



Contents lists available at ScienceDirect

Signal Processing: *Image Communication*journal homepage: www.elsevier.com/locate/image

A Bayesian network modeling approach for cross media analysis

Christina Lakka^a, Spiros Nikolopoulos^{a,c,*}, Christos Varytimidis^b, Ioannis Kompatsiaris^a^a Informatics and Telematics Institute, CERTH, 6th km Charilaou-Thermi Road, Thessaloniki, Greece^b School of Electrical and Computer Engineering, National Technical University of Athens, Greece^c School of Electronic Engineering and Computer Science, Queen Mary University of London, E1 4NS London, UK

ARTICLE INFO

Article history:

Received 7 March 2010

Accepted 26 January 2011

Available online 3 February 2011

Keywords:

Cross media analysis

Knowledge fusion

Bayesian networks modeling

Compound documents analysis

Video shot classification

ABSTRACT

Existing methods for the semantic analysis of multimedia, although effective for single-medium scenarios, are inherently flawed in cases where knowledge is spread over different media types. In this work we implement a cross media analysis scheme that takes advantage of both visual and textual information for detecting high-level concepts. The novel aspect of this scheme is the definition and use of a conceptual space where information originating from heterogeneous media types can be meaningfully combined and facilitate analysis decisions. More specifically, our contribution is on proposing a modeling approach for Bayesian Networks that defines this conceptual space and allows evidence originating from the domain knowledge, the application context and different content modalities to support or disprove a certain hypothesis. Using this scheme we have performed experiments on a set of 162 compound documents taken from the domain of car manufacturing industry and 118 581 video shots taken from the TRECVID2010 competition. The obtained results have shown that the proposed modeling approach exploits the complementary effect of evidence extracted across different media and delivers performance improvements compared to the single-medium cases. Moreover, by comparing the performance of the proposed approach with an approach using Support Vector Machines (SVM), we have verified that in a cross media setting the use of generative rather than discriminative models are more suited, mainly due to their ability to smoothly incorporate explicit knowledge and learn from a few examples.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

The automatic extraction of semantic metadata from multimedia content has been recognized as a particularly valuable task for various applications of digital content consumption. Current literature has made considerable progress in this direction especially for single-medium

scenarios. However, the methods proposed in the literature do not apply in cases where information is spread over different media types and unless considered simultaneously, its contribution cannot be fully exploited by the analysis process. Motivated by this, cross media analysis seeks to enhance semantic metadata extraction by exploiting information across media. Practically, the aim of such methods is to combine the evidence extracted from different media types and accumulate their effect in favor or against a certain hypothesis. These pieces of evidence can belong to different levels of granularity and used differently by the analysis mechanism. For instance, we can consider cross media analysis to be a general fusion problem that is carried out at different

* Corresponding author at: Informatics and Telematics Institute, CERTH, 6th km Charilaou-Thermi Road, Thessaloniki, Greece.

Tel.: +30 2311 257 752; fax: +30 2310 474 128.

E-mail addresses: lakka@iti.gr (C. Lakka),

nikolopo@iti.gr (S. Nikolopoulos),

chrisvar@image.ntua.gr (C. Varytimidis), ikom@iti.gr (I. Kompatsiaris).

levels of abstraction, namely result-level [1–3], extraction-level [4–6] and feature-level [7–9].

In the result-level approach, information from each data source is initially extracted separately and, still separately, transformed into conceptual information. Though result-level approaches are closer to human cognition and more suited for exploiting explicit knowledge (i.e., knowledge that is explicitly provided by experts in the form of rules, ontologies or other formal languages for knowledge representation), their major drawback is that each extractor has to produce its own internal evidence in order to extract the conceptual information. In the extractor-level approach the conceptual information is not extracted separately from each modality but instead, the analysis mechanism takes into account evidence from other modalities. Information coming from one medium may assist the information extraction module of another medium, using as input the output of another extractor. However, in contrast to the result-level approaches where knowledge is incorporated into the conceptual space, in this case it can only be exploited as part of a task specific mechanism. The feature-level approach consists in using all low-level features that can be extracted from each medium within the same analysis process. Initially, low-level features (e.g., text-tokens, named entities or image descriptors) are extracted separately from each modality and integrated into a common, concatenated representation. Subsequently, the common representation is used as input for the analysis process (i.e., classification, indexing, etc.). Feature-level analysis aims at exploiting the joint existence of low-level features into the same resource, but it is rather difficult to incorporate explicit knowledge in this case.

Our work is motivated by the need to boost the efficiency of cross media analysis using the knowledge explicitly provided by domain experts (i.e., domain knowledge). This was the reason for developing a method that operates on the result-level of abstraction and allows domain knowledge to become part of the inference process. Our method combines the soft evidence (soft in the sense that a confidence degree is attached to every piece of evidence) collected from different media types, to support or disprove a certain hypothesis made about the

semantic content of the analyzed resource. Soft evidence are obtained by applying single-medium analyzers on the low-level features of the different media types. Subsequently, these pieces of evidence are used to drive a probabilistic inference process that takes place in a Bayesian Network (BN). The structure and parameters of the BN are constructed by incorporating domain knowledge (expressed using ontologies) and application context (captured by conditional probabilities). We use the soft evidence to update the observable variables of the BN and verify or reject the examined hypothesis based on the posteriori probability of the remaining variables. Fig. 1 demonstrates the functional relations between the components of the proposed cross media analysis scheme.

The novelty introduced by the proposed method is that it manages to integrate into a common inference framework three types of information: (a) information obtained from the analysis of heterogeneous content (i.e., the output of single-medium analyzers supplied as soft evidence), (b) information about the domain that is provided explicitly, and (c) contextual information that is learned from sample data. Our contribution is on proposing a modeling approach for the BN that results in a conceptual space of likelihood estimates. In this space the evidence originating from the domain knowledge, the application context and the different content modalities can be meaningfully combined and facilitate semantic metadata extraction.

We show using content from a real world application taken from the car manufacturing industry as well as from the TRECVID2010 competition, that performing cross media analysis using the proposed method leads to significant improvements compared to the cases where single-medium analyzers act separately. We also prove experimentally that, in a cross media setting, the generative models outperform the discriminative ones in fusing the extracted evidence, mainly due to their ability in efficiently handling prior knowledge and learning from a few examples.

The rest of the manuscript is organized as follows. Section 2 details the proposed approach for modeling the BN and determining the conceptual space. Section 3 describes the implemented cross media analysis scheme

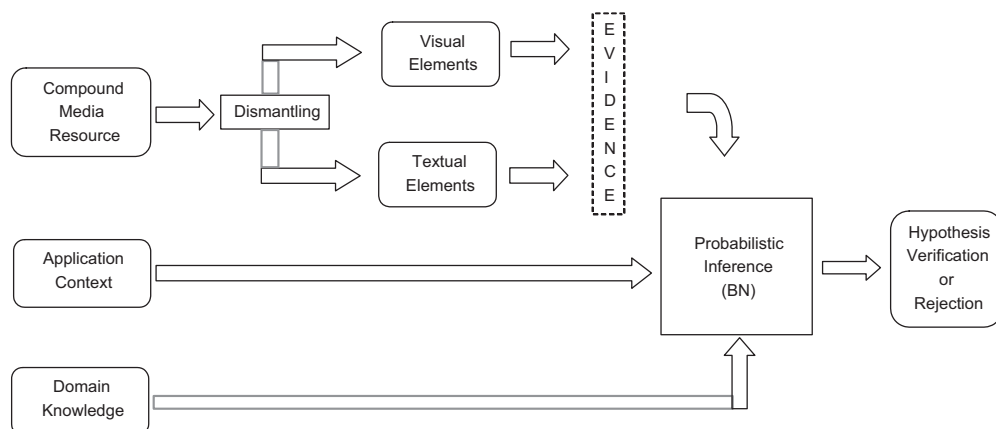


Fig. 1. Cross media analysis scheme.

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات