Spoken Language Processing Research Activities at NLPR

Prof. Bo Xu

Email:xubo@nlpr.ia.ac.cn

National Lab of Pattern Recognition(NLPR)
Institute of Automation
Chinese Academy of Sciences

Oct,2001

中国科学院自动化研究所
模式识别国家重点实验室

National Lab of Pattern Recognition
Institute of Automation,CAS

NLPR

Outline

Corner of SLP Room:
4 Research Staff
12 Ph.D candidates
7 M.S. Candidates

Corner of NLP Room:
3 Research staff
6 Ph.D Candidates
3 M.S.Candidates

NLPR

Long Histroy in Speech Recognition

- Mandarin LVCSR
  - Unified Triphone and Tone modeling
  - Onepass search
  - Very large text and speech corpus based training
  - Comparable accuracy to state-of-art system
- Spoken Dialogue System
  - ASR, Spoken Understanding, Dialogue Management, Language Generation, Speech Synthesis
  - RoadStar; Providing information about 400 interesting sites all over the China

NLPR

Natural Language Processing

- Spin off from Speech Processing
  - From speech recognition to speech understanding
  - Speech translation
  - Internet information retrieval
- Mainly focus spoken language processing
  - Rather than traditional NLP
- Get funding from various sources
  - 973, 863, NSF, Industry …

NLPR

Agenda:

- Spoken text Corpus
- Statistical Tagginig and Parsing
- Audio corpus retrieval
- **Speech Translation**

1. Spoken Text Corpus

- Written Text Corpus
  - Plain text corpus for N-gram in ASR(>5G)
  - Word tagged corpus
    - 13M with word segmentation and tagging
  - Special purpose corpus
    - Name corpus, affiliate name corpus, address corpus
- Spoken Text Corpus
  - Monolingual spoken text corpus
    - Roadstar and hotel reservation domain
  - Bilingual spoken text corpus
    - for statistical translation(Hotel reservation and Travel domain)

Learning from Corpus

- From the view of the meaning:
  - Meaning is simple and relates just one topic
  - can be represented in a simple form
- From the view of the structure
  - ill-formed
  - phrase have to conform to some strict linguistic rules
- Parsing ?
  - It's difficult to use syntactical-driven parsing!
- **Human-human dialogue vs. Human-Machine dialogue ?**

2. Statistical Tagging and Parsing

- Unified Word Trigram and POS Trigram modeling for Tagging

$$p(W,T) = P(W/T) \approx \prod_{i=1}^{n} p(w_i / t_i)\, p(t_i / t_{i-1} t_{i-2})$$

$$P(w,T) = P(T/W)P(W) \approx \prod_{i=1}^{n} p(t_i / w_i)\, p(w_i / w_{i-1} w_{i-2})$$

$$P^*(W,T) = \alpha \prod_{i=1}^{n} p(w_i/t_i)\, p(t_i/t_{i-1},t_{i-2}) + \beta \prod_{i=1}^{n} p(t_i/w_i)\, p(w_i/w_{i-1}w_{i-2})$$

Experiment condition

- No OOV (lexicon is collected from corpus)
- Vocabulary : 50000
- Tagging:
  - The first directory 19
  - the second directory 78
- Testing corpus is extracted from training corpus and are excluded in training
- Training corpus: 13M
- Test corpus: 40K words

NLPR

Segmentation and POS tagging

Types of Test Results	Segmentation Precision(%)	First Level Tagging Precision(%)	Second Level Tagging Precision(%)
Close Test without Language Model	97.78	96.33	93.24
Open Test without Language Model	96.79	96.32	93.10
Close Test with Language Model	99.48	96.28	93.21
Open Test with Language Model	98.06	96.32	93.07

3. Audio Corpus Retrieval

- Speech Classifying and Recognition
  - Speech/Nonspeech(Music, noise, …) classification
  - Speech Endpoint Detection
  - Background classification
  - Speaker Tracking
  - Speech Recognition
- Text Information Retrieval
  - Natural Language Query(and ,or and not operation)
  - Fuzzy based retrieval
  - Considering concept relation between the words

4. Multi-engine Speech Translation

- IF interfaced translation
  - Text-to-IF Understanding
  - IF-to-Text Generation
- Text Interfaced translation
  - Chinese-English Statistical Translation

IF-interfaced framework

C-STAR defined IF (Interchange Format) is a kind of
Underspecified Semantic Representation
that can be formalized as

IF ::= Speaker:[verify/request-verification/negate-]
Speech Act[+Concept*][Argument*]

Text->IF Understanding

- Spoken Language
  - Ungrammatical
  - Very simplified expression
- Basic idea
  - Taking word sequences as HMM input
  - HMM states represents the semantics
  - HMM state connection represents the grammar
  - By learning the HMM parameters, we can recover the semantic sequences from word sequences
  - Finally Mapping to semantic sequences

Semantic marking

- HEAD: information about sentence type

  TOPIC: main topic of a sentence

  REFERENCE WORD: topic identifier

  CASE: sub-topic of a sentence

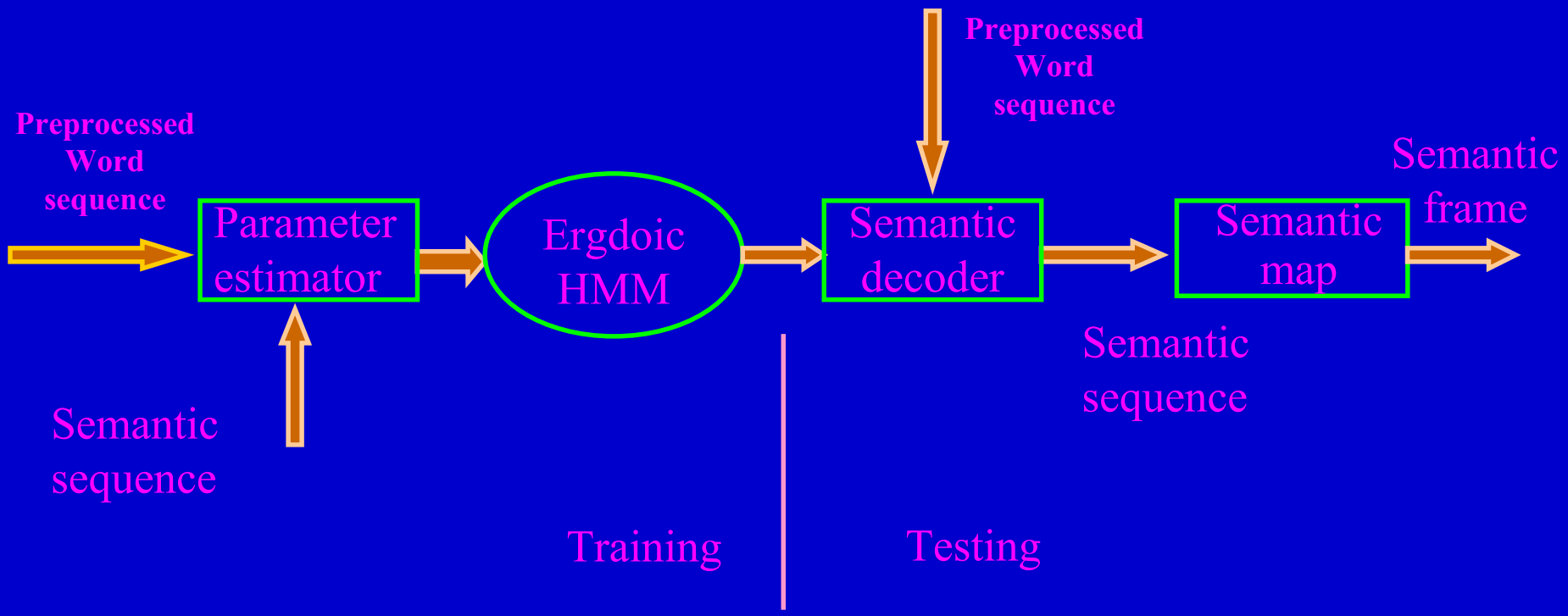
  CASE MARKER: case identifier

- 您好　我　要　订　一　个　单间

{h:greet} {null}{null}{t:reserve}{c:num}{m:num}{c:roomlevel}

NLPR

Diagram of HMM Understanding

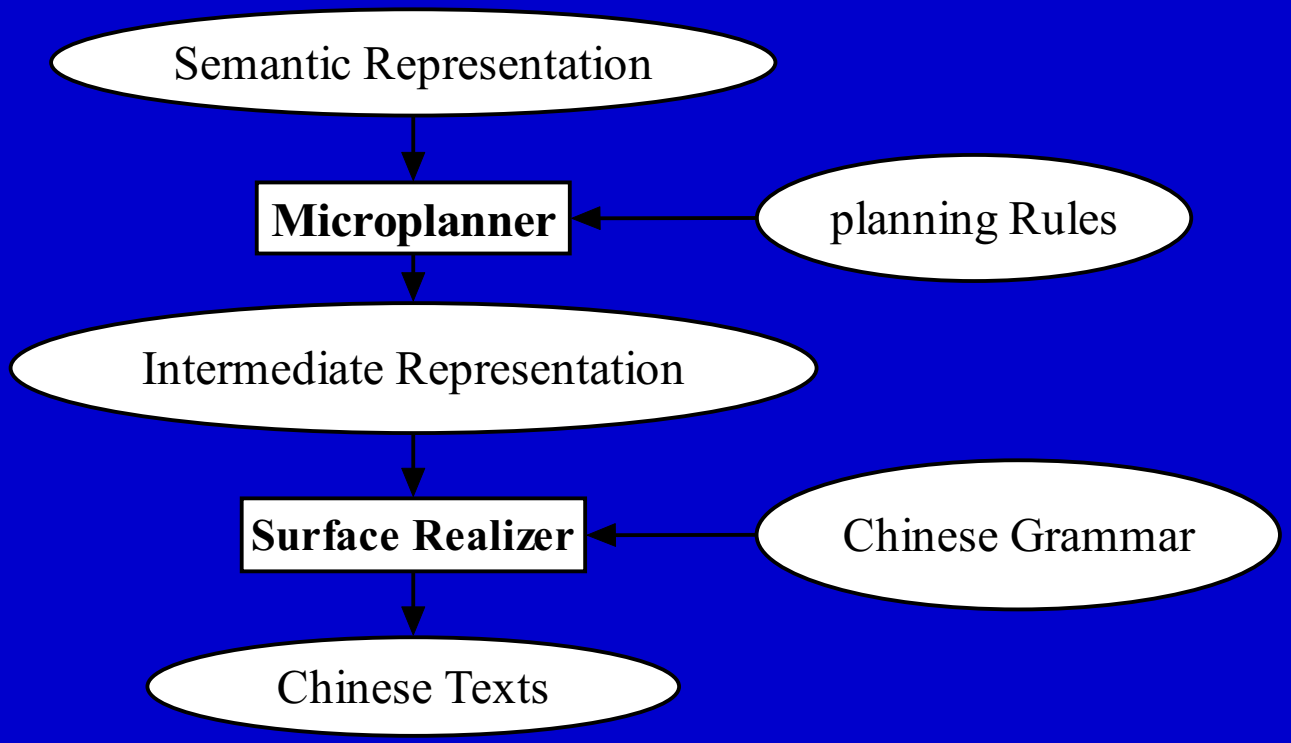


Experiment Result:

- Training corpus (1037 sentences)
  - Error rate 13%
- Test corpus (230 sentence)
  - Error rate 28%
- Main Problem
  - Data Sparse

NLPR

The Chinese Generator



Text-interfaced Framework

- Why statistical Translation
  - Example-based,template-based vs. Model-based
  - Feasibility to integrate the advantage of example-based or some rule-based ideas
  - Robustness to recognition error
  - If we can have a rough automatic evaluation method, we can adopt the methodology of ASR that achieve great successful in past ten years.

NLPR

EXPERIMENTS OF SPOKEN-LANGUAGE

- Bilingual Corpus
  - Training set
  - Test set

- Preprocessing
  - Sentence segment
  - Word segment
  - Categorization

TABLEI Training Set

		Chinese	English
Training	Sentences	3009	
	Words	15547	16935
	Vocabulary Size	804	726

TABLEII Test Set

		Chinese	English
Text	Sentences	100	
	Words	742	812
Speech	Item	51	
	Words	321	--

NLPR

EXPERIMENTS OF SPOKEN-LANGUAGE

- Performance measures
  - rank-A: Fair

    我还不熟悉你们宾馆在什么地方。

    I do not know where the hotel.

    I do not know where your hotel is.
  - rank-B: Acceptable

    我想问一下，就是说，我想订四间。

    I want to inquire, I mean, do I need to reserve four rooms.

    I want to inquire, I mean, I want to reserve four rooms.
  - rank-C: Nonsense

    您订哪天的房间?

    Which room are you sure that I reserved for tomorrow?

    When do you need it?

NLPR

EXPERIMENTS OF SPOKEN-LANGUAGE

TABLEIII RESULTS OF EXPERIMENT

	rank-A (%)	rank-B (%)	rank-C (%)
Text input	67	28	5
Speech input	21.6	43.1	35.3

- **Error analyses**
  - **Subordinate clause**
  - **Phrases and idioms**
  - **Data sparse …**

NLPR

TABLEIV DETAIL ANALYSES

		Spoken dialogues Translation		
		rank-A (%)	rank-B (%)	rank-C (%)
Speech Recognition	rank-A	68.7	28.1	3.2
	rank-B	4.3	43.4	52.3
	rank-C	0	16.7	83.3

Robustness to speech recognition error

NLPR

Translation result for written text

(1) <s> 他们 已经 放弃 了 一切 希望 。

<s> they <had done> abandon all hope .

<s> they <did> accepte our term <n-s> .

<s> they <did> (account for) five enemy plane <n-s> .

(2) <s> 她 被迫 放弃 了 那个 想法 。

<s> she <was> <be done> (be obliged to) abandon that idea .

<s> her brother <was> <be done> (be obliged to) abandon that idea .

(3) <s> 由于 缺乏 资金 ， 这位 科学家 放弃 了 他的 研究工作 。

<s> the scientist <did> abandon his research (for lack of) fund .

<s> the scientist <did> abandon his wife and his child .

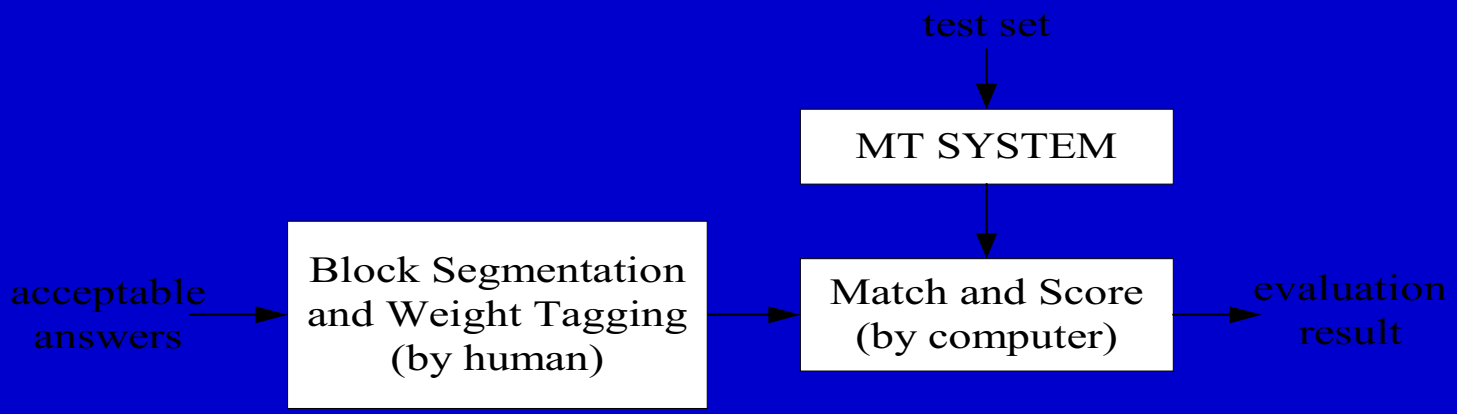
<s> the scientist <did> abandon his wife and his research (for lack of) fund .

(4) <s> 他 陷于 绝望 。

<s> he <did> (abandon oneself to) despair .

<s> he <did> (abandon oneself to) <doing> drink .

Automatic Evaluation of Output Quality



- **Preprocess of the acceptable answers**
  - **Information block**
    - It corresponds in some way to prosodic patterns and chunks.
    - The word order within a block is almost fixation; while the order in which a block occurs is much more flexible.
  - **Weight**
    - The weight of the block: main information; assistant information; complemental information; punctuation.
    - The weight of the word: center word; assistant word; structure word.

Automatic Evaluation of Output Quality

- **Automatic evaluation**
  - **Word match**
    - **Complete matching: the output word is as same as the one in the answer set .**
    - **Proton matching: the output word and the answer have the same etyma.**
    - **Meaning matching: the output word and the answer have the same meaning.**
  - **Evaluation score**

$$recall = \frac{\sum_{output}\left[weight_{block} \times \sum_{output}\left(maching \times weight_{word}\right)\right]}{\sum_{answer}\left(weight_{block} \times \sum_{answer} weight_{word}\right)}$$

$$precision = \frac{\sum_{output} maching}{SentenceLength_{output}}$$

$$score - F = \max_{i} \frac{(\beta^2 + 1) \times precision_i \times recall_i}{\beta^2 \times precision_i + recall_i} \qquad (in\ this\ exa\min ation\ \beta = 1)$$

Automatic Evaluation of Output Quality

- **Evaluation by human:**
  - **Comprehensibility**
  - **Fidelity**
  - **General evaluation**

$$Score_{human} = 0.33 \times comprehensibility + 0.67 \times fidelity$$

- **Examination:**
  - **The ability to distinguish the output quality.(see the figures in next page)**
  - **The ability to show the quality of MS systems.**

The number of training set	recall	precision	Score-F	PER
500	0.230182	0.255475	0.233182	0.193147
1000	0.398415	0.43529	0.404053	0.332356
2000	0.665772	0.721828	0.676731	0.565846
3000	0.753563	0.785515	0.757082	0.644348

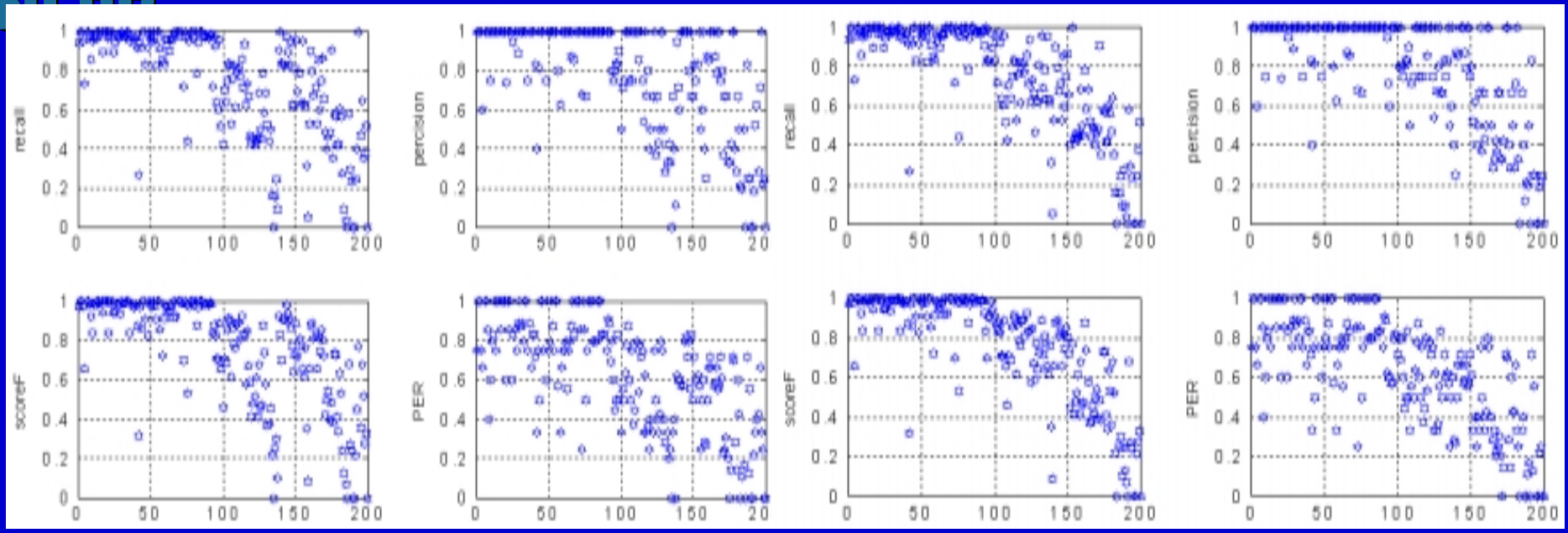


Fig. 1 for the comprehensibility

Fig. 2 for the fidelity

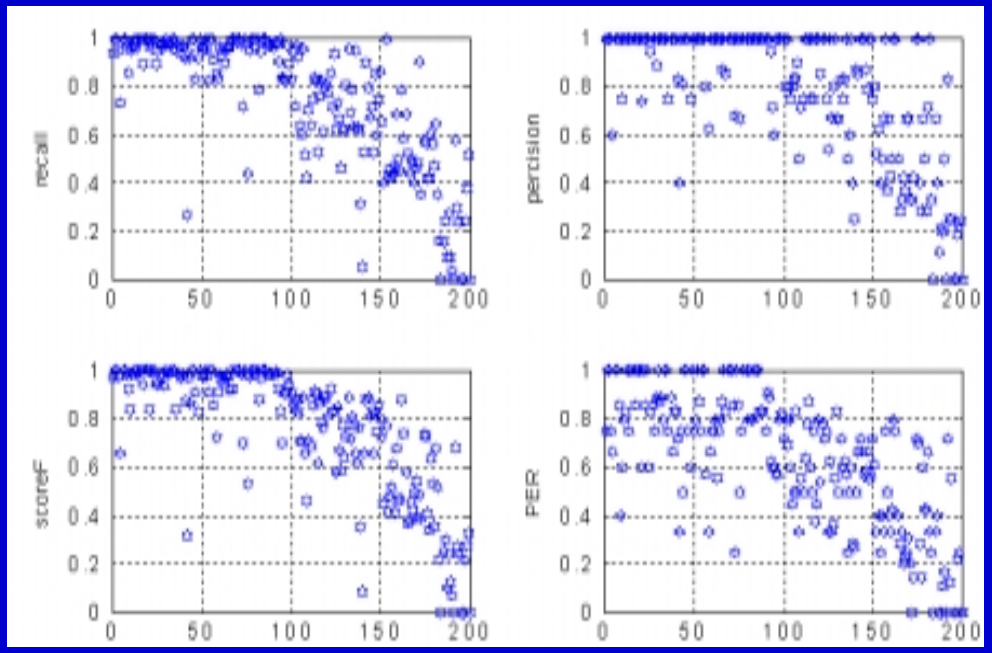


Fig. 3 for the general evaluation by human

NLPR

Automatic Evaluation of Output Quality

– The comparison of the four algorithm.

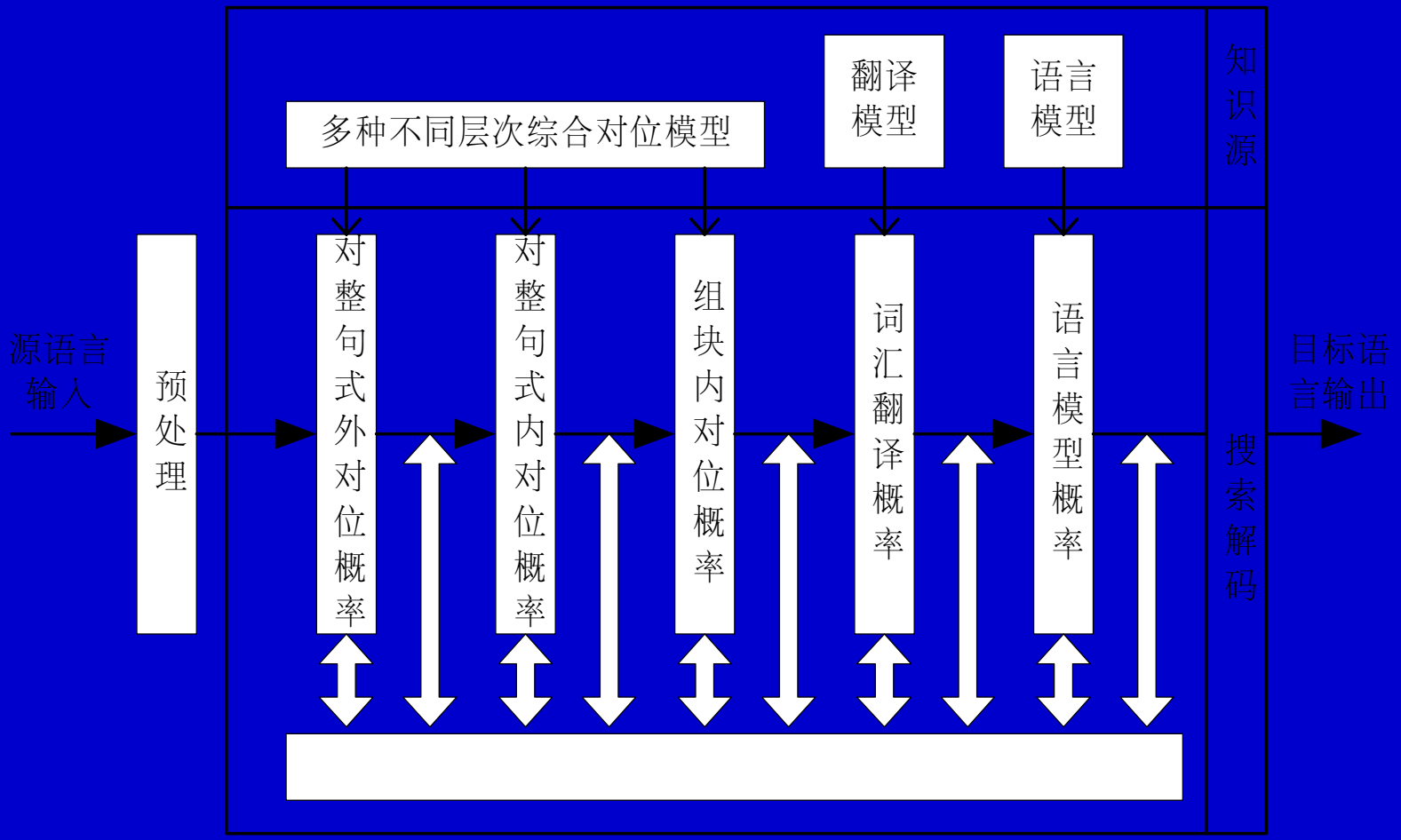
		average				DX			
		recall	precision	Score-F	PER	recall	precision	Score-F	PER
fidelity	A	0.9414	0.9509	**0.9435**	0.816	0.01071	0.01307	**0.01054**	0.02980
	B	0.7037	0.8192	**0.7433**	0.5734	0.03236	0.03282	**0.02384**	0.02506
	C	0.5209	0.5769	**0.5215**	0.3998	0.01562	0.06216	**0.01839**	0.04191
	D	0.1839	0.2453	**0.1852**	0.1737	0.03034	0.05245	**0.02828**	0.04313
Comprehensibility	A	0.8311	0.873	**0.844**	0.7207	0.05107	0.04550	**0.04671**	0.05894
	B	0.5858	0.6592	**0.6004**	0.4654	0.07901	0.09761	**0.07830**	0.06150
	C	0.5375	0.5160	**0.4838**	0.3445	0.04427	0.07134	**0.02904**	0.03454
	D	0	0	**0**	0	0	0	**0**	0

- **The comparison of our algorithm and the PER**

comparison	Score-F and the human evaluation	PER and the humanevaluation
The ratio of the difference (%)	17.5	32

NLPR

Alignment unit : Block and Sentence Pattern based Statistical Translation



NLPR

Translation System and Activities

Thanks !