Expert Systems with Applications 42 (2015) 246–258

Contents lists available at ScienceDirect

Expert Systems with Applications

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

A collaborative algorithm for semantic video annotation using a consensus-based social network analysis

Trong Hai Duong [a], Ngoc Thanh Nguyen [b,*], Hai Bang Truong [c], Van Huan Nguyen [d]

[a] International University, Vietnam National University-HCM, Viet Nam
[b] Institute of Informatics, Wroclaw University of Technology, Poland
[c] University of Information Technology, Ho Chi Minh City, Viet Nam
[d] Nguyen Tat Thanh University, Viet Nam

ARTICLE INFO

Article history:
Available online 7 August 2014

Keywords:
Video annotation
Semantic web
Social network
Consensus

ABSTRACT

Social TV represents a new form of shopping that enables consumers to view, select and buy products. This highlights the need for a collaborative video annotation technique. This paper proposes a collaborative algorithm for semantic video annotation using consensus-based social network analysis (SNA). The collaborative video annotation process is organized based on social networks. Here, the media content is shared with friends of friends who collaboratively annotate it. This study used an ontology-based approach to semantically describe the media content and allow sharing between users. A consensus-based method was used to reconcile conflicts between participants' annotations. The experimental results indicated that the use of SNA-based collaboration criteria to evaluate the collaborative process enhances the completeness and consistency of collaborative annotation. The more the collaboration criteria are satisfied by the collaborative group, the faster the group will reach a consensus. In addition, the consensus-based method is an effective approach for resolving conflicts on collaborative annotation.

1. Introduction

Social TV represents a new form of shopping that enables consumers to view, select and purchase products. For this, the sellers annotate videos and associate items with information from e-commerce systems in a semantic manner. On the other hand, manufacturers annotate the way items are produced. In addition, buyers provide their opinions on the items received. Therefore, there is a need for a collaborative video annotation method. *Collaborative video annotation is a process, in which a group of participants contribute annotations to specific videos of interest*. Annotation can be improved by collaboration between participants because they can benefit from each others' skill and knowledge. According to Duong and Jo (2010) and Duong, Nguyen, and Jo (2010), any effective collaboration should be satisfied criteria including *inclusive, egalitarian, interactive, coordinated*, and *trustworthy*. Here, the term *inclusive* means that there should be a sufficient number of participants. The term *egalitarian* indicates that these participants should be provided with as many opportunities as possible for their collaboration. *Interactive* suggests that there should be an easy way to establish contact with all collaborators within a collaborative group. The term *coordinated* means that any annotation information shared within the group should be easy to access and the term *trustworthy* indicates that the participants should be allowed to edit annotations based their behavior. Therefore, the workflow of a collaborative video annotation process is complex. According to Palau, Montaner, and Lopez (2004), SNA can highlight many relationships based on the aforementioned collaboration criteria and entail a range of centrality measures, such as *density, degree centrality, closeness centrality*, and *between centrality*. If the collaborative relationships in video annotation systems can be represented by means of social network, these centrality measures can be used to organize the collaborative annotation process and evaluate the collaboration criteria. For example, the *closeness centrality* can be used to identify those annotators occupying more advantageous positions for obtaining annotation support. The *density* is useful for evaluating the conditions for accessing and sharing annotation information between collaborators in a coordinated manner. In addition, consensus-based methods seem to be an effective approach for resolving conflicts in collaborative annotation.

This paper proposes a collaborative algorithm for semantic video annotation using a consensus-based SNA. The ontologies for semantically describing media content and facilitating sharing

* Corresponding author.
E-mail addresses: haiduongtrong@gmail.com (T.H. Duong), ngoc-thanh.nguyen@pwr.edu.pl (N.T. Nguyen), bangth@uit.edu.vn (H.B. Truong), conghuan_bkstudent@yahoo.com (V.H. Nguyen).

http://dx.doi.org/10.1016/j.eswa.2014.07.046
0957-4174/

between heterogeneous users or devices (e.g., smart-phone and PDAs) were considered. This study employed two ontologies. One ontology describes visual features of media objects, such as their color, texture, shape, motion and position (e.g., MPEG-7), and the other provides a knowledge base for a specific domain (domain ontology) for annotating the video content. The LSCOM ontology reported by Naphade et al. (2006), which is used in broadcast news video annotation, was considered for the domain ontology. Each media object in a specific video can be considered as an instance of a domain concept belonging to the LSCOM ontology. Here, the annotator chooses a relevant concept to describe a media object of interest and the collaborative video annotation process is organized based on a social network. The media content is shared within a large network of friends of friends who annotate the content based on the SNA. Annotators collaborate only with reliable users. Reliability is expressed through trust scores and other collaboration criteria. Once the annotator has a set of reliable users, he or she can benefit from their knowledge or skill for annotation purposes. When the annotator is unsure of the annotation information, he or she asks reliable users for their knowledge and opinions, and uses their trust scores to determine if the annotated object should be of interest. This study suggests that reliable users provide pertinent opinions while evaluating the entire collaborative process using collaboration criteria but that they can provide insufficient or conflicting knowledge. A consensus choice was used to resolve conflicts over annotation information between participants.

The remainder of this paper is organized as follows. Section 2 provides a literature review. Section 3 proposes a collaborative semantic annotation methodology using a consensus-based SNA. Section 4 presents the experimental results, and Section 5 reports the conclusions.

2. Related work

With the recent explosive growth of smart or social TV services, there have been considerable increases in social videos as well as traditional digital videos, such as TV programs and VOD offerings. On the other hand, it is difficult to find interesting and relevant media content and encourage interactions between users and Smart TV platforms due to the lack of semantic media content. This highlights the urgent need for semantic techniques to analyze the multimedia content, particularly those for a semantic video annotation.

2.1. Multimedia ontology

Ontologies have the following attributes: provide machine-processable information sources; enable users to organize information according to taxonomical concepts based on their own attributes; allow for a semantic description of the media content; and facilitate the sharing of media content. Each media object can be considered as an instance of a domain concept belonging to multimedia ontology, which offers advanced search functions that can include searches for similar or related information integrated from a range of social media sites (Hakeem, Lee, Javed, & Haering, 2009; Sebastine, Thuraisingham, & Prabhakaran, 2009; Wei, Zhao, Zhu, & Liu, 2010). Several multimedia ontologies have been proposed. Among them, the LSCOM ontology was proposed by Naphade et al. (2006) includes a set of 1,000 concepts built collaboratively by IBM, Carnegie Mellon University, and Columbia University in conjunction with CyC Corporation and various research communities. The aim is to create a framework for ongoing research on the semantic analysis of multimedia content. The Video Event Representation Language (VERL) of Francois, Nevatia, Hobbs, Bolles, and Smith (2005) is an event ontology used to describe the events and objects in videos. Here, video events and objects were annotated using the Video Event Markup Language (VEML). Annotations draw on the VEML for organizing content. Unambiguous data sharing between users enables marking up data streams using the VERL and the VEML. In addition, annotation data is accessible to automatic machine manipulations for indexing or inferring. The MPEG-7 multimedia content description standard already provides tools for representing media content fragments. The MPEG-7 Visual Part supports the objects color (e.g., dominant colors and color layouts), texture, shape (e.g., region/contour-based), and motion (local or global) descriptors (Sikora, 2011). The ultimate goal and objective of the MPEG-7 Visual Standard is to provide standardized descriptions of streamed or stored images or header bits that assist users or applications to identify, categorize or filter images or videos. Similarly, the MPEG-7 Multimedia Description Scheme (MDS) supports spatial (directional or topological) and temporal multimedia segment relationships as well as the hierarchical structures for the decomposition of multimedia segments.

2.2. Automatic semantic video annotation

In general, automatic semantic video annotation also referred to as video concept detection (Jeong, Hong, & Lee, 2011; Naphade & Smith, 2004; Park, Lee, Moon, Park, & Lee, 2007), video semantic analysis (Snoek, Worring, & Smeulders, 2005), and high-level feature extraction (Kraaij & Over, 2005; Wu & Li, 2011) can be performed using machine learning methods. This paper presents the representative learning-based video annotation methods as follows. According to Park et al. (2007) and Jeong et al. (2011), an automatic semantic video annotation system employs ontologies and semantic inference rules to facilitate video retrieval and consists of four functional modules: shot-level annotation, group-level annotation, scene-level annotation and video-level summarization. The visual features of objects in a video shot are first extracted using MPEG-7 visual descriptors, which are then mapped to the corresponding semi-concept values that are used as the basic units for inferring high-level concepts. The high-level concepts of objects are extracted automatically by applying shot-level inference rules to ontologies, and semantically-similar groups are parsed into one semantic scene based on the frequency and similarity of high-level concepts from each group. Finally, a video-level summary is generated by integrating and analyzing the high-level concepts of groups and scenes.

2.3. Collaborative semantic video annotation

Collaborative semantic video annotation methods focus on a web-based annotation interface. The participants collaboratively create and share video annotations through social networks. Several approaches can be used to achieve this. The IBM Efficient Video Annotation (EVA) system is a server-based tool for the semantic annotation of large video and image collections, and allows user access through web browsers Volkmer, Smith, and Natsev, 2005. The users can be assigned to specific tasks such that each annotator can have a personalized view of the collection. Tasks can be shared by a variety of users. The novel features of the system include its ability to collect aggregate-level user data during the annotation process and support inter-annotator analyses.

However, this system was designed with a few simplifying assumptions to promote consistence, simplicity, and speed of annotation:

- All annotations use an upper ontology and no free text is allowed;