SINAI at VideoCLEF 2009

José M. Perea-Ortega, Arturo Montejo-Ráez, M. Teresa Martín-Valdivia, L. Alfonso Ureña-López
SINAI research group. Computer Science Department. University of Jaén
Campus Las Lagunillas, Ed. A3, E-23071, Jaén, Spain
{jmperea,amontejo,maite,laurena}@ujaen.es

Abstract

This paper describes the second participation of the SINAI research group in the VideoCLEF track. This year we only participated in the subject classification task. A training collection was generated using the data provided by the VideoCLEF organization. Over this data, a supervised learning approach to classify the test videos was conducted. We have used Support Vector Machines (SVM) as classification algorithm and two experiments have been submitted, using the metadata files and without using them, during the generation of the training corpus. The results obtained show the expected increase in precision due to the use of metadata in the classification of the test videos.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval; H.3.4 Systems and Software

General Terms

Algorithms, Experimentation, Languages, Performance

Keywords

Image classification, Information Retrieval

1 Introduction

This paper presents the second participation of the SINAI research group at the VideoCLEF 2009 track. The goal of the track is to develop and evaluate tasks involving the analysis of multilingual video content [6]. This year we only participate in the subject classification task. It is about automatic tagging of videos with subject labels such as "*Archeology*", "*Dance*", "*History*", "*Music*" or "*Scientific Research*". A total of 46 subject labels have been defined. The classification process only makes use of the speech transcriptions of the videos and some metadata provided.

Our group have some experience in the field of the multimedia video retrieval [4] and image retrieval, participating the last years in several tasks of the ImageCLEF track [3, 2, 1]. With regard to the video categorization, we participated in VideoCLEF 2008, applying a simple approach to resolve the classification task: to use an Information Retrieval (IR) system as classifier. The speech transcriptions were used as textual queries and we generated a search collection based on documents retrieved using the Google[1] search engine. The results obtained showed that an IR

[1] http://www.google.com/

system can perform well as video classifier if the speech transcriptions of the videos have good quality [8].

This year we have submited some experiments following one main approach: supervised categorization using labeled samples. For that, a learning corpus has been generated using the data provided by the VideoCLEF organization. Then, we have applied Support Vector Machines (SVM) [5] as classification algorithm.

The following section describes how the training collection has been generated. In Section 3, we explain briefly the use of SVM as text classifier. In Section 4, we describe the experiments and we show the results. Finally, conclusions are presented in Section 5.

2 Generating the training corpus

The VideoCLEF 2009 training corpus consists of 262 XML files. These Automatic Speech Recognition (ASR) files belong to the VideoCLEF 2008 (50 files) and TRECVid[2] 2007 (212 files). In addition, there are some *metadata* files provided by the VideoCLEF organization. A fragment of a ASR file and a *metadata* file are showed in Figure 1 and Figure 2, respectively.

```
[...]
  <AudioSegment>
    <TextAnnotation>
        <FreeTextAnnotation>over</FreeTextAnnotation>
    </TextAnnotation>
    <MediaTime>
        <MediaTimePoint>T00:00:06:780F1000</MediaTimePoint>
        <MediaDuration>PT0H0M0S339N1000F</MediaDuration>
    </MediaTime>
  </AudioSegment>
  <AudioSegment>
    <TextAnnotation>
        <FreeTextAnnotation>naar</FreeTextAnnotation>
    </TextAnnotation>
    <MediaTime>
        <MediaTimePoint>T00:00:07:120F1000</MediaTimePoint>
        <MediaDuration>PT0H0M0S139N1000F</MediaDuration>
    </MediaTime>
  </AudioSegment>
  <AudioSegment>
    <TextAnnotation>
        <FreeTextAnnotation>het</FreeTextAnnotation>
    </TextAnnotation>
    <MediaTime>
        <MediaTimePoint>T00:00:07:259F1000</MediaTimePoint>
        <MediaDuration>PT0H0M0S110N1000F</MediaDuration>
    </MediaTime>
  </AudioSegment>
  <AudioSegment>
    <TextAnnotation>
        <FreeTextAnnotation>gebouw</FreeTextAnnotation>
    </TextAnnotation>
[...]
```

Figure 1: Fragment of a ASR file

```
[...]
  <asset>
     <assets_id>17785</assets_id>
     <title>NOORDERLICHT (expr_id:22205)</title>
     <creator>VPRO</creator>
     <creator>Dijk, Jochgem van</creator>
     <description>BG_34978-out.wmv</description>
     <description>Teleblik</description>
     <description>INTERVIEWS met: prof. Robert Behringer, fysicus
         Duke University, (op het terrein van een cementfabriek)
         over het verrassend gedrag van  korrelstructuren, hoe deze
         zich manifesteren als vaste stof maar ook als vloeistof of
         gas; de geschiedenis van het onderzoek; nieuwe onderzoeksimpulsen
         door het model van Per Bak en de chaostheorie; hoe men alleen de
         complexiteit begrijpt en worstelt om een goede fundamentele theorie
         op te stellen; dr. Eric Clement, fysicus Universita Pierre et Marie
         Curie, toont een aantal proeven waarbij wiskundige vormen in
         korrelstructuren ontstaan oa die laat zien hoe moeilijk het is
         korrels van verschillende grootte met elkaar te mengen, vergelijkt
         zijn proeven met die van Faraday en vertelt over de oorzaken van
         de wiskundige vormen en hoe moeilijk het in de industrie is met
         korrelstructuren te werken oa het mengen bij de aanmaak van beton.
         Parijs: Clement op scooter door stad, bij cementfabriek en igm
         collega op werk; cementfabriek; souvenirdoosje met dwarrelsneeuw;
         zandloper wordt omgedraaid en loopt, vallend zand;
     </description>
     <description_abstract>Wetenschappelijk magazine met uiteenlopende
         onderwerpen. Programma met reportages over wetenschappelijke
         onderwerpen. Deze aflevering gaat over het zoeken naar
         natuurkundige wetmatigheden in korrelstructuren zoals zand.
     </description_abstract>
     <publisher>VPRO</publisher>
[...]
```

Figure 2: Fragment of a *metadata* file

With regard to the ASR files, we have extracted the content of the *FreeTextAnnotation* labels, generating a TREC file per document. Besides, we have added the content of the *description_abstract* labels from the *metadata* files. The collection of training documents was processed filtering the stop words and applying a stemmer. Because all the original files are in Dutch language, we have used the Snowball stop word list for Dutch[3], which contains 101 stop words, and the Snowball Dutch stemmer[4].

3 Using SVM as text classifier

Automatic tagging of videos with subject labels can be seen as a categorization problem, using the speech transcriptions of the test videos like the documents to classify. One of the successful uses of SVM algorithms is the task of text categorization into fixed number of predefined categories based on their content. Commonly utilized representation of text documents from the field of

[2] http://www-nlpir.nist.gov/projects/trecvid/
[3] http://snowball.tartarus.org/algorithms/dutch/stop.txt
[4] http://snowball.tartarus.org/algorithms/dutch/stemmer.html

Experiment	MAP	R-prec
Using metadata	**0.0028**	0.0089
Without using metadata	0.0023	0.0061

Table 1: SINAI results at VideoCLEF 2009

information retrieval (IR) provides a natural mapping for construction of Mercer kernels utilized in SVM algorithms.

For the experiments and analysis carried out in this paper, the Rapid Miner[5] framework was selected. This toolkit provides several machine learning algorithms such as SVM and techniques along with other interesting features.

4 Experiments and results

The experiments carried out in this paper are a first approximation to the automatic tagging of videos using a text classifier. Two experiments have been submited: using the metadata files provided by the VideoCLEF organization and without using them, during the generation of the training corpus. The results obtained are showed in Table 1.

In order to evaluate the quality of the results, we have used two usual measures: the Mean Average Precision (MAP) and the R-precision. Analyzing the results, we can see that the use of metadata during the generation of the training corpus improves about 21.7% the average precision of the classification of the test videos.

5 Conclusions

The use of metadata as an valuable source of information in text categorization has been already applied some time ago, for example, in the categorization of full-text papers enriched by its bibliographic records [7].

We expect to continue this work by applying a multi-label classifier, instead the multiclass SVM algorithm used so far.

Acknowledgements

This work has been supported by the Regional Government of Andalucía (Spain) under excellence project GeOasis (P08-41999), under project on Tourism (FFIEXP06-TU2301-2007/000024), the Spanish Government under project Text-Mess TIMOM (TIN2006-15265-C06-03) and the local project RFC/PP2008/UJA-08-16-14.

References

[1] Díaz-Galiano, M.C., García-Cumbreras, M.A., Martín-Valdivia, M.T., Montejo-Ráez, A. and Ureña-López, L.A. Using Information Gain to Improve the ImageCLEF 2006 Collection. In Carol Peters, Paul Clough, Fredric C. Gey, Jussi Karlgren, Bernardo Magnini, Douglas W. Oard, Maarten de Rijke, and Maximilian Stempfhuber, editors, *CLEF*, volume 4730 of *Lecture Notes in Computer Science*, pages 711–714. Springer, 2006.

[2] Díaz-Galiano, M.C., García-Cumbreras, M.A., Martín-Valdivia, M.T., Montejo-Ráez, A. and Ureña-López, L.A. Integrating MeSH Ontology to Improve Medical Information Retrieval. In Carol Peters, Valentin Jijkoun, Thomas Mandl, Henning Müller, Douglas W. Oard, Anselmo

[5] Rapid Miner is available from `http://rapid-i.com/`

Peñas, Vivien Petras, and Diana Santos, editors, *CLEF*, volume 5152 of *Lecture Notes in Computer Science*, pages 601–606. Springer, 2007.

[3] Díaz-Galiano, M.C., García-Cumbreras, M.A., Martín-Valdivia, M.T., Ureña-López, L.A. and Montejo-Ráez, A. SINAI at ImageCLEFmed 2008. In Carol Peters, editor, *Proceedings of the Cross Language Evaluation Forum (CLEF 2008)*, 2008.

[4] Díaz-Galiano, M.C., Perea-Ortega, J.M., Martín-Valdivia, M.T., Montejo-Ráez, A. and Ureña-López, L.A. SINAI at TRECVID 2007. In Paul Over, editor, *Proceedings of the TRECVID 2007 Workshop (TRECVID 2007)*, 2007.

[5] Joachims, T. Text categorization with support vector machines: learning with many relevant features. In Claire Nédellec and Céline Rouveirol, editors, *Proceedings of ECML-98, 10th European Conference on Machine Learning*, number 1398, pages 137–142, Chemnitz, DE, 1998. Springer Verlag, Heidelberg, DE.

[6] Larson, M., Newman, E. and Jones, G. Overview of VideoCLEF 2009: New Perspectives on Speech-based Multimedia Content Enrichment. In Francesca Borri, Alessandro Nardi, and Carol Peters, editors, *Working Notes of CLEF 2009*, September 2009.

[7] Montejo-Ráez, A., Ureña-López, L.A. and Steinberger, R. Text categorization using bibliographic records: beyond document content. *Sociedad Española para el Procesamiento del Lenguaje Natural*, (35), 2005.

[8] Perea-Ortega, J.M., Montejo-Ráez, A., Martín-Valdivia, M.T., Díaz-Galiano, M.C. and Ureña-López, L.A. SINAI at VideoCLEF 2008. In Carol Peters, editor, *Proceedings of the Cross Language Evaluation Forum (CLEF 2008)*, 2008.