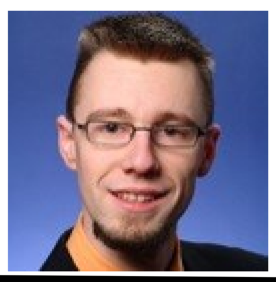
# Joint Optimization of User-desired Content in MDS by Learning from User Feedback

Avinesh P.V.S. and Christian M. Meyer

AIPHES
ADAPTIVE PREPARATION OF INFORMATION FROM HETEROGENEOUS SOURCES

UBIQUITOUS KNOWLEDGE PROCESSING

## Overview

**Motivation:**

- **No one best** summary for all needs
- **Low $\kappa$ of content selection**
- **Automatic methods** produce **low quality** summaries compared to humans

**Objective:**

- Creation of **user-desired summaries** using interactive learning methods.

**Contributions:**

- **Interactive loop** to integrate feedback
- **AL** and **joint optimization** techniques to collect user feedback

**Applications:** Journalistic aid, Interactive annotation tool

## Interactive Models

**Baseline Model***: max $\sum_i w_i c_i$

**Novel User Feedback Models**

**Accept Model: (ACCEPT)**

$$\forall i \in I_0^t. \quad w_i = MAX$$
$$\forall i \in Q_0^t - I_0^t. \quad w_i = 0$$

**Accepted concepts** **Ignored concepts**

**Joint ILP Model: (JOINT)**

Explore concepts **lacking feedback**

$$\max \sum_{i \notin Q_0^t} w_i c_i - \sum_{i \in Q_0^t} w_i c_i$$

**Lacking feedback** **Received feedback**

**Active Learning Uncertainty Model: (AL)**

Explore concepts with **high uncertainty**

$$\max \sum_{i \notin Q_0^t} u_i w_i c_i$$

**Uncertainty**

**Active Learning Certainty Model: (AL+)**

Explore concepts based on **positive prediction of acceptance of a concept**

$$\max \sum_{i \notin Q_0^t} l_i (1 - u_i) w_i c_i$$

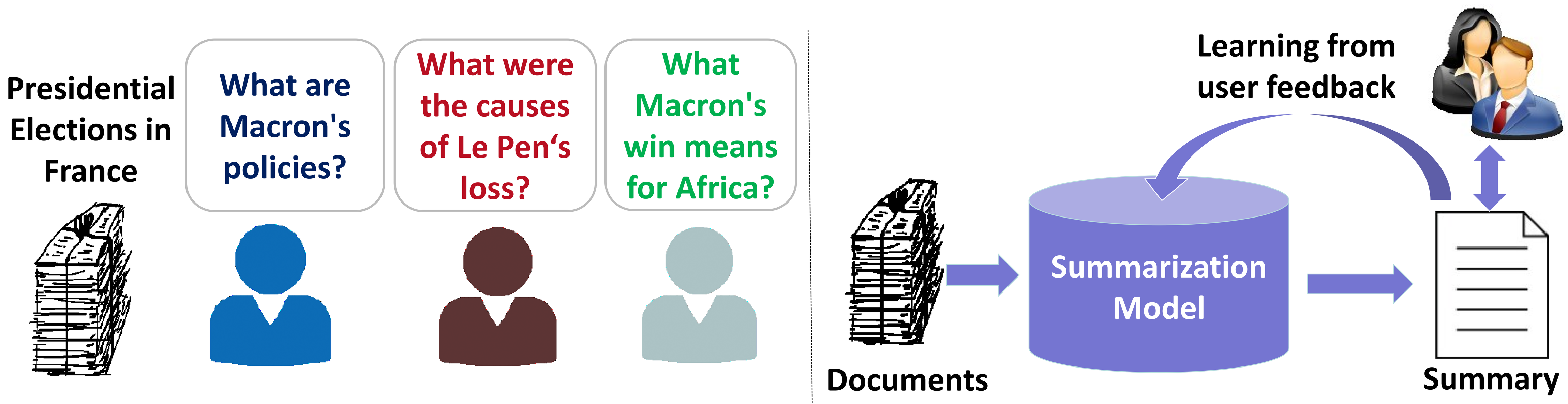
**Prediction** **Certainty**

$$\text{Where } l_i = \begin{cases} 0 & if\ f^{(t)}(\Phi(\tilde{x}_i)) = -1 \\ 1 & if\ f^{(t)}(\Phi(\tilde{x}_i)) = 1 \end{cases}$$

**Feature Function** **Concept Feature Vector**

*SOA ICSI system (Boudin et al. (2015 ))

## Interactive Personalized Summarization




**Generic summaries cannot satisfy all needs** **Interactive Personalized Summarization**

**Best of both the worlds: Automatic (System) and Manual (Human) Summarization**

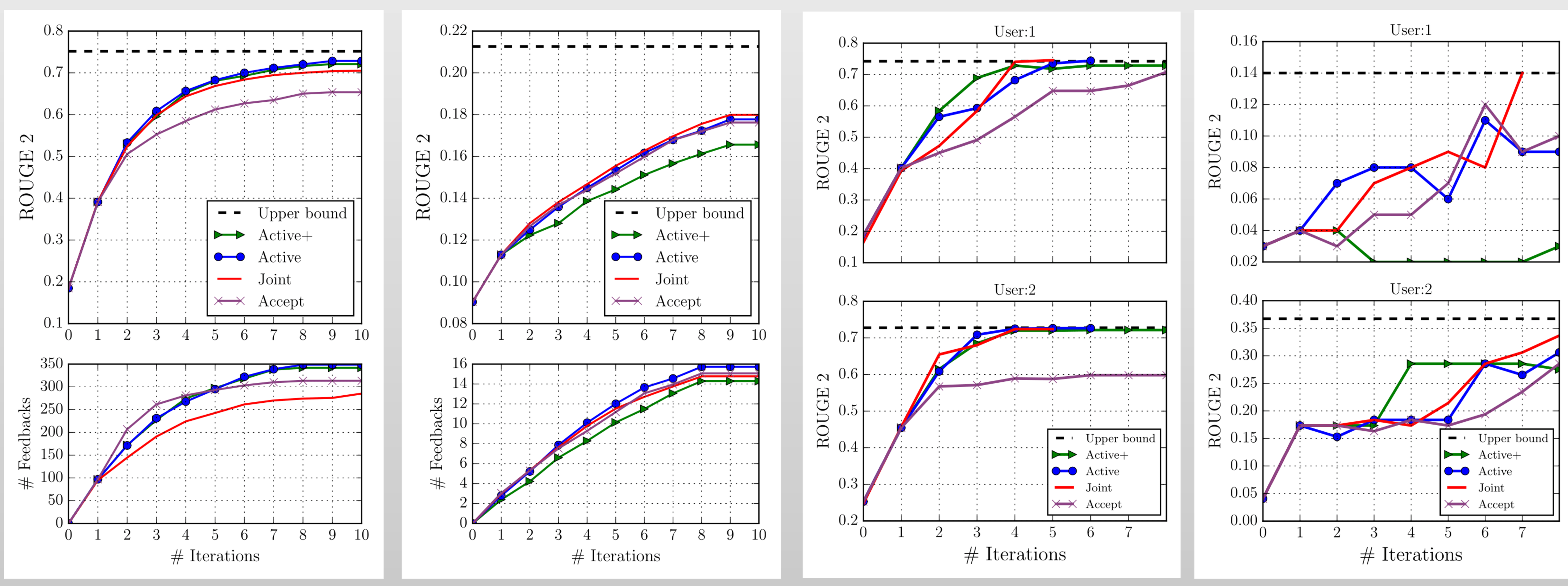
## Experiments & Analysis

Evaluate: The coverage of the user-desired content in the summary

**≈ Evaluate: To reach the upper bound for a user's reference summary**

| Datasets | ICSI | | | UB | | | Accept | | | JOINT | | | AL | | | AL+ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | R1 | R2 | SU4 | R1 | R2 | SU4 | R1 | R2 | SU4 | R1 | R2 | SU4 | R1 | R2 | SU4 | R1 | R2 | SU4 |
| **Concept Notion: Bigrams** | | | | | | | | | | | | | | | | | | |
| **DBS** | .451 | .183 | .190 | **.848** | **.750** | **.532** | .778 | .654 | .453 | .815 | .707 | .484 | **.833** | **.729** | .498 | .828 | .721 | **.500** |
| **DUC'04** | .374 | .090 | .118 | **.470** | **.212** | **.185** | .442 | .176 | .165 | **.444** | **.180** | **.166** | .440 | .178 | .160 | .427 | .166 | .154 |
| **DUC'02** | .350 | .085 | .110 | **.474** | **.216** | **.187** | .439 | .178 | .161 | .444 | .182 | .165 | **.448** | **.188** | .165 | **.448** | .184 | **.170** |
| **DUC'01** | .333 | .073 | .105 | **.450** | **.213** | **.181** | .414 | .171 | .156 | .418 | .167 | .149 | **.435** | **.186** | **.163** | .426 | .181 | .158 |
| **Concept Notion: Phrases** | | | | | | | | | | | | | | | | | | |
| **DBS** | .403 | .135 | .154 | **.848** | **.750** | **.532** | .691 | .531 | .430 | .742 | .597 | .419 | **.776** | **.652** | **.448** | .767 | .629 | .440 |
| **DUC'04** | .374 | .090 | .118 | **.470** | **.212** | **.185** | .441 | .176 | .160 | .441 | .179 | .162 | **.444** | **.180** | **.162** | .422 | .164 | .150 |
| **DUC'02** | .350 | .085 | .110 | **.474** | **.216** | **.187** | .436 | .181 | .162 | .444 | .183 | .165 | **.446** | **.185** | **.168** | .442 | .182 | .162 |
| **DUC'01** | .333 | .073 | .105 | **.450** | **.213** | **.181** | .410 | .165 | .153 | .417 | .170 | .156 | **.433** | **.182** | **.161** | .420 | .179 | .154 |

**DBS** **DUC'04** **DBS (7)** **DUC'04 (d30051t)**




## Conclusions

- **Interactively collecting feedback** steers a general summary to a **personalized summary.**
- **JOINT** model **consistently converges** to the upper bound with minimal feedback.
- **AL** model **balances well** the trade-off between **faster convergence** and **amount of feedback**.
- **AL+** model performs well when there is **sufficient amount of feedback.**
- **Future work: Sampling strategies** using AL and **propagation methodologies**

## Try it out, get in touch

**Code and data:** https://github.com/UKPLab/acl2017-interactive_summarizer

**Questions or comments**: avinesh@aiphes.tu-darmstadt.de

RUPRECHT-KARLS-UNIVERSITÄT HEIDELBERG
Heidelberger Institut für Theoretische Studien
HITS
TECHNISCHE UNIVERSITÄT DARMSTADT

**Acknowledgements:** This work has been supported by the German Research Foundation as part of the Research Training Group "Adaptive Preparation of Information from Heterogenous Sources" (AIPHES) under grant No. GRK 1994/1.