

USING ASSOCIATION RULE MINING TO ENRICH SEMANTIC CONCEPTS FOR VIDEO RETRIEVAL

Nastaran Fatemi, Florian Poulin, Laura E. Raileanu
HEIG-VD, Univ. of Applied Science of Western Switzerland, Yverdon-les-Bains, Switzerland

Alan F. Smeaton
CLARITY: Centre for Sensor Web Technologies, Dublin City Univ., Ireland

Keywords: Mining multimedia data, Association rule mining, Video indexing, Inter-concept relations.

Abstract: In order to achieve true content-based information retrieval on video we should analyse and index video with high-level semantic concepts in addition to using user-generated tags and structured metadata like title, date, etc. However the range of such high-level semantic concepts, detected either manually or automatically, is usually limited compared to the richness of information content in video and the potential vocabulary of available concepts for indexing. Even though there is work to improve the performance of individual concept classifiers, we should strive to make the best use of whatever partial sets of semantic concept occurrences are available to us. We describe in this paper our method for using association rule mining to automatically enrich the representation of video content through a set of semantic concepts based on concept co-occurrence patterns. We describe our experiments on the TRECVID 2005 video corpus annotated with the 449 concepts of the LSCOM ontology. The evaluation of our results shows the usefulness of our approach.

1 INTRODUCTION AND CONTEXT

Indexing video with high-level semantic concepts, such as events, locations, activities, people or objects is an important requirement of video retrieval applications if we are to move beyond information retrieval on video based on user tags and structured metadata. Currently, mainstream video access is based around exploiting video metadata and user-generated content such as tags and text comments and such approaches are the basis for systems such as YouTube. During the last decade, the automatic identification of high level semantic concepts in video has received a lot of attention yet has still proven to be hard to achieve with good degrees of reliability and is not present in mainstream video search.

A common approach to concept detection has been to model them based on the occurrence of low-level visual features such as colour, texture and motion, in sets of both positive and negative example video segments, relative to the high level concept. Approaches using Support Vector Machines (SVMs) (Ebadollahi et al., 2006), Hidden Markov Models

(HMMs) (Dimitrova et al., 2000), and Bayesian Networks (BN) (Pack and Chang, 2000) for example have been used to train semantic concept classifiers based on low-level features as inputs. Generally, these have varying degrees of accuracy and reliability and hence varying usefulness in the retrieval process. Manual annotation is normally still used for semantic concepts and in applications such as broadcast news or film archives, video retrieval is based almost exclusively on manual annotation of video coupled with structural metadata (Smeaton, 2007). Manual annotations have their own drawbacks. In (Volkmer et al., 2005) Volkmer & al. highlight the difficulty and the timeliness of manually annotating video. They argue that the efficiency of manual annotation is still lower for more complex concepts. However, image and video sharing and retrieval tools such as Flickr and YouTube, and the expansion of social Web applications have created new incentives in generating an increasing amount of manual annotations, or tags. These annotations are still strongly dependent on the users, the annotation context and the subsequent application. In an information retrieval context, applications usually have to do with a “partial set of annotations”, partial in the

sense that the annotation set is incomplete in its description of video content when considering the vocabulary of possible concepts. This problem is also relevant when tackling automatically generated annotations. Hence, even though there has been a lot of work to improve the performance of automatic or manual concept detection, our goal should be to make the best use of whatever partial sets of semantic concepts are available in order to maximise overall retrieval potential. Furthermore, one thing that we can leverage is that we know that there exist inter-concept relationships based on how they are used in indexing, even when there are only a small number of them. In this paper we address the issues of exploiting such relationships, increasing the coverage and quality of a partial set of semantic concepts for video archives, not by building new automatic classifiers but by leveraging existing ones and mining previous annotations using association rules.

The paper is organised as follows. In the next section we present a brief summary of related work. In section 3 we introduce the video corpus we used and our approach to enriching the set of semantic concepts used to index it. Section 4 presents the results of our experiments, including an evaluation of the impact of enrichment from association rule mining and the impact of removing synonym concepts. Finally section 5 presents our conclusions.

2 RELATED WORK

A number of recent research works have studied the question of how to best exploit a partial set of video annotations in the retrieval process. (Hauptmann et al., 2007) discuss finding a trade-off between the number of concepts to be detected and the utility of these concepts in the retrieval process. They show that only a few thousand semantic concepts could be sufficient to support highly accurate video retrieval and argue that when sufficiently many concepts are used, even low detection accuracy can potentially provide good retrieval results if the concepts are combined in a reasonable way. (Lin and Hauptmann, 2006) identify the semantic concepts of a large scale ontology which are likely to benefit many queries. They show that frequent concepts play a more vital role in video retrieval than rare concepts. Unlike rare concepts that benefit none or only one specific topic, frequent concepts can help multiple searches, either by filtering out irrelevant results, or by promoting relevant shots.

(Koskela and Smeaton, 2006) highlight the importance of inter-concept relations for semantic analysis in multimedia repositories and propose several meth-

ods to analyse concept similarities. They measure visual similarity by comparing the statistical distribution of the concept-models (clusters or latent aspects), trained from low-level features. The co-occurrence of pairs of concepts is calculated using the distance between the vectors of concept occurrences in the shots. Semantic similarity between concepts is then calculated based on human assessments and hierarchical relations are exploited based on a manually constructed hierarchical ontology. The authors reveal the usefulness of co-occurrence of concepts in the context of assisted annotation and automatic concept detection. (Garnaud et al., 2006) present an application of co-occurrence relations in an assisted video annotation tool, comparing different approaches to assist concept annotation. They evaluate the ability of an untrained user to perform fast and exhaustive annotation and conclude that concept recommendation based on co-occurrence, gives best results.

More traditional data mining techniques can be used to discover patterns of semantic concept use in video which may benefit subsequent video information retrieval. (Xie and Chang, 2006) describe their application of four data mining techniques: frequent itemset mining, k-means clustering, hidden Markov modeling and hierarchical hidden Markov models (HHMMs). They use the TRECVID'05 corpus annotated with 39 LSCOM-Lite concepts and evaluate the discovered patterns using a 192-concept subset of LSCOM. They highlight the difficulties of computational load when using frequent itemsets and conclude that HHMMs has the best average prediction among all models, but that different models seem to excel for different concepts depending on the concept prior and the ontological relationship. However the impact of different ontological relationships on the obtained results is unknown.

(Yan et al., 2006) use probabilistic directed and undirected graphical models to mine the relationships between video concepts. Their experiments, also on the TRECVID'05 development data, show the effectiveness and potential of using undirected models to learn concept relations. (Zha et al., 2007) propose to refine video annotations by exploiting pairwise concurrent relations among semantic concepts. They construct a concurrent matrix to explicitly represent such relations. Through spreading the scores of all related concepts to each other iteratively, detection accuracy is improved.

In recent work of (Dasiopoulou et al., 2008), a reasoning framework based on fuzzy description logics is used to enhance the extraction of image semantics. Explicit semantic relationships among concepts are represented using assertions of description log-

ics. Automatic detection is realised at both object and scene levels. The reasoning framework is then used to infer new concepts and to resolve inconsistencies in order to lead into a semantically meaningful description of the images.

(Ken-Hao et al., 2008) propose the use of association and temporal rule mining for post-filtering results obtained by automatic concept detectors. This work has a different goal and approach comparing to ours. (Ken-Hao et al., 2008) use association and temporal rules to improve the performance of automatic detectors. They discover association rules on the TRECVID'05 corpus annotated with 101 concepts of the MediaMill ontology. Among the 101 concepts of the MediaMill vocabulary, they find 32 concepts that have statistically significant rules for inference. When applying merely association rules (with no temporal smoothing) the results of automatic detection are improved for 24 concepts.

In our work we apply association rules on manually detected concepts with the goal of concept enrichment, i.e. discovering new concepts. We focus on a large vocabulary, the LSCOM 449 concepts. Our results show that association rule mining can be used with reasonable computational cost to automatically add new concepts with good performance.

3 SEMANTIC CONCEPT ENRICHMENT

3.1 Video Corpus

We use an older TRECVID corpus from 2005 (TRECVID'05) (Over et al., 2005a) because it has been manually annotated using semantic concepts from the Large Scale Concept Ontology for Multimedia (LSCOM) (Naphade et al., 2006). TRECVID'05 consists of a collection of broadcast TV news videos captured in October and November 2004 from 11 broadcasting organisations. We use the development set of TRECVID'05 consisting of 80 hours of video segmented into 43,907 shots. A detailed description of the corpus can be found in (Over et al., 2005b). LSCOM is a large multimedia concept lexicon with more than 1,000 concepts defined of which 449 have been used to annotate the TRECVID'05 development set in an effort by TRECVID participants (Volkmer et al., 2005).

Annotations on the TRECVID'05 corpus are done at the shot level. For each of the 43,907 shots and each of the 449 LSCOM concepts, there are three possible types of judgment namely “positive” (meaning

that the concept appears in the shot), “negative” (the concept does not appear in the shot), or “skip” (the shot remains unannotated). In our analysis we only take into account the positive and the negative judgments. For each shot the judgments are done only on their selected keyframes, which raises the issue of how representative these keyframes actually are, but that is beyond the scope of the present work.

3.2 Semantic Concept Mining and Enrichment

The use of a rich set of semantic annotations holds much potential for video retrieval and the goal of *enrichment* is to add new semantic concepts, automatically derived from the set of existing ones, to the video index. To determine new ones to be added to a video's representation, our approach is to use *association rule mining*. We proceed in two steps. In the first step we discover association rules among semantic concepts used in video indexing by analysing co-occurrences of the concepts in a set of fully annotated video shots. In the second step, we automatically enrich concepts. i.e. using the previously discovered association rules we automatically derive semantic concepts missing from the original annotation. We now describe these two steps in detail.

3.2.1 Mining Rules from Fully Annotated Shots

Association rules are used to identify groups of data items that typically co-occur frequently. They can reveal interesting relationships between items and can be used to predict new ones. Mining association rules has previously been applied to information stored in databases. Datasets in which an association rule is to be found is viewed as a set of tuples, where each tuple contains a set of items. Let $I = \{i_1, i_2, \dots, i_m\}$ be a set of items and $D = \{t_1, t_2, \dots, t_n\}$ a database of transactions, where $t_i = \{i_{i1}, i_{i2}, \dots, i_{ik}\}$ and $i_{ij} \in I$. An *association rule* is an implication of the form $A \Rightarrow B$, where $A, B \subset I$ are sets of items called *itemsets* and $A \cap B = \emptyset$. The *support* of an item (or set of items) is the percentage of transactions in which that item (or items) occurs. The *support* for an association rule $A \Rightarrow B$ is the percentage of transactions in the database that contain both A and B . The *confidence* for an association rule $A \Rightarrow B$ is the ratio of the number of transactions that contain both A and B to the number of transactions that contain A .

The association rule problem is to identify all association rules that satisfy a user specified *minimum support* and *minimum confidence* and this is solved in two steps. Firstly, all itemsets whose support is

greater than the given minimum are discovered and these are called *frequent itemsets*. Frequent itemsets are then used to generate interesting association rules where a rule is considered as *interesting* if its confidence is higher than the minimum confidence.

We use *Frequent Pattern Trees* (Han et al., 2000), to mine the frequent itemsets and then apply the *genRules* algorithm (Agrawal and Srikant, 1994) to these to generate interesting association rules. As the computational complexity of *genRules* depends largely on the maximum number of items in the rules, we simplified the algorithm by calculating only those rules which have a single item in the consequent. This is a sensible approach as in the following step we then enrich the absent concepts, one by one. Our simplified approach is presented in Algorithm 1 is not recursive.

Algorithm 1. *Simplified genRules algorithm.*

In : C : target concept for enrichment
In : m : minimum confidence threshold
In : F : list of frequent itemsets
Out : R : list of association rules having C
as consequent and confidence $\geq m$

```

for all  $i \in F$  do
  if  $C \in i$  and  $\text{size}(i) > 1$  then
    if  $\text{Support}(i)/\text{Support}(i - C) \geq m$  then
       $R = R \cup (i - C \Rightarrow C)$ 
    end if
  end if
end for

```

3.2.2 Applying Rules to Partially Annotated Shots

We treat each shot as a transaction and the set of semantic concepts corresponds to the set of items, I . Therefore, any subset of semantic concepts annotated positively in a shot corresponds to an itemset. Alg. 2 presents our process for video concept enrichment.

Algorithm 2. *Enrichment algorithm.*

In : C : target concept for enrichment
In : T : list of altered transactions (i.e. without C)
In : R : list of rules sorted by
decreasing value of confidence
Out : E : list enriched by applying R to T

```

for all  $t \in T$  do
   $e \leftarrow$  copy of  $t$ 
  for all  $r \in R$  do
    if antecedant of  $r \subseteq t$  then
       $e = e \cup C$ 
      skip the remaining rules
    end if
  end for
end for

```

end for

$E = E \cup e$

end for

This uses an *altered set* of transactions, which consists of the set of shots annotated with all the concepts except C which corresponds to the target of the enrichment algorithm. For each transaction of the altered set, we apply the discovered association rules having C as consequent, if the transaction contains the antecedant of the rule. For example, a rule $A, B \Rightarrow C$ would be used to add C to every transaction containing A and B . The resulting set of transactions is called the *enriched set*. Association rules used are sorted in decreasing value of their confidence which allows us to stop iterating as soon as a rule is applied.

4 EXPERIMENTS AND RESULTS

4.1 Cross-validation

We used a 5-fold cross-validation approach, meaning that the whole corpus of 43,907 transactions was randomly partitioned into five subsamples, a single part, the *validation set*, was retained and the remaining four parts, the *training set*, were used as training data. The process was repeated five times, with each of the five parts used once as validation data. Results were calculated as an average of the five rotations.

4.2 Enrichment Evaluation

To measure the enrichment performance of a concept, we calculate two measures, precision and recall. We compare two sets of transactions: the totally annotated set of transactions, T_{ref} , and the set of transactions after the enrichment, T_{enr} . T_{ref} corresponds to the validation set, and T_{enr} corresponds to the enriched set. A transaction in T_{enr} is relevant to a concept C , if it contains C and its corresponding transaction in T_{ref} contains the concept too. A transaction in T_{enr} is non-relevant, if it contains C , but its corresponding transaction in T_{ref} does not contain the concept. Precision corresponds to the number of relevant enriched transactions divided by the total number of transactions. Recall corresponds to the number of relevant transactions divided by the number of all transactions in T_{ref} containing the concept. We note that, depending on the value of the minimum confidence used in *genRules*, different values of precision and recall are obtained for the same concept. This can be explained by the fact that the higher the confidence of applied rules, the lower the probability of producing false positive enriched transactions,

which in turn causes a higher precision. At the same time, the lower the minimum confidence threshold, the bigger the number of applied rules and therefore the greater the probability to generate true positive transactions. This in turn causes higher recall. We therefore calculate different pairs of precision-recall values for different values of minimum confidence thresholds. Also, in order to measure the impact of the enrichment in terms of both recall and precision, we use the harmonic mean of these values, the *F-Measure*. We choose the best F-measure value obtained for each concept enrichment as a representative evaluation measure. In practice, the question of whether to favour precision or recall would be decided by the application.

4.3 Description of Results Graphs

We generated results representing recall, precision and the F-measure for the enriched concepts. In these experiments the minimum support used by the FPTree algorithm is fixed to 0.001 which allows us to produce a wide range of frequent itemsets (including concepts appearing relatively rarely in the collection) within reasonable computation time. With such a value, concepts appearing at least in 36 shots of the training set are taken into account, corresponding to 285 of the LSCOM concepts. The complete set of graphs are provided in (Fatemi et al., 2007). As an example to illustrate our findings, we use the concept *Soldiers*. An excerpt of the rules derived to enrich this concept is presented below:

$$\begin{aligned} \{Person, Military_Personnel, Rifles\} &\Rightarrow Soldiers \\ \{Person, Rifles, Military_Base\} &\Rightarrow Soldiers \\ \{Person, Military_Personnel\} &\Rightarrow Soldiers \end{aligned}$$

Table 1 shows an excerpt of the concepts used for the enrichment of the LSCOM concept *Soldiers* and their distributions in the applied rules. To correctly interpret these distribution values, when enriching a transaction with a given concept selected from among the applicable rules only the most confident rules will genuinely imply the concept. Figure 1 shows the evaluation of enrichment performance for this concept. The x-axis presents the minimum confidence values ranging from 0.5 to 1.0. As can be seen, for the concept *Soldiers* the graph shows that the best recall value, is obtained when a minimum confidence of 0.5 is used, while in terms of precision the best obtained value is when a minimum confidence of 1.0 is used. The F-measure combines both precision and recall is more representative if recall and precision are equally significant and so when analysing the results we focus on the best F-measure obtained among F-measures

Table 1: Concepts used to enrich *Soldiers*.

Percentage	Concept
86.4%	<i>Person</i>
74.2%	<i>Military_Personnel</i>
34.0%	<i>Rifles</i>
25.5%	<i>Armed_Person</i>
...	...

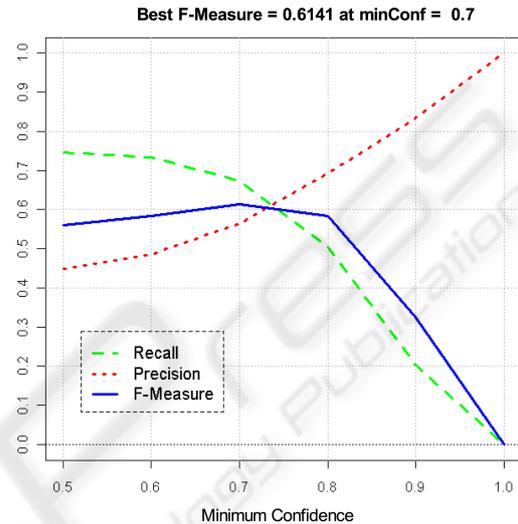


Figure 1: Performance of the *Soldiers* enrichment.

for different minimum confidence values. As shown in the graph of the Figure 1, for the concept *Soldiers* the best F-measure is obtained with a minimum confidence value of 0.7 and is 0.6141.

4.4 Impact of Removing Synonym Concepts

As can be seen in Table 1, among the concepts used frequently in the enrichment of *Soldiers* we find the concept *Military_Personnel* which is contained in 74.2% of the applied rules. This concept is indeed a synonym of *Soldiers*. We consider that the automatic enrichment of a concept is interesting if the video does not already contain synonyms of the concept. In order to examine the impact of the synonymy relations on the final results, we produced all the enrichment results in two ways: firstly by taking into account all concepts, and secondly by ignoring synonyms of the concepts in the collection. In the latter case, we kept only the most frequent concept in each synonym group and in order to determine the synonym groups we used an automatic technique based on association rule mining (Fatemi et al., 2007).

To globally analyse the impact of synonyms on the performance of the enrichment, we compared

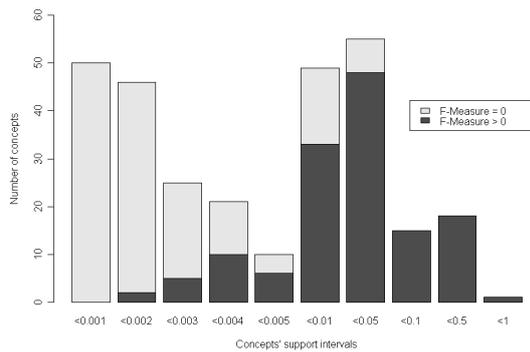


Figure 2: Distribution of the best F-Measures over support values.

the values of precision, recall and F-measure of all enriched concepts before and after removal of synonyms. We distinguish two kinds of enriched concepts: those involved in synonymy relations, and those which are not. For concepts not involved in a synonymy relation, there is no real difference between the quality of results obtained whether keeping the synonyms or ignoring them, whereas for concepts involved in a synonymy relation there is a significant difference. For example, the best F-measure of the *BoatShip* concept, is 0.5617 when considering all the concepts, and 0.5578, when ignoring synonyms.

We compared the best F-measure scores of these categories of concepts before removing their synonyms to those obtained after removing their synonyms and we obtained an average decrease of 30%. An example of a concept whose enrichment is totally biased when keeping synonyms is *Backpack*. When keeping synonyms, the *Backpack* concept is enriched with a recall and a precision of 1.0. When ignoring synonyms, recall falls to 0.0. This is due to the distribution of the concept *Backpack* and its synonym *Backpackers*, in the corpus. In fact, the only association rule employed to enrich *Backpack* is the rule *Backpackers* \Rightarrow *Backpack*. These two concepts always appear together. This illustrates to what point keeping synonyms can actually bias the results so for our set of enrichment experiments, we ignored them.

4.5 Overall Performance Evaluation

Among the 285 enriched concepts, 137 obtain a positive F-measure and 148 concepts an F-measure equal to 0. Figure 2 shows the distribution of the support intervals of the 285 concepts. This figure distinguishes between the concepts with positive best F-measure and those with 0 F-measure. The x-axis corresponds to the classes of support values. The classes were chosen to reflect the distribution of the support values of concepts in the corpus and have a non-uniform

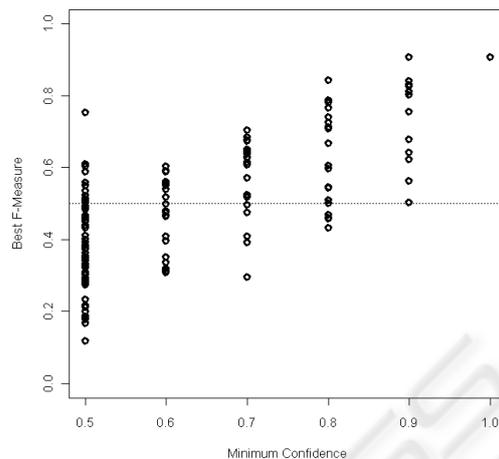


Figure 3: Distribution of the best F-Measures over minimum confidence values.

distribution. There are concepts having a very low support i.e ≤ 0.005 and concepts with a relatively high support i.e > 0.01 . The y-axis corresponds to the cumulated number of concepts. We observe that the majority of the 0 F-measure values correspond to concepts having low support values. More precisely, from among the 285 concepts considered in our experiments, 95% of concepts obtaining 0 F-measure have a support less than 0.01. Also 60% of the concepts obtaining a positive F-measure have a support greater than 0.01.

We then analysed the 137 enriched concepts having a positive F-measure. As discussed in 4.3, the values of the F-measures vary as a function of the minimum confidence of the applied rules. The best F-measures of concept enrichments are obtained at different minimum confidence values. Figure 3 illustrates the distribution of the best F-measures over the minimum confidence values for the 137 concepts. We observe that the highest best F-measure values are obtained at highest minimum confidence values. Figure 4 shows the histogram of values of best F-measures obtained for the 137 enriched concepts having positive F-measure. For 66 concepts we obtain a best F-measure greater than 0.5, representing good performance quality. Table 2 provides the complete list of the 137 concepts enriched with positive F-measure grouped by their best F-measure value intervals.

5 CONCLUSIONS AND FUTURE WORK

We have presented an approach to automatically deriving semantic concepts from existing video annota-

Table 2: F-Measure scores for various concepts.

F-Measure	Concept
[0.9, 1.0]	<i>Civilian_Person, Weapons</i>
[0.8, 0.9]	<i>Residential_Buildings, Outdoor, Adult, Individual, Sports, News_Studio</i>
[0.7, 0.8]	<i>Face, Rifles, Ground_Vehicles, Suits, Waterscape_Waterfront, Athlete, Armored_Vehicles, Exploding_Ordinance, Ties, Female_Anchor, Vehicle</i>
[0.6, 0.7]	<i>Single_Person_Female, Indoor_Sports_Venue, Female_Person, Armed_Person, Flags, Smoke, Car, Actor, Interview_Sequences, Road, Sitting, Male_Anchor, Singing, Soldiers, Building, Body_Parts, Speaking_To_Camera, Caucasians, Apartment_Complex, Group</i>
[0.5, 0.6]	<i>Grandstands_Bleachers, Head_And_Shoulder, Soccer, Explosion_Fire, Election_Campaign, Entertainment, Standing, Boat_Ship, Talking, Overlaid_Text, Ground_Combat, Address_Or_Speech, Head_Of_State, Furniture, Running, Sky, Government_Leader, Child, Microphones, Windows, Congressman, Airplane, Election_Campaign_Address, Attached_Body_Parts, Hu_Jintao</i>
[0.4, 0.5]	<i>Golf_Player, Golf_Course, George_Bush, Windy, Politics, Walking_Running, Non-us_National_Flags, Meeting, People_Marching, Commercial_Advertisement, Fields, Host, Crowd, Male_Reporter, Airplane_Flying, Lawn, Urban_Scenes, Dark-skinned_People, Single_Family_Homes, Dancing, Celebrity_Entertainment, Text_On_Artificial_Background, Business_People, Reporters, Female_Reporter, Asian_People, Office_Building, Computer_Or_Television_Screens, Guest</i>
[0.3, 0.4]	<i>Male_News_Subject, Interview_On_Location, Glasses, Streets, Fighter_Combat, Walking, Sidewalks, Police_Private_Security_Personnel, Commentator_Or_Studio_Expert, Funeral, Landscape, Scene_Text, Mug, Agricultural_People, Oceans, Boy, Telephones, Us_Flags, Weather, Newspapers, Greeting, Mountain, Vegetation, Yasser_Arafat, Dresses, Celebration_Or_Party, Muslims, Store, Dresses_Of_Women</i>
[0.2, 0.3]	<i>Flowers, First_Lady, Nighttime, Conference_Room, Girl, Cheering, Maps, Network_Logo, Truck</i>
[0.1, 0.2]	<i>Beards, Demonstration_Or_Protest, Hill, Room, Street_Battle, Computers</i>
[0.0, 0.1]	-

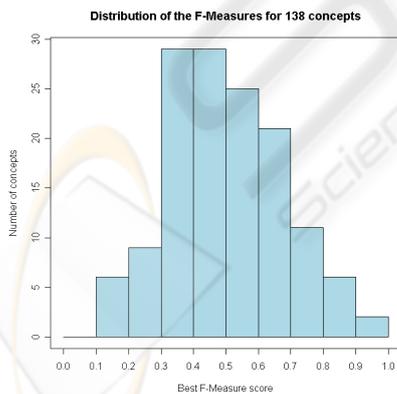


Figure 4: Distribution of the 137 concepts having a positive F-Measure.

tions by association rule mining. We described experiments on a video corpus manually annotated using a large scale multimedia concept ontology, LSCOM. In order to fairly evaluate our concept enrichment, we also studied the impact of synonymy relations in the

results obtained. We showed that enrichment experiments are biased for the concepts which have synonyms, and we realised experiments and evaluated the results when removing such synonyms. Analysis of the results showed that among the 449 LSCOM concepts that we studied, there are 66 which can be automatically enriched with an F-measure greater than 0.5. This suggest the usefulness of automatic enrichment in concept annotation for video.

There are several directions for us to continue this work. It would be useful to determine a minimal set of concepts which can benefit a maximum number of enrichments. Manual as well as automatic annotation systems can then focus only on detecting concepts of the minimal set and rely on the automatic enrichment procedure to derive the rest. Optimisation methods should determine the minimum number of concepts required to produce a maximum number of enrichments with good quality. Another possible way to improve results would be to mine association rules on a set of consecutive shots. The shot granularity used

in these experiments could be too small to discover all possible associations of concepts. This can be particularly penalising for concepts which have low support in the corpus. Finally, it would be interesting to realise the enrichment experiments based on both the absence and the presence of concepts in shots. This could increase the complexity of the procedure and would require new optimisations to keep the computation manageable.

ACKNOWLEDGEMENTS

This work is supported by the RCSO-TIC strategic reserve funds of Switzerland, under grant HES-SO/18453. Alan Smeaton is supported by Science Foundation Ireland under grant 07/CE/I1147.

REFERENCES

- Agrawal, R. and Srikant, R. (1994). Fast algorithms for mining association rules. In *Proc. 20th Int. Conf. Very Large Data Bases (VLDB)*, pages 487–499. MorganKaufmann.
- Dasiopoulou, S., I.Kompatsiaris, and M.G.Strintzis (2008). Using fuzzy dls to enhance semantic image analysis. In *SAMT'08: 3rd International Conference on Semantic and Digital Media Technologies*.
- Dimitrova, N., Agnihotri, L., Jasinschi, R., Zimmerman, J., Marmaropoulos, G., McGee, T., and Dagtas, S. (2000). Video scouting demonstration: smart content selection and recording. In *ACM Multimedia*, pages 499–500.
- Ebadollahi, S., Xie, L., Chang, S.-F., and Smith, J. R. (2006). Visual event detection using multi-dimensional concept dynamics. In *IEEE International Conference on Multimedia and Expo (ICME 06)*, Toronto.
- Fatemi, N., Raileanu, L., and Poulin, F. (2007). LSVAM - Large Scale Video Annotation Mining: Second part. *Internal Report ICT-COM/2007-3*.
- Garnaud, E., Smeaton, A. F., and Koskela, M. (2006). Evaluation of a video annotation tool based on the lscm ontology. In *SAMT 2006: Proceedings of The First International Conference on Semantics And Digital Media Technology*, pages 35–36.
- Han, J., Pei, J., and Yin, Y. (2000). Mining frequent patterns without candidate generation. In Chen, W., Naughton, J., and Bernstein, P. A., editors, *2000 ACM SIGMOD Intl. Conference on Management of Data*, pages 1–12. ACM Press.
- Hauptmann, A., Yan, R., and Lin, W.-H. (2007). How many high-level concepts will fill the semantic gap in news video retrieval? In *CIVR '07: Proceedings of the 6th ACM international conference on Image and video retrieval*, pages 627–634.
- Ken-Hao, L., Ming-Fang, W., Chi-Yao, T., Yung-Yu, C., and Ming-Syan, C. (2008). Association and temporal rule mining for post-processing of semantic concept detection in video. *IEEE Transactions on Multimedia, special issue on Multimedia Data Mining*, 10(2):240–251.
- Koskela, M. and Smeaton, A. F. (2006). Clustering-based analysis of semantic concept models for video shots. In *ICME*, pages 45–48.
- Lin, W.-H. and Hauptmann, A. (July 2006). Which thousand words are worth a picture? experiments on video retrieval using a thousand concepts. *Multimedia and Expo, 2006 IEEE International Conference on*, pages 41–44.
- Naphade, M., Smith, J. R., Tesic, J., Chang, S.-F., Hsu, W., Kennedy, L., Hauptmann, A., and Curtis, J. (2006). Large-Scale Concept Ontology for Multimedia. *IEEE MultiMedia*, 13(3):86–91.
- Over, P., Ianeva, T., Kraaij, W., and Smeaton, A. F. (2005a). TRECVID 2005 - An Overview. In *TRECVID 2005: Proceedings of the TRECVID Workshop*, Md., USA. National Institute of Standards and Technology.
- Over, P., Ianeva, T., Kraaij, W., and Smeaton, A. F. (2005b). Trecvid 2005 - an overview. In *Proceedings of TRECVID 2005*.
- Pack, S. and Chang, S.-F. (2000). Experiments in constructing belief networks for image classification systems. In *Proceedings of the International Conference on Image Processing*.
- Smeaton, A. F. (2007). Techniques used and open challenges to the analysis, indexing and retrieval of digital video. *Information Systems Journal*, 32(4):545–559.
- Volkmer, T., Smith, J., and Natsev, A. (2005). A web-based system for collaborative annotation of large image and video collections: an evaluation and user study. *Proceedings of the 13th annual ACM international conference on Multimedia*, pages 892–901.
- Xie, L. and Chang, S.-F. (2006). Pattern mining in visual concept streams. In *IEEE International Conference on Multimedia and Expo (ICME 06)*, Toronto.
- Yan, R., Yu Chen, M., and Hauptmann, A. G. (2006). Mining relationship between video concepts using probabilistic graphical models. In *ICME*, pages 301–304.
- Zha, Z.-J., Mei, T., Hua, X.-S., Qi, G.-J., and Wang, Z. (2007). Refining video annotation by exploiting pairwise concurrent relation. In *MULTIMEDIA '07: Proceedings of the 15th international conference on Multimedia*, pages 345–348.