Information Processing and Management 54 (2018) 985–1001

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

Image annotation tactics: transitions, strategies and efficiency

Jing Chen[a], Dan Wang[a], Iris Xie[b], Quan Lu[c,*]

[a] School of Information Management, Central China Normal University, Wuhan 430079, China
[b] School of Information Studies, University of Wisconsin-Milwaukee, WI 53201, USA
[c] Center for Studies of Information Resources, Wuhan University, Wuhan 430072, China

ARTICLE INFO

Keywords:
Image semantics
Image annotation
Human-computer interaction
Annotation behavior

ABSTRACT

Human interpretation of images during image annotation is complicated, but most existing interactive image annotation systems are generally operated based on social tagging, while ignoring that tags are insufficient to convey image semantics. Hence, it is critical to study the nature of image annotation behaviors and process. This study investigated annotation tactics, transitions, strategies and their efficiency during the image annotation process. A total of 90 participants were recruited to annotate nine pictures in three emotional dimensions with three interactive annotation methods. Data collected from annotation logs and verbal protocols were analyzed by applying both qualitative and quantitative methods. The findings of this study show that the cognitive process of human interpretation of images is rather complex, which reveals a probable bias in research involving image relevance feedback. Participants preferred applying scroll bar (Scr) and image comparison (Cim) tactics comparing with rating tactic (Val), and they did fewer fine tuning activities, which reflects the influence of perceptual level and users' cognitive load during image annotation. Annotation tactic transition analysis showed that Cim was more likely to be adopted at the beginning of each phase, and the most remarkable transition was from Cim to Scr. By applying sequence analysis, the authors found 10 most commonly used sequences representing four types of annotation strategies, including Single tactic strategy, Tactic combination strategy, Fix mode strategy and Shift mode strategy. Furthermore, two patterns, "quarter decreasing" and "transition cost," were identified based on time data, and both multiple tactics (e.g., the combination of Cim and Scr) and fine tuning activities were recognized as efficient tactic applications. Annotation patterns found in this study suggest more research needs to be done considering the need for multi-interactive methods and their influence. The findings of this study generated detailed and useful guidance for the interactive design in image annotation systems, including recommending efficient tactic applications in different phases, highlighting the most frequently applied tactics and transitions, and avoiding unnecessary transitions.

1. Introduction

Millions of images become accessible to people with the new development of Internet technology. In order to retrieve the digital images that exactly satisfy users' needs, annotating the content with semantically meaningful labels is of great necessity. Prior image annotation approaches, regardless of interactive ones, such as relevance feedback techniques (Nezamabadi-Pour & Kabir, 2009; Rui, Huang, Ortega, & Mehrotra, 1998), or non-interactive ones, such as multiple feature spaces (Sun, 2013; Xu, Tao, & Xu, 2015) and other statistical models (Feng & Lapata, 2010; Qi & Han, 2007), fail to tackle the challenge of semantic gap fundamentally (Ke, Li, &

* Corresponding Author
E-mail address: mrluquan@whu.edu.cn (Q. Lu).

https://doi.org/10.1016/j.ipm.2018.06.009
Received 16 January 2018; Received in revised form 20 June 2018; Accepted 23 June 2018

Cao, 2011). On one hand, users of current human-computer interactive image annotation tools are supposed to annotate an image by tagging, while ignoring whether the description expressed by tags matches image semantics. On the other hand, human interpretation of images is complicated, while its process and mechanisms remain uncertain (Ivasic-Kos, Ipsic, & Ribaric, 2015). At the same time, how users perceive images in semantic ways is not very well understood. Consequently, more insight into human-computer interactive image annotation behavior is needed, so that users' interpretation process in annotation of images can be understood, narrowing the image semantic gap, ultimately resulting in better image annotation experiences for all.

At present, there are three major interactive methods usually used to investigate an individual's image interpretation: (a) normative rating method (Scherer, 2005), (i.e. typically features a rating scale from 1-5, 1-9, etc.); (b) semantic visualization based scroll bar method (Schmidt & Stock, 2009), i.e. dragging the scroll bar to express the image's semantic meaning; (c) ranking visualization based comparison method (Hardoon & Pasupa, 2010), i.e. comparing with other images when evaluating the current image. Different methods represent users' diverse image cognition processes, and their own strengths and weaknesses. Rating method is characterized by quantifying evocation into specific value(s). The scroll bar and image comparison methods are relatively perceptual. In the scroll bar method, users are supposed to adjust the scroll bar to express perceived semantic intensity. When performing a comparison method, users' image interpretation may be influenced by reference image(s) since they would estimate differences in generating an appropriate annotating result. The existing studies on image semantic interpretation are all conducted under one single method, such as rating (Grühn & Scheibe, 2008), scroll bar (Schmidt & Stock, 2009) or image comparison (Junge & Reisenzein, 2015), which may produce a bias in collecting image relevance feedback. Users may not express the image interpretation accurately by just rating, the most widely used annotation method. In addition, this kind of single annotation method application seems to limit the comprehensive investigation of users' sophisticated annotation behavior. Hence, the combination of multiple interactive methods helps researchers understand the complicated process of image interpretation.

Moreover, in the field of information retrieval (IR), systems usually offer users multiple interactive activities, such as formulating queries, scanning websites, and so on (Baskaya, Keskustalo, & Järvelin, 2013; Niu & Kelly, 2014). Searching behavior patterns (e.g., frequently applied actions, transition of these actions) generated by selecting or combining these operations enable researchers to study general users' searching behaviors and strategies (Joho, Jatowt, & Blanco, 2015; Xie, 2011). Therefore, this approach of user behavior investigation in search fields has provided reference for image annotation. In this study, the authors intend to create an interactive annotation environment that provides various methods to explore users' image annotation behavior.

With respect to the essential components of the annotation process, the authors offered an explanation of annotation moves, tactics, and strategies for this study after merging definitions proposed by Bates (1979, 1990) and Marchionini (1995) for search moves, tactics, and strategies. "Move" is a common alternative word for "tactic". Annotation moves are basic thoughts or actions in the annotation process. Tactics indicate a move or moves, including annotation choices and actions that users apply to advance their annotation. Annotation strategies represent patterns of sequential tactics which imply users' plans for the annotation. In the annotation process, annotation tactics can be identified on the basis of annotation moves (e.g., dragging the slider, considering the position of the scroll bar, evaluating the value, and comparing with the reference image.) (Schmidt & Stock, 2009; Zhang, Fu, Liang, Chi, & Feng, 2010), which is similar to the search process (Shiri & Revie, 2003). Since the objective of this study is to figure out users' image interpretation processes, annotation moves discussed here are mainly related to the user's cognitive involvement. In addition to these typical moves, tactics, such as fine tuning (Cuzzola, Jovanović, Bagheri, & Gašević, 2015), are critical to the success of the annotation process. In that most works of the search strategies are in relation to topic refinement (Chen & Dhar, 1991; Xie & Joo, 2010), the act of fine tuning plays the same role as refinement, which is the basis to classify annotation strategies into two types in this study. The first type represents the common patterns of tactic selection without fine tuning, while the second type reveals the tactic chains including fine tuning. Regardless of whether a user refines the annotation result or not, the strategy he or she employs implies his or her plan for the annotation.

The annotation process is a complicated and dynamic one, in which tactics adopted by users may change during the process. To understand the process, merely identifying tactics is not enough. It is also critical to look into the transitions among annotation tactics. Each transition focuses on detecting changes between two annotation actions. In addition, a chain of tactic transitions within a strategy illustrates the annotation process. Unfortunately, few researchers have investigated transitions in annotation tactics. Parallel to search tactic research, where transitions have been conducted via discourse or log analysis (Rieh & Xie, 2006; Xie & Joo, 2010), the occurrences of transitions in tactics could also be derived from utterances or log data to highlight a user's successive decision-making process when accomplishing an annotation task.

This study focuses on exploring users' annotation tactics, transitions, strategies and efficiencies by investigating users' behaviors in dealing with annotation tasks. The analysis of annotation tactics helps researchers acquire a better understanding of the nature of human's interpretation process to images. The study of annotation tactics, transitions, strategies and efficiencies offers an opportunity to recognize users' image annotation behavior, thus provide insights into annotation behavior research, and finally improve existing annotation systems to assist users optimize their annotation behaviors and enhance their image annotation experience.

2. Related Work

2.1. Image Annotation

The semantic gap, in most of the previous research, has been defined as the difference between the subjective users' understanding of an image and the objective computer's interpretation of the users' annotation (Bahmanyar, Murillo Montes De Oca, & Datcu, 2015). Early solutions attempting to fill the semantic gap is to use relevance feedbacks from users (Nezamabadi-Pour & Kabir,