

Title: Identifying challenges during colonoscopy procedure through the application of Latent Dirichlet Allocation (LDA) method to interviews with endoscopists

Juan Francisco Ortega-Morán^{1,*}, Luis Bote-Curiel¹, Blas Pagador¹, Águeda Azpeitia², Cristina L. Saratxaga³, Francisco M. Sánchez-Margallo¹

¹ Jesús Usón Minimally Invasive Surgery Centre, Ctra. N-521, Km 41.8, 10071, Cáceres, Spain

² Biobanco Vasco, Fundación Vasca de Investigaciones e Innovación Sanitaria (BIOEF), Ronda de Azkue, 1, 48902, Barakaldo, Spain

³ Grupo de visión artificial, TECNALIA, Basque Research and Technology Alliance (BRTA), Parque Científico y Tecnológico de Bizkaia, Geldo Auzoa, 700 Edificio, 48160, Derio, Spain

Introduction: The high incidence and mortality rate of colorectal cancer require new technologies to improve early diagnosis of the disease. In order to improve the design and development of such new systems, it is important to previously identify the colonoscopy procedural constraints. For that, interviews to endoscopists, as the main actors involved in the colonoscopy procedure, are often performed. The aim of this study is to identify challenges during colonoscopy procedure through the application of Latent Dirichlet Allocation (LDA) method to interviews with endoscopists.

Methods: The LDA method is a statistical analysis technique widely used to identify latent topics in a collection of documents. LDA allows getting the most relevant words that appear in each topic and representing them in word cloud charts. In these charts, the larger the probability of occurrence of a specific word is, the bigger and bolder it appears in the chart. In this study, LDA method has been applied to interviews performed to endoscopists.

Results: Six endoscopists from two different Spanish hospitals were interviewed. After applying the LDA method, two topics that best describe the interviews have been found: polyps and patient. In fact, a histogram of the 100 most used words in the interviews shows them as the two most repeated words (36 and 22, respectively). A further detailed analysis of both topics in the interviews shows that one of the main problems during the colonoscopy procedure is the detection of polyps, and that the technical problems for that are mainly related to the patients (poor bowel preparation prior to colonoscopy, and anatomy of the colon).

Conclusions: Firstly, LDA method shows an interesting potential to identify hidden patterns within free-text data, including interviews, so far scarcely used in the health care area. Secondly, results shows that the most important challenge during colonoscopy procedure is to see what you cannot see. On the one hand, educational initiatives for patients to improve their compliance of bowel preparation should be encouraged, but also for health care providers. On the other hand, although mechanical add-on devices and accessory device-based systems are used to maximize the colon exposure, the challenge in polyp detection could be the achievement of an endoscopic tip with 360° of vision, either with a rotating tip or with a system that allows peripheral as well as focused vision, for the improvement of the visual field behind the colonic folds.

Keywords: Latent Dirichlet Allocation (LDA); Colorectal cancer; Detection; Interviews; Colonoscopy