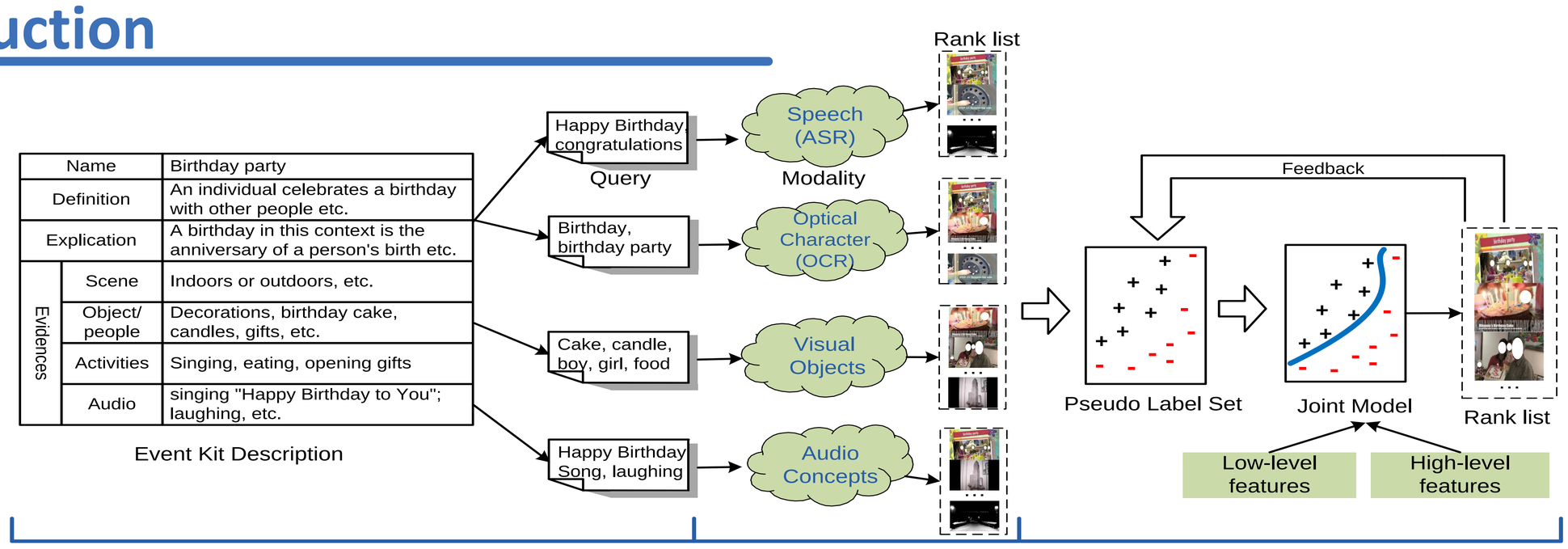
Zero-Example Event Search using MultiModal Pseudo Relevance Feedback (MMPRF)

Lu Jiang, Teruko Mitamura, Shoou-I Yu, Alexander G. Hauptmann
Carnegie Mellon University

Introduction



MMPRF significantly contributes to CMU Team's final submission in TRECVID-13 Multimedia Event Detection.

MMPRF in a nutshell:

- Construct a pseudo label set.
- Train a joint model on the pseudo label set using both high-level and low-level features.
- Feedback the rank list of the joint model to establish the pseudo label set for the next iteration.

Pseudo Label Construction

$$\arg\max_{\mathbf{y}} \sum_{i=1}^{m} \ln L(\mathbf{y}; \Omega, \Theta_i)$$

$$\text{s.t. } \mathbf{A}^T\mathbf{y} \leq \mathbf{g};\ \mathbf{y} \in \{0,1\}^{|\Omega|}$$

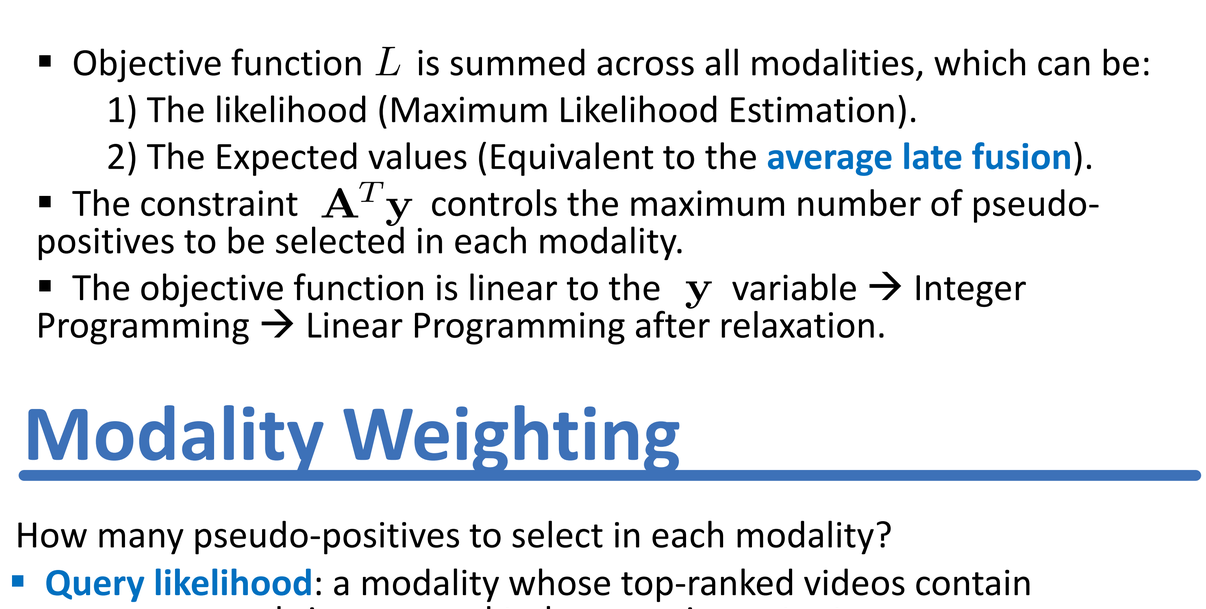
- Objective function L is summed across all modalities, which can be:
  1) The likelihood (Maximum Likelihood Estimation).
  2) The Expected values (Equivalent to the **average late fusion**).
- The constraint $\mathbf{A}^T\mathbf{y}$ controls the maximum number of pseudo-positives to be selected in each modality.
- The objective function is linear to the $\mathbf{y}$ variable → Integer Programming → Linear Programming after relaxation.

Modality Weighting

How many pseudo-positives to select in each modality?

- **Query likelihood**: a modality whose top-ranked videos contain more query words is supposed to be more important.
- Find **indicative words** in the event kit description. For example, the occurrence of words "narration/narrating" and "process" in the event kit description indicates an "accurate ASR event".

Pseudo Positive Videos



Experimental Results

Dataset: TRECVID (MED) 2013 development and MEDTest set.

Events	Method	Single split	Ten splits
Pre-Specified	Without PRF	3.9	4.9 ± 0.8
	Rocchio	5.7	7.4 ± 1.1
	Relevance Model	2.6	3.4 ± 0.5
	CPRF	6.4	8.3 ± 0.9
	Learning to Rank	3.4	4.2 ± 0.7
	MMPRF1	9.0	11.8 ± 1.1
	MMPRF2	**10.1**	**13.6 ± 1.2**
Ad-Hoc	Without PRF	4.0	6.4 ± 0.6
	Rocchio	5.6	6.3 ± 0.9
	Relevance Model	2.3	3.7 ± 0.8
	CPRF	5.9	9.1 ± 1.0
	Learning to Rank	4.3	6.0 ± 0.9
	MMPRF1	7.0	10.9 ± 1.0
	MMPRF2	**8.3**	**12.1 ± 1.1**

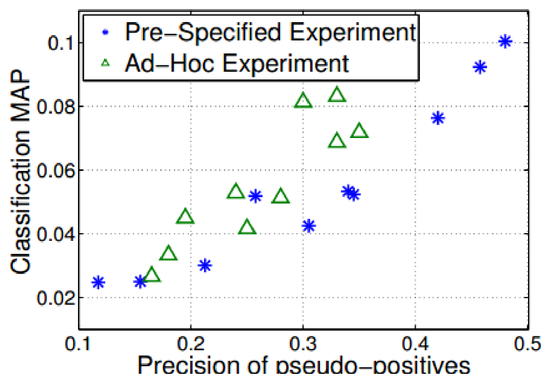
MMPRF1: w/o modality weighting. MMPRF2: w/ modality weighting.

Improve the baseline Without PRF:

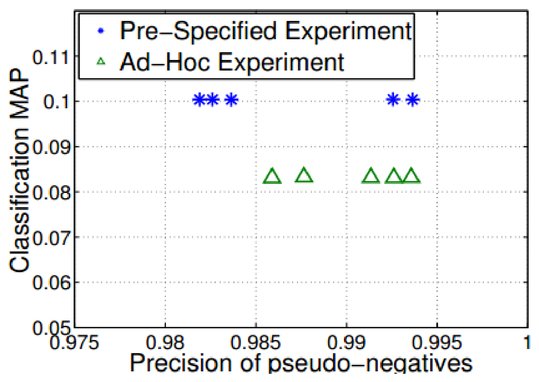
by a **relative 158%** (**absolute 6.2%**) on Pre-Specified events

by a **relative 107%** (**absolute 4.3%**) on Ad-Hoc events.

Statically significantly better than other baseline methods.



(a) Pseudo-positives



(b) Pseudo-negatives

Pseudo label set	Top k^+	Pre-Specified P@N	Pre-Specified MAP	Ad-Hoc P@N	Ad-Hoc MAP
Without PRF	-	-	3.90	-	4.00
ASR	10	0.34	5.33	0.28	5.13
ASR	20	0.26	5.18	0.20	4.49
OCR	10	0.42	7.63	0.33	6.88
OCR	20	0.35	5.23	0.24	5.28
SIN/DCNN	10	0.16	2.50	0.18	3.33
SIN/DCNN	20	0.12	2.48	0.17	2.67
Late Fusion	10	0.30	4.25	0.35	7.18
Late Fusion	20	0.21	3.00	0.25	4.16
MMPRF-2	10	0.48	**10.05**	0.33	**8.32**
MMPRF-2	20	0.45	9.23	0.3	8.13

Conclusions

- MultiModal Pseudo Relevance Feedback (MMPRF) is a **first** attempt to use both high-level and **low-level features** in MED EK0.
- MMPRF offers a solution to conduct PRF on **multiple ranked lists**. Empirically it significantly **outperforms all baseline methods** on MEDTest.
- Modality weighting is beneficial.

This work was partially supported by Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center contract number D11PC20068. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the U.S. Government.

Carnegie Mellon University