Information Processing and Management 52 (2016) 783–800

Contents lists available at ScienceDirect

Information Processing and Management

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

A multimodal query expansion based on genetic programming for visually-oriented e-commerce applications

Patrícia C. Saraiva [a], João M.B. Cavalcanti [a,*], Edleno S. de Moura [a], Marcos A. Gonçalves [b], Ricardo da S. Torres [c]

[a] Institute of Computing, Federal University of Amazonas, AM, Brazil
[b] Department of Computer Science, Federal University of Minas Gerais, MG, Brazil
[c] Institute of Computing, University of Campinas, SP, Brazil

ARTICLE INFO

Article history:
Received 6 November 2014
Revised 23 February 2016
Accepted 7 March 2016
Available online 20 April 2016

Keywords:
Content-based image retrieval
Multimodal query expansion
Genetic programming
E-commerce

ABSTRACT

We present a novel multimodal query expansion strategy, based on genetic programming (GP), for image search in visually-oriented e-commerce applications. Our GP-based approach aims at both: learning to expand queries with multimodal information and learning to compute the "best" ranking for the expanded queries. However, different from previous work, the query is only expressed in terms of the visual content, which brings several challenges for this type of application. In order to evaluate the effectiveness of our method, we have collected two datasets containing images of clothing products taken from different online shops. Experimental results indicate that our method is an effective alternative for improving the quality of image search results when compared to a genetic programming system based only on visual information. Our method can achieve gains varying from 10.8% against the strongest learning-to-rank baseline to 54% against an adhoc specialized solution for the particular domain at hand.

1. Introduction

Recent technological advances have contributed to the creation of new opportunities for content-based image retrieval (CBIR) applications. Such is the case for systems that search for products in online stores, popularized by the wide-spread use of mobile devices, such as tablets and smartphones that can access applications using remote resources, as well as by modern and secure payment mechanisms that allow users to buy goods using their mobile devices.

Contributing to this scenario, we can see an increasing number of e-commerce companies, which demand constant improvements in search technologies to support their operations. Additional evidence that this specific problem is increasing in importance can be found in recent reports summarizing e-commerce activities. These reports show a fast growing market share of segments such as fashion, apparel and accessories along with furniture and home furnishings, where visual search can play an important role. For instance, these last two segments combined represent approximately 25% of e-commerce sales in the US in 2012, according to the eMarketer report (eMarketer, 2012). Garment, including clothing, and accessories,

* Corresponding author. Tel.: +559281142535.
E-mail addresses: patricia.saraiva@gmail.com (P.C. Saraiva), john@icomp.ufam.edu.br (J.M.B. Cavalcanti), edleno@icomp.ufam.edu.br (E.S. de Moura), mgoncalv@dcc.ufmg.br (M.A. Gonçalves), rtorres@ic.unicamp.br (R.d.S. Torres).

http://dx.doi.org/10.1016/j.ipm.2016.03.001
0306-4573/

is the second biggest category in e-commerce shopping in the US and is growing in importance in most of the largest e-commerce markets worldwide.

In this context, we propose a new machine-learning CBIR approach for visually-oriented e-commerce applications in which image search is a key component. In this specific domain, we consider that the system does not have access to textual information from the user. A common scenario occurs, for instance, when a user takes a photo of a product in a store with his/her cell phone and wants to find similar items to the one she found at the store. This type of query is useful when the user is searching for products such as clothing, shoes, handbags, watches and accessories. The visual presentation of this sort of product is essential for the consumer's purchase decision. Considering that there are usually several visually similar products in the product database, visual search can help in the specific task of finding the desired product and improving the user experience. As one example of our target application, we mention popular apps currently available such as ebay fashion,[1] look4color[2] in the US, and netshoes click[3] in Brazil. In these applications, users search for products by giving as input photos taken from other products. Such applications are growing in importance since mobile devices with online access to apps are increasingly available.

One of the key properties that makes our target application different from traditional image search applications is the final goal for the search task. The query in our target application is an image that represents a product, but what makes a product relevant, or not, to the user may be coded in other attributes not present in its image. Sometimes products in the database may be relevant to the user, even though their image representation is not so close to the image provided in the query. Examples of this problem include differences in the approach to produce the product photos, such as folded versus unfolded t-shirts. Another example occurs when products with distinct colors and textures may also be considered relevant in specific cases, given their styles and so on.[4]

Our proposed solution allows finding relevant products related to an image query even though their image representation is not (very) similar to the query. Our solution to the problem is to perform a multimodal expansion, using the initial query to infer other attributes that are relevant to the query, such as product's category and textual description. The key idea is to use the visual information to produce an initial ranking and then extract accurate multimodal information from the results that may be used as an expansion to the initial query.

Our strategy exploits a self-reranking machine learning approach based on automatic multimodal query expansion. The challenge here is that, different from most previous work, we do not assume that the query is defined in terms of both image visual content and textual descriptions. In fact, the constraints usually imposed by our target application imply that only the visual information is available, i.e., a photo of the desired product. The key idea of our method is to expand the initial image query with information about the inferred category of the query image along with textual content automatically associated with other visually similar images. This type of information is usually largely available in online product catalogs, helping us in the goal of improving the quality of the overall retrieval process. Thus, we expand the initial query image with multimodal information, producing a new ranking based on the expanded query.

The expansion and actual ranking of images exploit a Genetic Programming (GP) approach to perform both: the expansion of the initial query and the computation of a new ranking based on it. We propose and experiment four alternatives to using GP for deriving multimodal query expansion methods. GP is used to find the best possible multimodal combination from the available pieces of evidence. We chose GP for a number of reasons, including: (i) excellent effectiveness in previous CBIR studies (Andrade, Almeida, Pedrini, & da S. Torres, 2012; Calumby, da S. Torres, & Gonçalves, 2014; Faria et al., 2010; Ferreira et al., 2011; Torres et al., 2009), mainly when exploiting multimodal information; (ii) capability to find near-optimal solutions in large search spaces, as is the case here; (iii) and capability to deal with multiple objectives at the same time (in our case query expansion and effective ranking). After learning a ranking function in an offline process with GP, we apply the function at query processing time without any extra overhead. As far as we know, GP has never been applied in the scenario we deal with in this paper, i.e., only the visual information is initially available for multimodal query expansion. The challenge of exploiting only visual aspects comes from the difficulty of mapping low-level features obtained by means of image processing algorithms to high-level concepts found in images, the well-known *semantic gap* problem (Liu, Zhang, Lu, & Ma, 2007).

Experimental results indicate that our new GP multimodal query expansion approach is able to significantly improve the overall quality of results of e-commerce visual search applications when compared to the application of GP without expansion. This demonstrates that the idea of performing a multimodal machine learning based automatic expansion for image queries is very promising.

In a previous work (dos Santos, Cavalcanti, Saraiva, & de Moura, 2013), we have addressed a similar problem, i.e., searching products using only image queries.[5] At that opportunity, however, no learning paradigm was explored. Machine learning solutions are both: (i) more principled, with a lot of theoretical background, and (ii) more flexible, able to easily

[1] www.mobile.ebay.com (As of October 2015).

[2] www.look4color.com (As of October 2015).

[3] https://itunes.apple.com/br/app/netshoes-conecta-voce-ao-esporte/id886443637?mt=8 (As of October 2015).

[4] We remind the reader that in our target application the system takes only an image, usually a product photo, as input.

[5] In fact, as far as we know, this is the only work that takes advantage of multimodal information to expand visual queries. Other works use multimodal information to expand textual queries.