VIDEO ANNOTATION PROPAGATION USING ACTIVE LEARNING

DAVID TEKSEN AKSUN *supervised by Prof. Dr. Zehra Çataltepe*
ISTANBUL TECHNICAL UNIVERSITY - *Faculty of Computer Engineering and Informatics*

TÜBİTAK

OBJECTIVES

To annotate videos effectively for person identification using active learning:

1. Implement a ITU Annot Framework for efficient development.
2. Query the system after active learning with time and labels.
3. Develop a faster method.
4. Get better performance.
5. Use a mathematically sound active learning procedure.

ACTIVE LEARNING SYSTEM

The general flow of active learning:

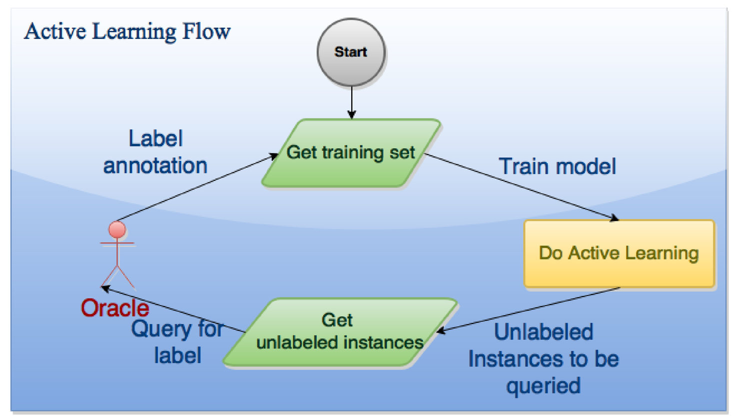

Figure 3: Active learning flow [1]

$F-measure$ is used for evaluating performance:

$$f-measure = 2\left(\frac{precision \cdot recall}{precision + recall}\right) \quad (1)$$

$$precision = \frac{|relevanttracks \cap retrievedtracks|}{|retrievedtracks|} \quad (2)$$

$$recall = \frac{|relevanttracks \cap retrievedtracks|}{|relevanttracks|} \quad (3)$$

INTRODUCTION

Many videos containing useful information are available on the web. Useful information such as the names of the people who have appeared or spoken should be known easily. Yet, human annotators are costly. In addition they are slow. Camomile project's goal is to find an efficient way to resolve the problem of human annotation.
In the project **Tensor MVS** is proposed and implemented. Tensor MVS uses minimum variance selection strategy and selects the track with the least amount of variance. Then that track is sent to the oracle for receiving annotation.

ANNOTATION BY TENSOR MVS

The system was tested on the REPERE dataset [2].

Figure 4: Repere Challange Data [2]

Given video tracks, propagate annotations using Tensor MVS. MVS is used for active learning selection. The most informative points are where the track variance is low; when variance is low, confusion between labels increases. Tensor is useful for propagation of labels by exploiting similarities between tracks. A sample result after running Tensor MVS is given in figure .

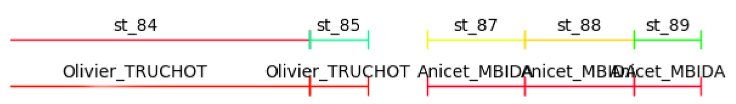

Figure 5: Labeling of tracks after completion of active learning

EXPERIMENTAL RESULTS

BCS, Hard-Hybrid methods are state of the art methods prior to Tensor MVS, while Random is used for benchmarking. In 20 steps optimality is reached [2]. Random selects random track, BCS selects from biggest cluster with least annotations, Hard-Hybrid uses BCS for the first n steps and MDS for the later $20 - n$ steps. MDS is a query by committee method using agglomerative clustering and k-medoids in combination with stable matching.

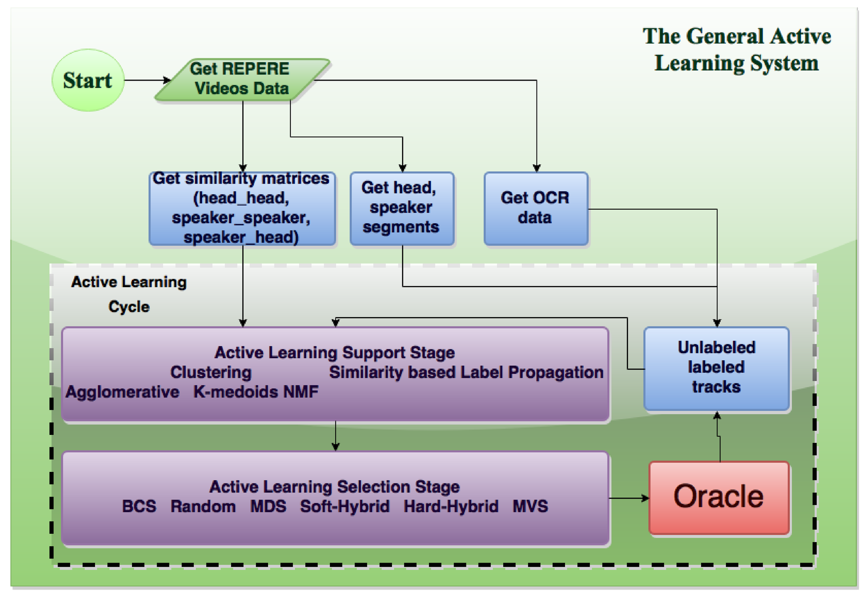

Figure 1: Active learning flow [1]

Data: segments, similarity matrices
Result: f-measure
preprocess;
for *step* $\leftarrow 1$ **to** 20 **do**
  tensor based label propagation;
  minimum variance selection;
  get label;
  calculate f-measure;
end
postprocess;

Algorithm 1: Tensor MVS

1. **MVS has better f-measure performance:** MVS performs in the range of 92.9%, while Hard-Hybrid method performs in the range of 89.8%.
2. **MVS has lower f-measure performance standard deviation:** MVS performs in the range of 0.29% at the 20^{th} step.
3. **MVS calculation can be done very fast:** MVS is 144 faster compared to Hard-Hybrid, Random , BCS! MVS finished in 150 seconds.
4. **MVS is inherently parallel:** All videos are analyzed independently and they can be parallelized.

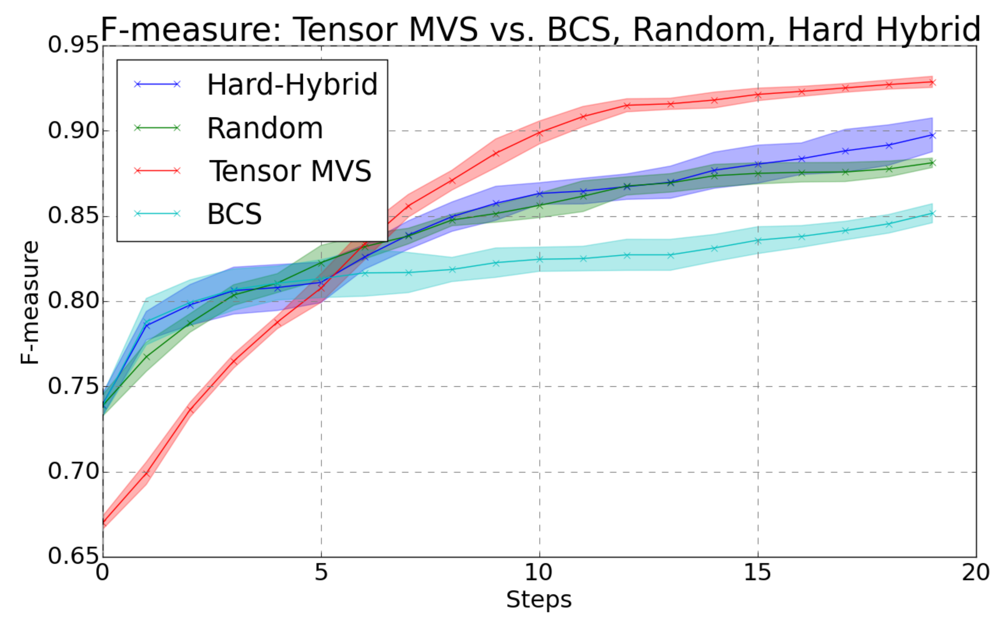

Figure 2: Comparison of the results of Tensor MVS method using speaker modality and speaker annotation for 20 steps

CONCLUSION

- The results clearly depict that Tensor MVS performs much better for 20 steps.
- Tensor MVS is dramatically faster compared to older methods. The speed can be enhanced with full parallelization.
- Analyzed videos can be queried via time and labels.
- The new framework allows new methods to be implemented easily.
- Each video can be analyzed independently.

REFERENCES

[1] B. Settles. *Active Learning*. Morgan & Claypool Publishers, 1st edition, 2012.

[2] M. Budnik, J. Poignant, L. Besacier and G. Quenot Automatic Propagation of Manual Annotations for Multimodal Person Identification in TV Shows. *Context Based Multimedia Indexing (CBMI)*, pp.1,4, June 2014.

FUTURE RESEARCH

camomile

The Camomile project will be over by September 2015. Yet many contributions can be done afterwards.
Different configurations of Tensor MVS can be analyzed using grid search for performance.
Collective NMF methods can be tested more effectively with different configurations. These methods can also be tested on videos with different languages.

ACKNOWLEDGEMENTS

I thank Prof. Dr. Zehra Çataltepe for her support and guidance. I thank Abdurrahman Namlı and Mateusz Budnik for their support. I also extend my gratitude to all the faculty members.
Research and work on this project is conducted under the CHIST-ERA Camomile project which is funded by TUBITAK project 112E176.