- [8] Sweller, J. Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, 22(2), (2010), 123-138.
- [9] van Merriënboer, J., Clark, R., & de Croock, M. Blueprints for complex learning: The 4C/ID-model. *Educational Technology Research and Development*, 50(2), (2002), 39-61.