Different Contributions to Cost-Effective Transcription and Translation of Video Lectures

Thesis

Presented by:

Joan Albert Silvestre Cerdà

Supervisors:
Dr. Alfons Juan Císcar
Dr. Jorge Civera Saiz

Machine Learning and Language Processing
Departament de Sistemes Informàtics i Computació
Universitat Politècnica de València

January 27, 2016

Outline

Introduction

Explicit Length Modelling for SMT

Efficient Audio Segmentation for Speech Detection

The transLectures-UPV Platform

Recommender Systems for Online Learning Platforms

LM Adaptation Using External Resources for ASR

Demos

Conclusions

Outline

Introduction

Explicit Length Modelling for SMT

Efficient Audio Segmentation for Speech Detection

The transLectures-UPV Platform

Recommender Systems for Online Learning Platforms

LM Adaptation Using External Resources for ASR

Demos

Conclusions

2 Introduction

Explicit Length Modelling for SMT

Efficient Audio Segmentation for Speech Detection

The transLectures-UPV Platform

Recommender Systems for Online Learning Platforms

LM Adaptation Using External Resources for ASR

Demos

Conclusions

Introduction

- Internet has brought new opportunities to academic institutions.
- Multimedia repositories as fundamental knowledge assets.
- Subtitles are really needed in these repositories.
- Most repositories are neither transcribed nor translated.
- **Cost-effective** transcription and translation of video repositories.

3 Introduction

Explicit Length Modelling for SMT

Efficient Audio Segmentation for Speech Detection

The transLectures-UPV Platform

Recommender Systems for Online Learning Platforms

LM Adaptation Using External Resources for ASR

Demos

Conclusions

Scientific and Technological Goals

- To propose an approach to explicit length modelling for SMT.
- To develop efficient audio segmentation systems.
- To design a system architecture for ASR and SMT integration.
- To improve adaptation techniques for ASR and SMT.
- To design recommender systems using speech transcriptions.
- To evaluate these contributions in real-life scenarios.

Outline

Introduction

Explicit Length Modelling for SMT

Efficient Audio Segmentation for Speech Detection

The transLectures-UPV Platform

Recommender Systems for Online Learning Platforms

LM Adaptation Using External Resources for ASR

Demos

Conclusions

Introduction

5 Explicit Length Modelling for SMT

Introduction

Log-linear modelling

Experiments

Conclusions

Efficient Audio Segmentation for Speech Detection

The transLectures-UPV Platform

Recommender Systems for Online Learning Platforms

LM Adaptation Using External Resources for ASR

Demos

Conclusions

Introduction

- Length modelling is a well-known problem.
- Focus on **explicit** length modelling for SMT.
- Comparative study on phrase length modelling for SMT.
- Two novel length models for phrase-based SMT are presented.

Introduction

Explicit Length Modelling for SMT

6 Introduction

Log-linear modelling

Experiments

Conclusions

Efficient Audio Segmentation for Speech Detection

The transLectures-UPV Platform

Recommender Systems for Online Learning Platforms

LM Adaptation Using External Resources for ASR

Demos

Conclusions

SMT and Log-linear Modelling

Search the most likely translation $\hat{y}$:

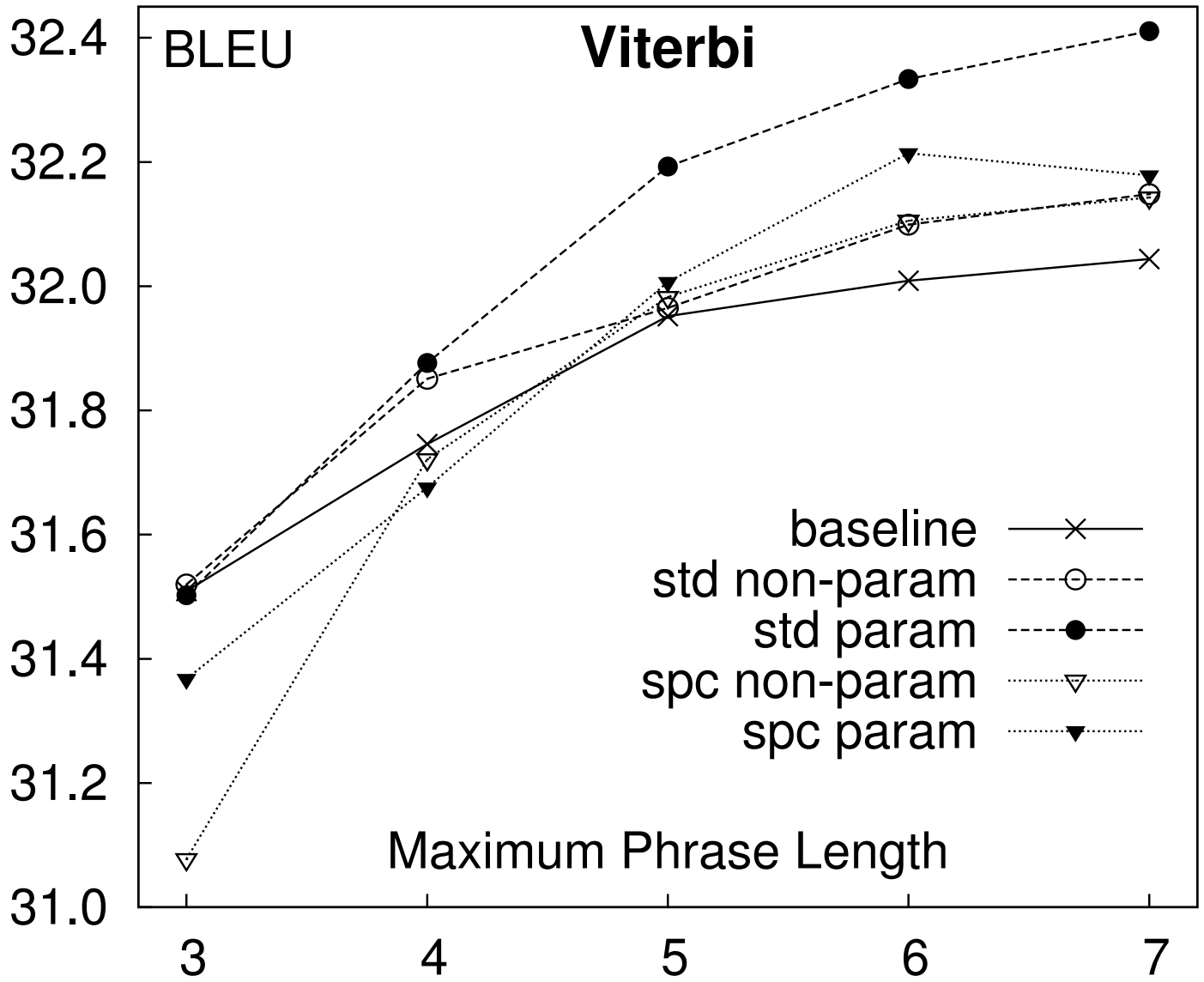
$$\begin{aligned}\hat{y} &= \arg\max_y \; p(y \mid x) \\ &\approx \arg\max_y \; \frac{1}{Z(x)} \exp\left(\sum_i \lambda_i f_i(x,y)\right) \\ &= \arg\max_y \; \sum_i \lambda_i f_i(x,y)\end{aligned}$$

where feature functions $f_i(x,y)$ are logs of:

- Phrase-based translation models: $p(\bar{y} \mid \bar{x})$, $p(\bar{x} \mid \bar{y})$.
- Lexical models: $l(\bar{y} \mid \bar{x})$, $l(\bar{x} \mid \bar{y})$.
- Language model: $p(\bar{y})$.
- Reordering models.
- **Phrase-length models**: std and spc (param/non-param).

Different Contributions to Cost-Effective Transcription and Translation of Video Lectures
J.A. Silvestre Cerdà
Introduction
Explicit Length Modelling for SMT
Introduction
7 Log-linear modelling
Experiments
Conclusions
Efficient Audio Segmentation for Speech Detection
The transLectures-UPV Platform
Recommender Systems for Online Learning Platforms
LM Adaptation Using External Resources for ASR
Demos
Conclusions

Europarl En→Es (train 1M, test 2K)



Different Contributions to Cost-Effective Transcription and Translation of Video Lectures
J.A. Silvestre Cerdà
Introduction
Explicit Length Modelling for SMT
Introduction
Log-linear modelling
8 Experiments
Conclusions
Efficient Audio Segmentation for Speech Detection
The transLectures-UPV Platform
Recommender Systems for Online Learning Platforms
LM Adaptation Using External Resources for ASR
Demos
Conclusions

Conclusions

- Two novel phrase-length models for phrase-based SMT.
- **Statistically significant improvements** on all language pairs.
- Length models behave differently depending on the task.
- Trade-off between model complexity and data sparseness.

Introduction
Explicit Length Modelling for SMT
Introduction
Log-linear modelling
Experiments
9 Conclusions
Efficient Audio Segmentation for Speech Detection
The transLectures-UPV Platform
Recommender Systems for Online Learning Platforms
LM Adaptation Using External Resources for ASR
Demos
Conclusions

Outline

Introduction

Explicit Length Modelling for SMT

Efficient Audio Segmentation for Speech Detection

The transLectures-UPV Platform

Recommender Systems for Online Learning Platforms

LM Adaptation Using External Resources for ASR

Demos

Conclusions

Introduction

- The temporal cost of ASR depends on the input length.
- Only speech segments should be delivered to ASR systems.
- A prior segmentation can provide a better transcription quality.
- A **fast GMM-HMM Audio Segmentation** system is proposed.
- Albayzin Audio Segmentation Evaluation 2012.

Introduction
Explicit Length Modelling for SMT
Efficient Audio Segmentation for Speech Detection
11 Introduction
System Description
Experiments
Conclusions
The transLectures-UPV Platform
Recommender Systems for Online Learning Platforms
LM Adaptation Using External Resources for ASR
Demos
Conclusions

System description

- AS can be seen as a simplified case of ASR.
- Reduced set of acoustic classes $\mathscr{C}$ (i.e. speech, noise, music).
- Search for a sequence of class labels $\hat{c}$ so that

$$\begin{aligned}\hat{c} &= \arg\max_{c \in \mathscr{C}^*} p(c \mid x) \\ &= \arg\max_{c \in \mathscr{C}^*} p(x \mid c)\, p(c)\end{aligned}$$

where:

$p(x \mid c)$ GMM-HMM based acoustic model.
$p(c)$ n-gram language model.

Introduction
Explicit Length Modelling for SMT
Efficient Audio Segmentation for Speech Detection
Introduction
12 System Description
Experiments
Conclusions
The transLectures-UPV Platform
Recommender Systems for Online Learning Platforms
LM Adaptation Using External Resources for ASR
Demos
Conclusions

Albayzin Audio Segmentation Evaluation 2012

- Albayzin corpus: 111h train, 18h blind test.

	SER			
	Speech	Music	Noise	Overall
test	1.9	36.8	46.5	26.5

- Real Time Factor (RTF) values close to zero.
- Final standings for the Albayzin 2012 Competition:

Pos.	System	SER
1	AHOLAB-EHU	26.3
2	**MLLP-UPV**	**26.5**
3	GTM-UVIGO	28.1
4	...	> 33

- MLLP was the fastest among the two best (RTF 0.001 vs 1.6).

Different Contributions to Cost-Effective Transcription and Translation of Video Lectures
J.A. Silvestre Cerdà
Introduction
Explicit Length Modelling for SMT
Efficient Audio Segmentation for Speech Detection
Introduction
System Description
13 Experiments
Conclusions
The transLectures-UPV Platform
Recommender Systems for Online Learning Platforms
LM Adaptation Using External Resources for ASR
Demos
Conclusions

Conclusions

- Simple AS approach based on current ASR technology.
- **Excellent performance** when detecting speech segments.
- Improvable results when dealing with noise and music.
- **Extremely fast** segmentation.
- Among the best two systems in the Albayzin 2012 Evaluations.

Introduction

Explicit Length Modelling for SMT

Efficient Audio Segmentation for Speech Detection

Introduction

System Description

Experiments

14 Conclusions

The transLectures-UPV Platform

Recommender Systems for Online Learning Platforms

LM Adaptation Using External Resources for ASR

Demos

Conclusions

Outline

Introduction

Explicit Length Modelling for SMT

Efficient Audio Segmentation for Speech Detection

The transLectures-UPV Platform

Recommender Systems for Online Learning Platforms

LM Adaptation Using External Resources for ASR

Demos

Conclusions

Introduction

Explicit Length Modelling for SMT

Efficient Audio Segmentation for Speech Detection

15 The transLectures-UPV Platform

Introduction

System Architecture

Conclusions

Recommender Systems for Online Learning Platforms

LM Adaptation Using External Resources for ASR

Demos

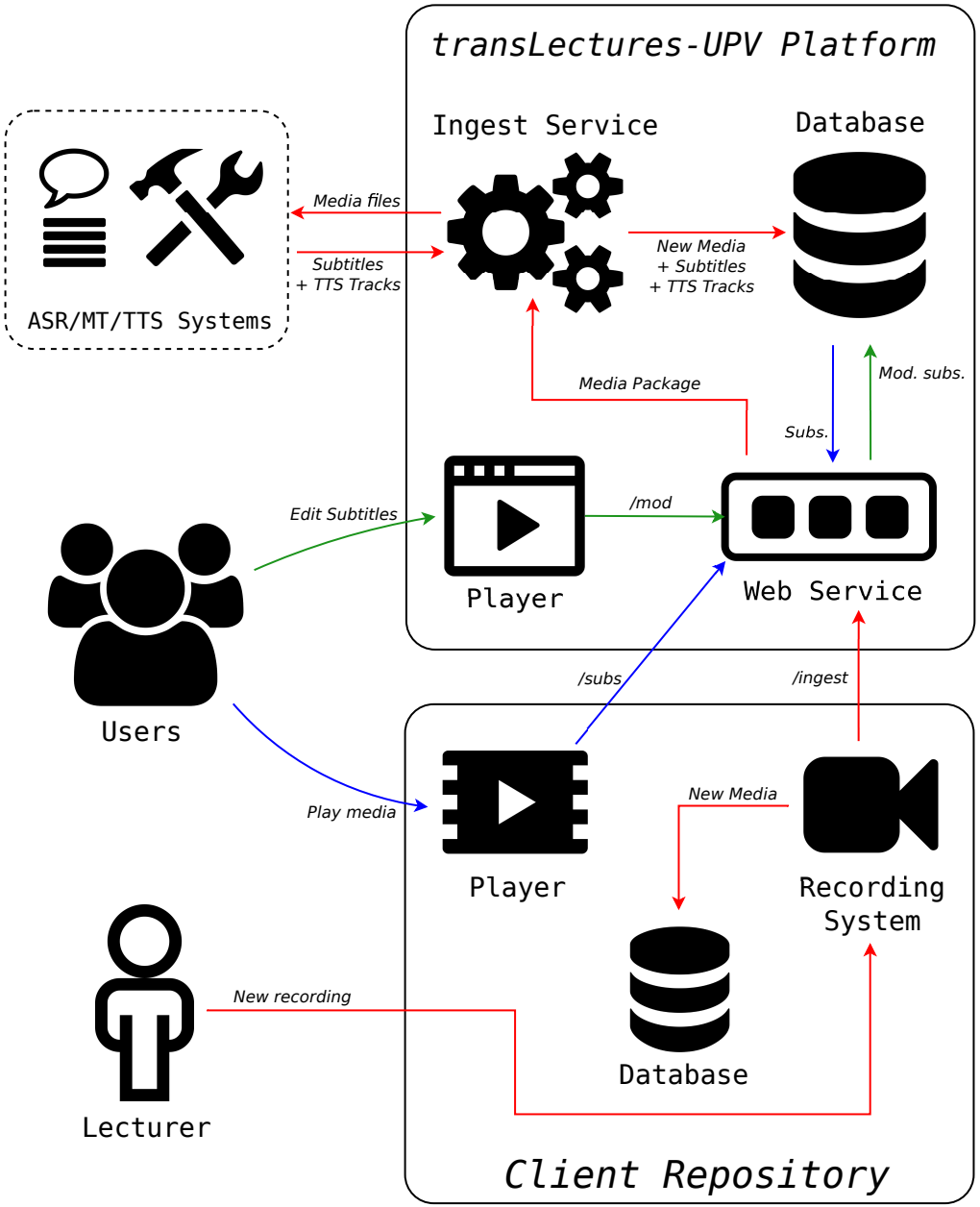
Conclusions

Introduction

- System architecture to **integrate ASR and MT** technologies.
- **Collaborative framework** to review automatic subtitles.
- Adopted in the EU project trans**Lectures**.
  1. Improvement of transcription & translation by massive adaptation.
  2. Improvement of transcription & translation by intelligent interaction.
  3. Integration into Opencast to enable real-life evaluation.
- Focus on the integration with the poliMedia video repository.

Introduction
Explicit Length Modelling for SMT
Efficient Audio Segmentation for Speech Detection
The transLectures-UPV Platform
16 Introduction
System Architecture
Conclusions
Recommender Systems for Online Learning Platforms
LM Adaptation Using External Resources for ASR
Demos
Conclusions

System Architecture and Use Cases



Different Contributions to Cost-Effective Transcription and Translation of Video Lectures

J.A. Silvestre Cerdà

Introduction

Explicit Length Modelling for SMT

Efficient Audio Segmentation for Speech Detection

The transLectures-UPV Platform

Introduction

17 System Architecture

Conclusions

Recommender Systems for Online Learning Platforms

LM Adaptation Using External Resources for ASR

Demos

Conclusions

The transLectures-UPV Platform (TLP) and poliMedia

- TLP is the **open source toolkit** implementing the architecture.
- In production at the UPV's **poliMedia** video lecture repository.
- Service for the distribution of multimedia educational content.
- Courses in videos accompanied by time-aligned slides.
- Statistics of the poliMedia repository (September 2015):

Lectures	15436
Duration (hours)	3079
Avg. Lecture Length (minutes)	12
Speakers	1759
Avg. Lectures per Speaker	8

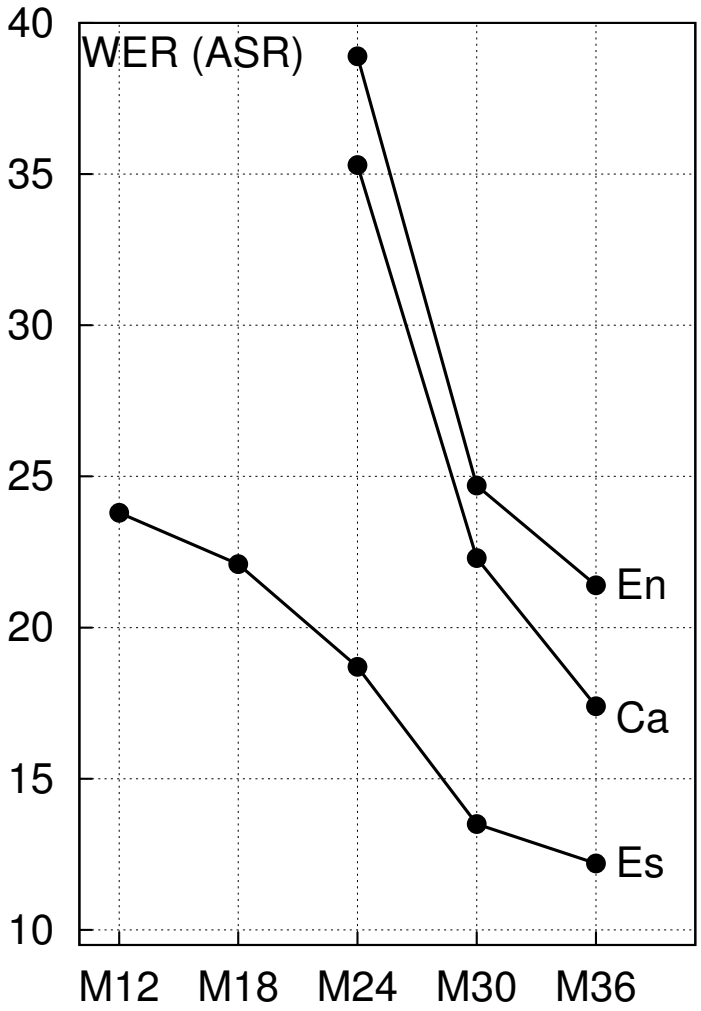
  - Spanish (88%), English (7%), Catalan (3%), others (2%).

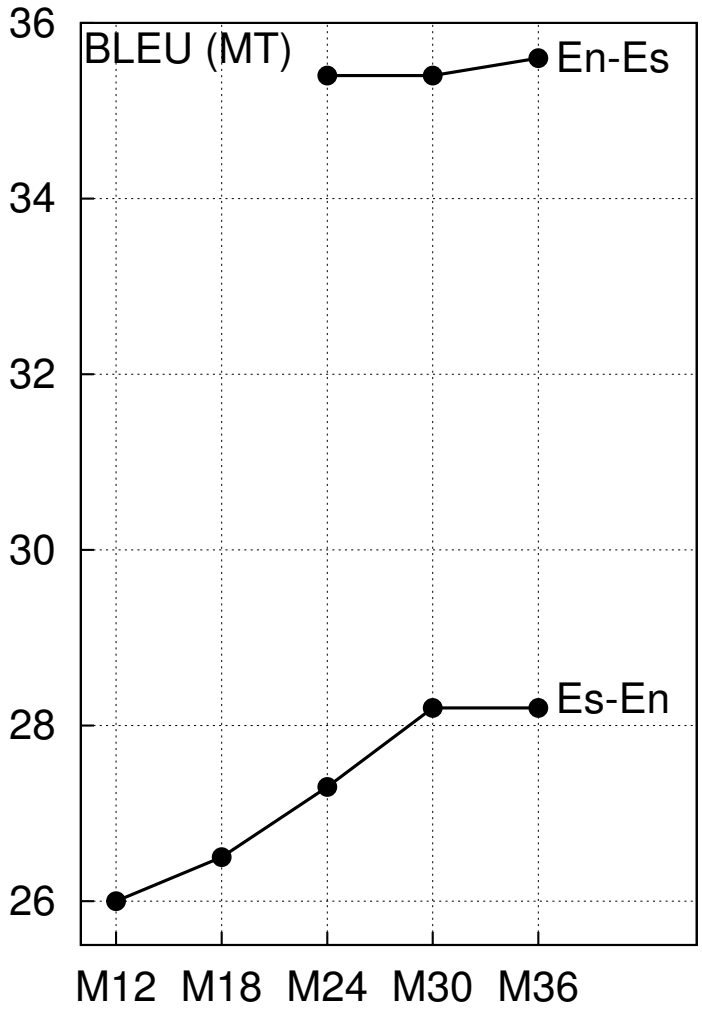
Different Contributions to Cost-Effective Transcription and Translation of Video Lectures
J.A. Silvestre Cerdà
Introduction
Explicit Length Modelling for SMT
Efficient Audio Segmentation for Speech Detection
The transLectures-UPV Platform
Introduction
18 System Architecture
Conclusions
Recommender Systems for Online Learning Platforms
LM Adaptation Using External Resources for ASR
Demos
Conclusions

TLP integration into poliMedia and VideoLectures.NET

- In M12 (October 2012):
  - First Spanish (Es) ASR system.
  - First Es→En MT system.
  - All Spanish lectures were transcribed and translated into English.
- In M24 (October 2013):
  - First Catalan (Ca) and English (En) ASR systems.
  - First Ca↔Es, Ca↔En and En→Es MT systems.
  - All poliMedia lectures subtitled in Spanish, English and Catalan.
  - Any newly recorded lecture was automatically processed in TLP.
- poliMedia was **re-transcribed** and **re-translated** every 6 months.
- Platform assesment: automatic and human evaluations.
- Integration into VideoLectures.NET was similar.

TLP integration: automatic evaluations





TLP integration: user evaluations at the UPV

- **Docència en Xarxa (DeX)** programme of the UPV.
- DeX 2013: Post-editing was the preferred reviewing method.
- DeX 2014:

	Es	Es→En
Lecturers	39	10
Error (%)	12	42
RTF	3	12
RTF from scratch	10	30
Effort red. (%)	**70**	**60**

Conclusions

- System architecture for cost-effective automatic subtitling.
- Implemented as the open source **transLectures-UPV Platform**.
- Integration into pM showed **savings up to 2/3 of user effort**.
- Basis of MLLP's Transcription and Translation Platform (TTP):
  - More than 200 users (orgs) and 1000 videos (250h) uploaded.
  - 10 transcription languages and 14 translation pairs.
  - Support for the EMMA EU project.
  - Support for the Translation Centre for the Bodies of the EU (CdT).
  - Support for the SUBurbia EU project.
  - Under study by many interested organisations.

Introduction
Explicit Length Modelling for SMT
Efficient Audio Segmentation for Speech Detection
The transLectures-UPV Platform
Introduction
System Architecture
22 Conclusions
Recommender Systems for Online Learning Platforms
LM Adaptation Using External Resources for ASR
Demos
Conclusions

Outline

Introduction

Explicit Length Modelling for SMT

Efficient Audio Segmentation for Speech Detection

The transLectures-UPV Platform

Recommender Systems for Online Learning Platforms

LM Adaptation Using External Resources for ASR

Demos

Conclusions

Introduction

Explicit Length Modelling for SMT

Efficient Audio Segmentation for Speech Detection

The transLectures-UPV Platform

23 Recommender Systems for Online Learning Platforms

Introduction

System Design

Evaluation

LM Adaptation Using External Resources for ASR

Demos

Conclusions

Introduction

- Recommender Systems (RS) are often needed by users.
- The use of speech transcripts for recommendation was studied.
- **La Vie PASCAL2 project:**
  - Main goal was to improve recommendations in VideoLectures.NET.
  - Baseline RS based on lecture keywords.
  - New RS based on SVMs and exploiting speech transcripts.

Introduction

Explicit Length Modelling for SMT

Efficient Audio Segmentation for Speech Detection

The transLectures-UPV Platform

Recommender Systems for Online Learning Platforms

24 Introduction

System Design

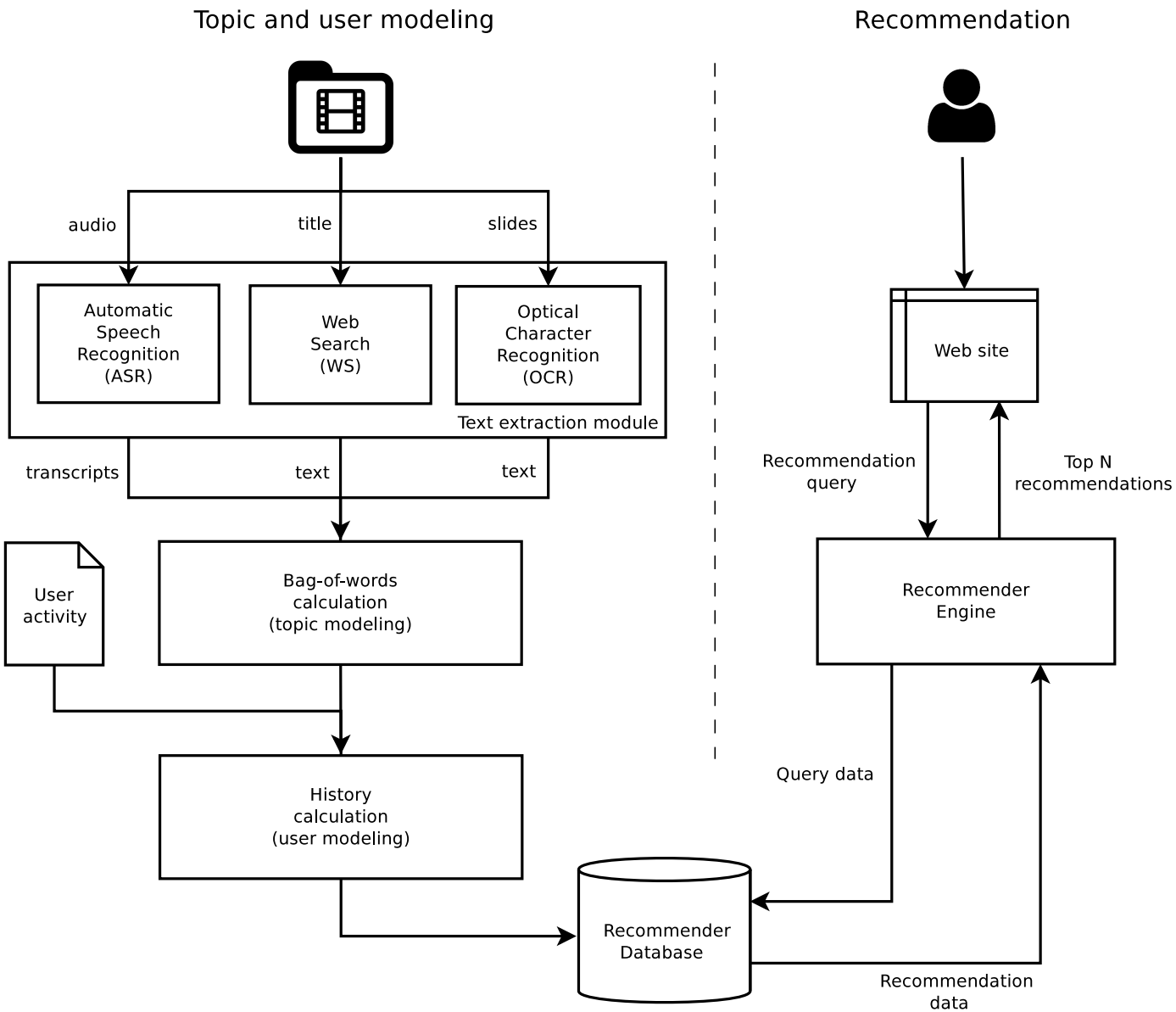
Evaluation

LM Adaptation Using External Resources for ASR

Demos

Conclusions

Recommender System Overview



Evaluation

- Comparison with the baseline RS.
- User clicks on recommended videos were logged.
- Results after a 6-month evaluation were not conclusive.
- An in-depth analysis of the logs is still pending.

Introduction

Explicit Length Modelling for SMT

Efficient Audio Segmentation for Speech Detection

The transLectures-UPV Platform

Recommender Systems for Online Learning Platforms

Introduction

System Design

26 Evaluation

LM Adaptation Using External Resources for ASR

Demos

Conclusions

Outline

Introduction

Explicit Length Modelling for SMT

Efficient Audio Segmentation for Speech Detection

The transLectures-UPV Platform

Recommender Systems for Online Learning Platforms

LM Adaptation Using External Resources for ASR

Demos

Conclusions

Introduction

Explicit Length Modelling for SMT

Efficient Audio Segmentation for Speech Detection

The transLectures-UPV Platform

Recommender Systems for Online Learning Platforms

27 LM Adaptation Using External Resources for ASR

Introduction

Experiments

Demos

Conclusions

Introduction

- ASR performance can be greatly improved using in-domain data:
  - Lecture slides and related documents, if available.
  - Metadata such as title, keywords or abstract.
- LM adaptation by **document retrieval**:
  - PDF files retrieved from search queries based on lecture titles.
  - Per-lecture retrieval methods: exact and extended search.
  - Individual LMs are estimated on each data source separately.

Introduction
Explicit Length Modelling for SMT
Efficient Audio Segmentation for Speech Detection
The transLectures-UPV Platform
Recommender Systems for Online Learning Platforms
LM Adaptation Using External Resources for ASR
28 Introduction
Experiments
Demos
Conclusions

Evaluation and conclusions

- Spanish polimedia corpus: 107h train, 3h test.
- Comparison of search methods against the baseline system:

	WER	Δ%
Baseline (BL)	15.7	-
BL + Exact (5 docs)	14.6	7
BL + Extended (5 docs)	14.4	8
BL + Extended (10 docs)	14.4	8
BL + Extended (20 docs)	14.2	10
BL + OCR Slides	13.8	12
BL + OCR Slides + Extended (20 docs)	13.5	14

- Simple yet effective method to retrieve related documents.

Outline

Introduction

Explicit Length Modelling for SMT

Efficient Audio Segmentation for Speech Detection

The transLectures-UPV Platform

Recommender Systems for Online Learning Platforms

LM Adaptation Using External Resources for ASR

Demos

Conclusions

Demos

- TLP integration with poliMedia:
  - Live: `http://media.upv.es`
- MLLP's Transcription and Translation Platform:
  - Live: `http://ttp.mllp.upv.es`

Outline

Introduction

Explicit Length Modelling for SMT

Efficient Audio Segmentation for Speech Detection

The transLectures-UPV Platform

Recommender Systems for Online Learning Platforms

LM Adaptation Using External Resources for ASR

Demos

Conclusions

Main contributions

- An explicit phrase-length modelling approach for SMT.
- A simple segmentation approach for fast speech detection.
- The transLectures-UPV Platform (TLP) for ASR & MT integration.
- Integration of TLP into poliMedia (UPV), UC3M, etc.
- Support for the MLLP's TTP: EMMA, CdT, SUBurbia, etc.
- A new approach to video lecture recommendation.
- A new document retrieval technique for LM adaptation.

Introduction
Explicit Length Modelling for SMT
Efficient Audio Segmentation for Speech Detection
The transLectures-UPV Platform
Recommender Systems for Online Learning Platforms
LM Adaptation Using External Resources for ASR
Demos
Conclusions
33 Main contributions
Future work
Publications

Future work

- Explicit length modelling for SMT:
  - Perform a full Viterbi-like iterative training method.
  - Smooth Viterbi counts with extract counts.
  - Study alternative weight optimisation methods to MERT.
- Audio segmentation for speech detection:
  - Measure impact on transcription quality in terms of WER.
  - Adopt a hybrid DNN-HMM approach.
- The transLectures-UPV Platform (TLP):
  - To extend TLP to give full support to MOOCs.
  - To explore other applications (i.e. film industry).
- Recommender systems for online learning platforms:
  - Retrain RS using better speech transcriptions.
  - Extend the system to provide cross-lingual recommendations.
- LM adaptation using external resources:
  - Consider also retrieving web pages (HTML).
  - Adaptation speaker's vocabulary.

Publications

This work has derived **9 scientific publications**:

- Explicit Length Modelling for SMT (2):
  - 1 International Journal (Pattern Recognition)
  - 1 International Conference (IbPRIA)
- Efficient Audio Segmentation for Speech Detection (1):
  - 1 Competition (Albayzin evaluations)
- Transcription and Translation Platform (4):
  - 1 International Journal (Speech Communication)
  - 2 International Conferences (IEEESMC, EC-TEL)
  - 1 National Conference (IberSpeech)
- Recommender Systems for Online Learning Platforms (1):
  - 1 National Conference (IberSpeech)
- LM Adaptation Using External Resources for ASR (1):
  - 1 National Conference (IberSpeech)

Thank you for your attention!

Gràcies per la vostra atenció!

Gracias por vuestra atención!

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

MLLP | Machine Learning and Language Processing