Goal-Driven Multi-Turn Dialog Processing: From Call Routing Search to Entropy Minimization Dialogue Management

Chin-Hui Lee
School of ECE, Georgia Tech
Atlanta, GA 30332-0250, USA
chl@ece.gatech.edu

Collaboration with colleagues at BL, NUS, I2R & Tsinghua

Talk Outline

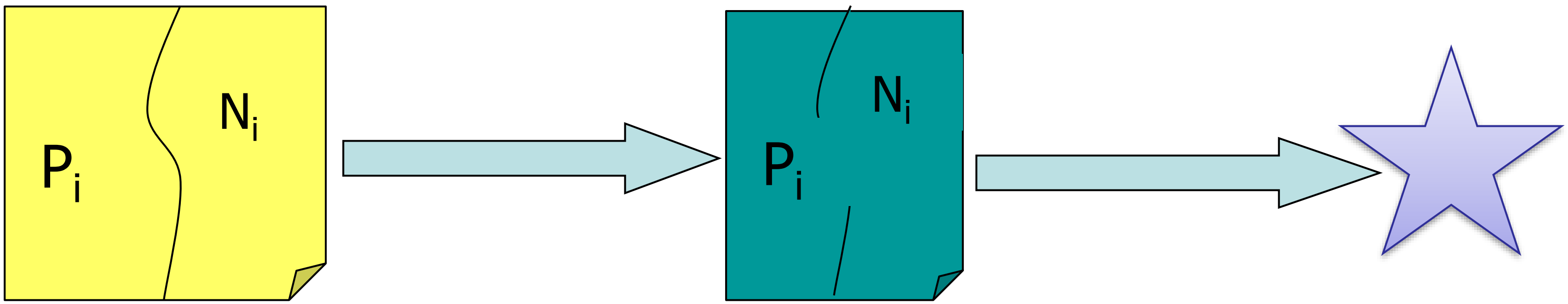
- Discriminative text categorization: unification
  - For speech, music, image & video via tokenization
- Call routing (CR) based on text categorization (TC)
  - Search with collaborative dialogues: USAA banking
  - Human-like machines outperform human agents
- A probabilistic representation of multi-turn dialogue
  - Dynamic stochastic dialogue state modeling, no training
- From call routing to multi-turn, goal-driven dialogue
  - Entropy minimization dialogue management (EMDM)
  - Experimental illustration and result analysis
- Summary

Text Categorization (TC) Unification

1. Training

Feature Extraction (LSA) & Reduction (SVD)

Discriminative Classifier Learning (MFoM)



Training set for each category C_i , i = 1,…,M

Doc. in new feature space

Classifier T_i for category C_i

2. Testing

Unknown Document → LSA & SVD → **Feature** → Text Classification → **TC Results**

Adopt information retrieval (IR), Tokenization: media to text documents

A Binary Classification TC Illustration

- *ModApte* split version of *Reuters-21578* task
  - Lexicon: 10118 words, remove 319 stop-words and words occurred less than 4 times
  - Experiments setup: 7,770/3,019 training/test documents, 90 topics, some with only few positive training instances
  - Gao, *et al*, SIGIR2003, my first paper from NUS, maximal figure of merit (MFoM) discriminative training (DT) is key
  - Using simple LDF as classifiers, DT on weight vectors

	***k*-NN**	**SVM**	**Binary F_1-MFoM**
micR	**0.834**	**0.812**	**0.857**
micP	**0.881**	**0.914**	**0.914**
micF$_1$	**0.857**	**0.860**	**0.884**
macF$_1$	**0.524**	**0.525**	**0.556**

Binary vs. Multi-Category MFoM DT (Gao, *et al*, ICML 2004, ACM T-IS 2006, Binary MFoM better than SVM, SIGIR2003)

Category	# of Training instances	Binary MFoM	MC MFoM
Income	9	0.429	0.600
Oat	8	0.167	0.500
Platinum	5	0.286	0.833
Potato	3	0.333	0.750
Sun-meal	**1**	**0.000**	**0.667**

- **F_1** -based comparison (Gao, *et al*, ICML2004): Multi-Class MFoM works better for training with little positive samples

Maximal Figure-of-Merit (MFoM) Learning

- Distance based loss: $l_k(X_i, \Lambda) = l(d_k) = 1/\{1 + \exp[-a(d_k + b)]\}$
- Overall empirical type I error

maximal separation

$$L_{k1}(\Lambda) = 1/V_{k1} \sum_{i=1}^{V} l_k(X_i, \Lambda) 1(X_i \in C_k)$$

- Overall empirical type II error

(Gao & Lee, SIGIR2003)

$$L_{k2}(\Lambda) = 1/V_{k2} \sum_{i=1}^{V} [1 - l_k(X_i, \Lambda)] 1(X_i \notin C_k)$$

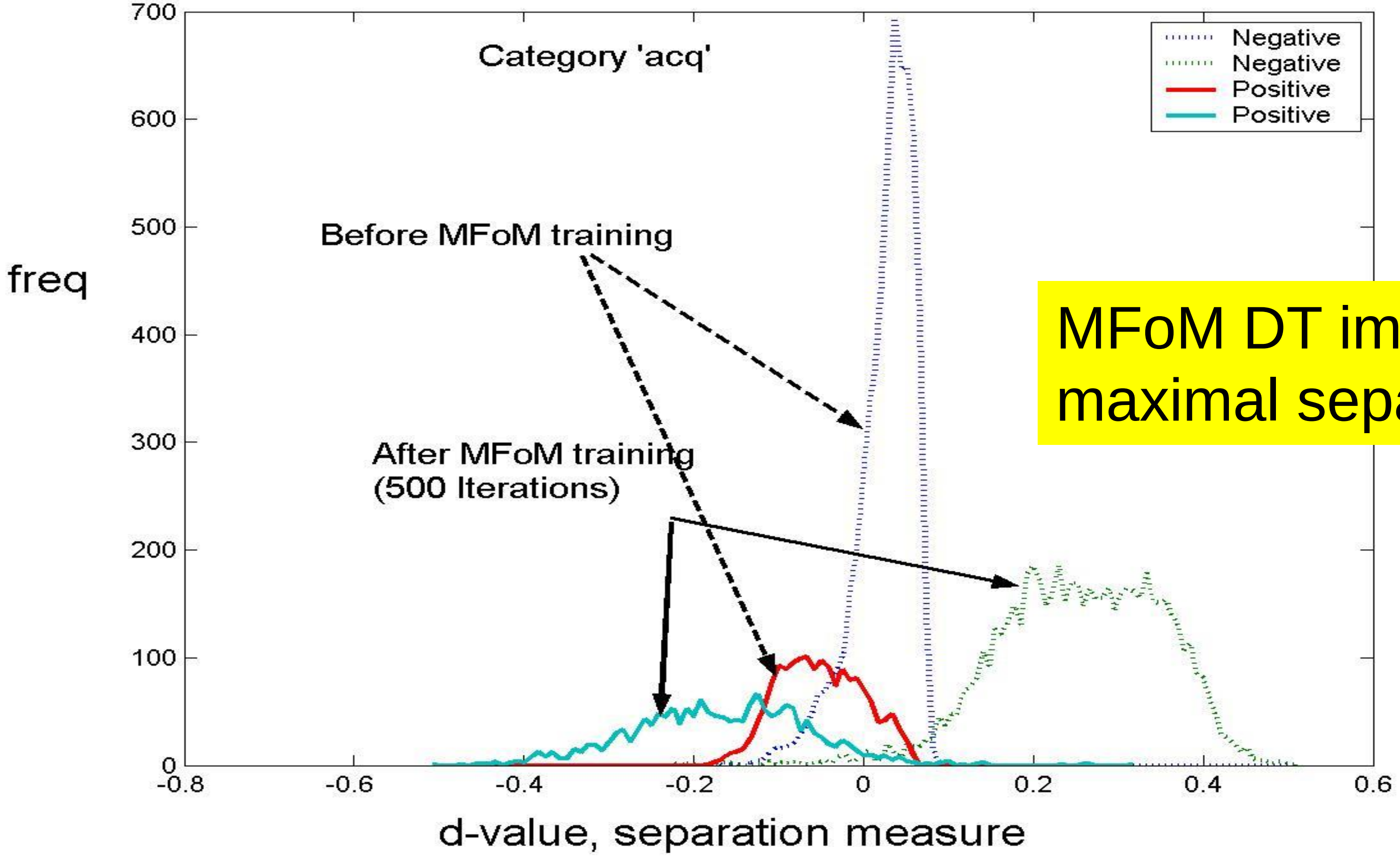
- Overall empirical loss to be minimized (any figure of merit or FoM: precision, recall, F_1 etc.): e.g., AUC

$$U = \sum_{i=}^{M} \sum_{j=1}^{N} I\left(x_i, y_j\right) \Big/ MN$$

(Gao & Lee, ICPR2006)

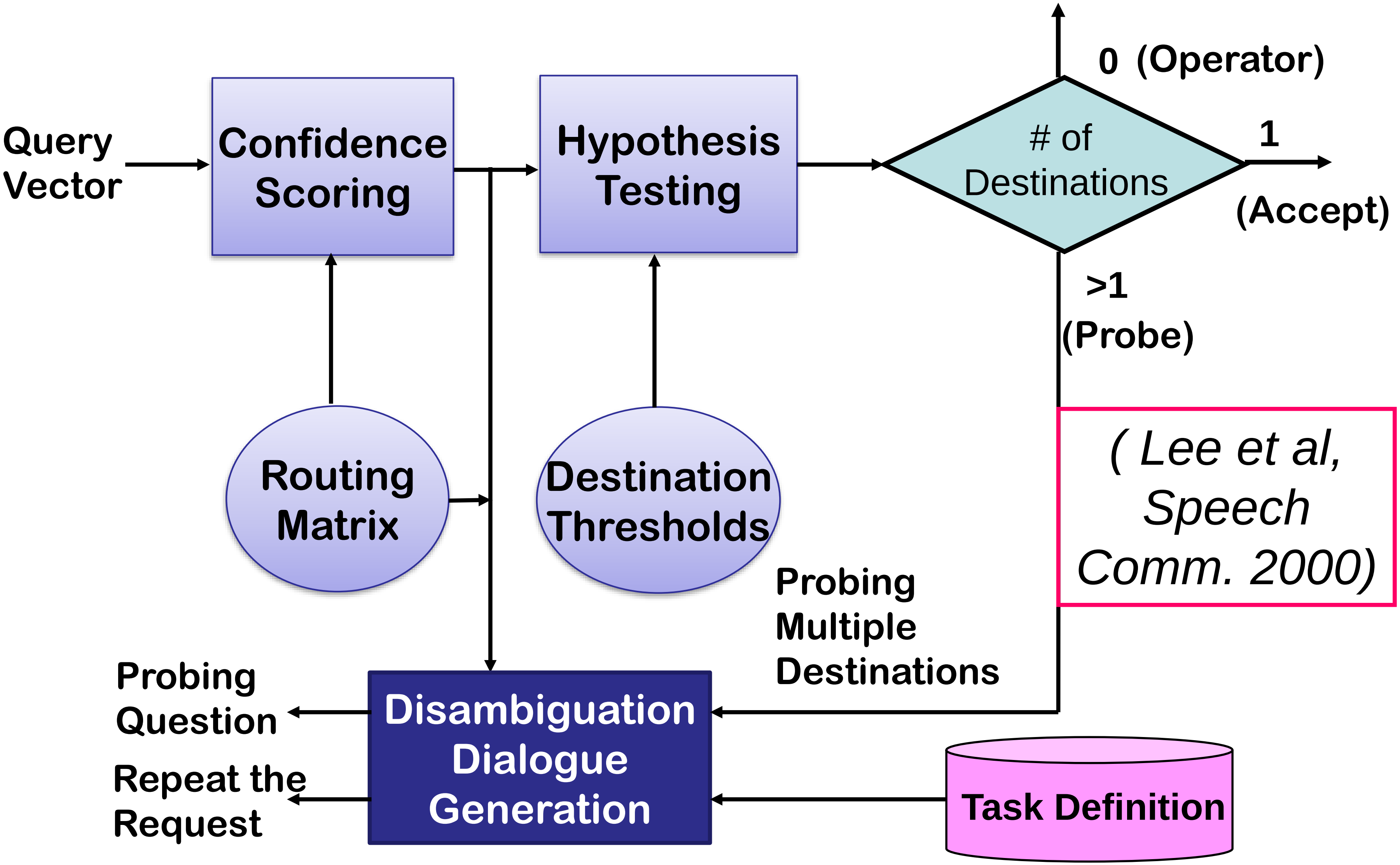
- Epoch-based generalized probabilistic descent (GPD)
  - 5000 iterations

Class Separation before & after MFoM (Gao, *et al*, SIGIR 2003, ICML 2004, ACM T-IS 2006)



MFoM DT implies maximal separation

A Dialogue-Based Call Router



Task Analysis – USAA Banking
(Last Major Project at BL -- ASR , NLP & BU)

- USAA Banking task: utilizing text categorization
  - Mostly veterans and their families (lots of naïve users)
  - Call directors handles over 1000 lines, need to double the agents and the space for the equipment
  - Call directors cost about 80% in a call center, cutting down connection time means big savings
  - 23-40 destinations for automation (cover +99% traffic)
- Catch-all number (Natural Language Call Routing)
  - People call for many purposes (ambiguous request)
  - Call directors are not well-trained (high turnover rate)
- Task could be very challenging: high ASR errors

Vector-Based Routing Matrix (from IR)

Adopt information retrieval (IR), Tokenization: media to text documents

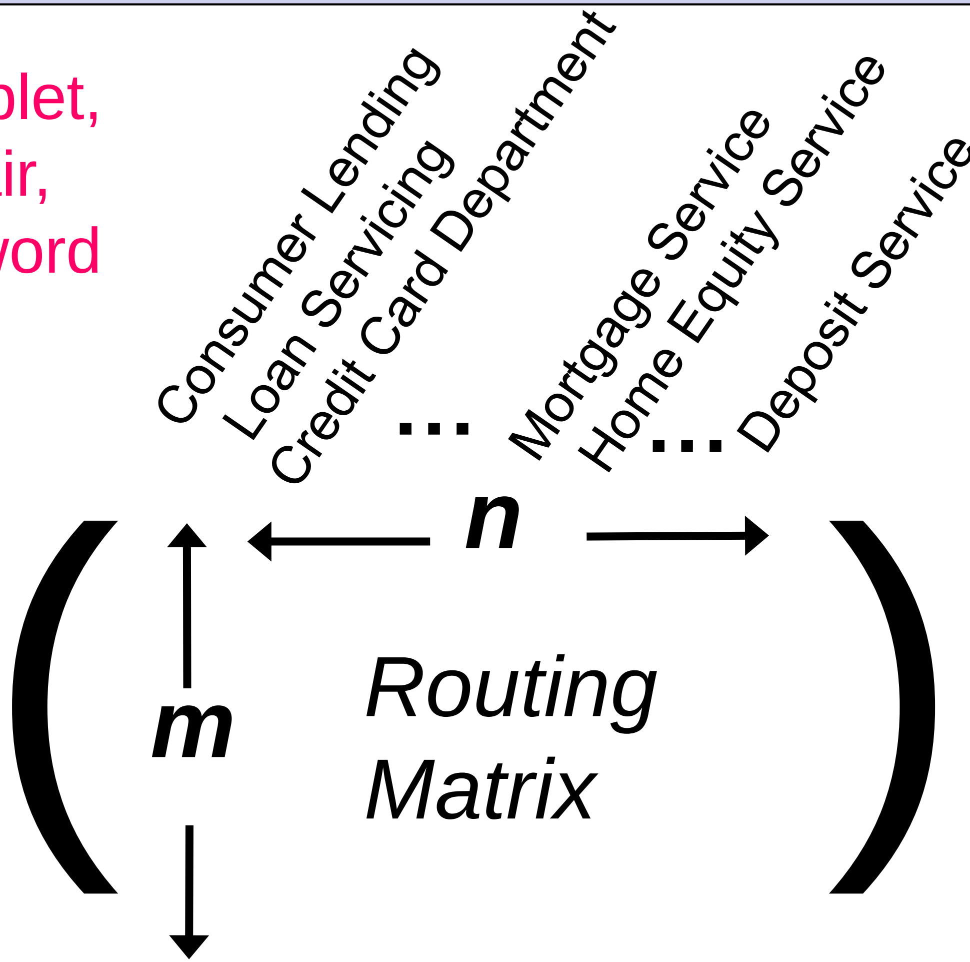
Features: **trigram** = word triplet,
bigram = word pair,
unigram = single word

Forming Query Vectors

trigrams > 3 times: home,equity,loan; new,auto,loan; ⋮

bigrams > 3 times: bank,card; current,rate; ⋮

unigrams > 2 times: annuity; ⋮



In call routing, multiple word co-occurrence increases indexing power

Examples of User Requests

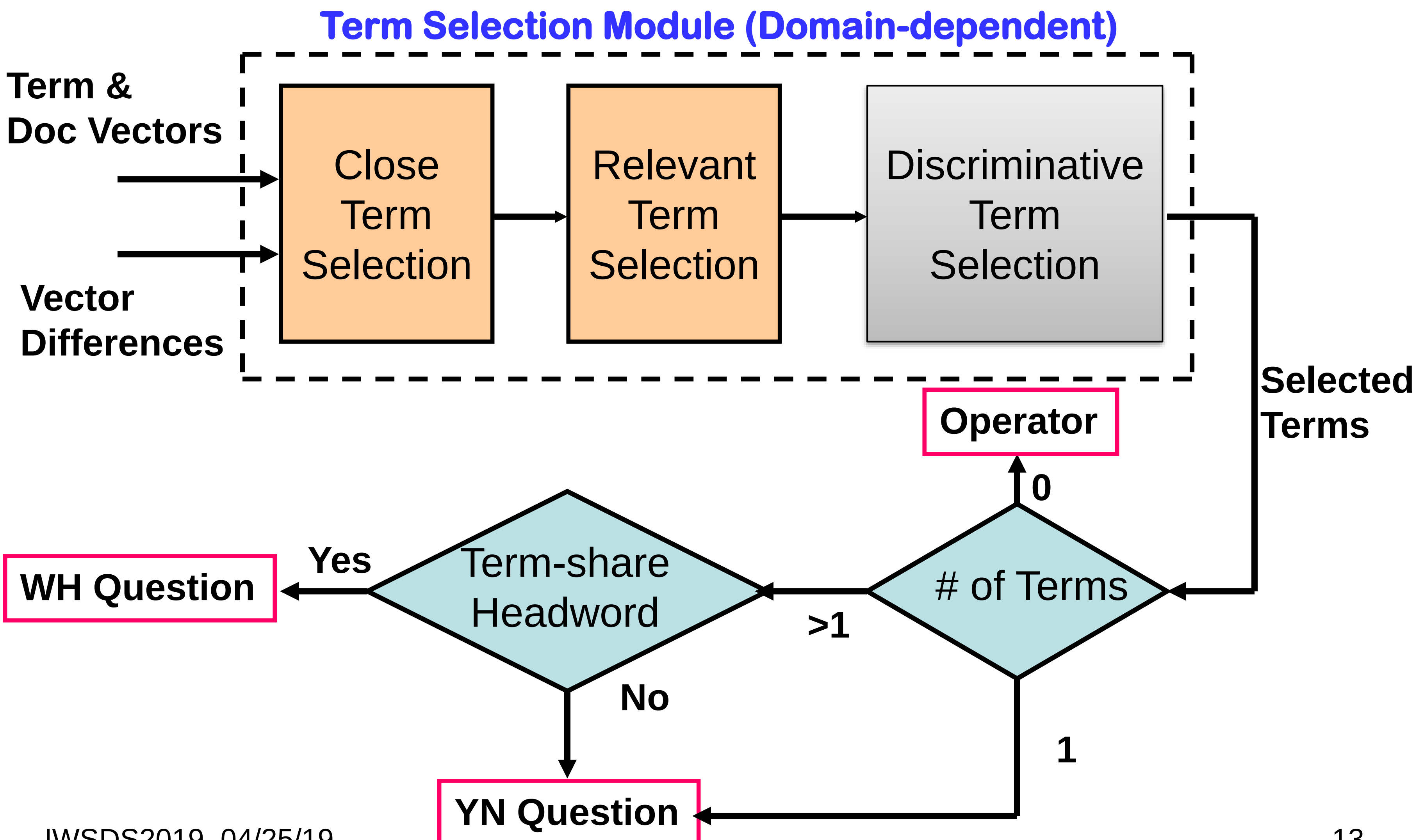
Category	Query Examples
1. Direct Request	"Yes ma'am. I'm trying to find someone in *deposit services."*
-	"Uh, please connect me to *credit card services."*
2. Activity Request	"Yes I need to speak to someone about *wiring money to my checking account*."
-	"Um I need the *blue book value* of a vehicle I am thinking about buying."
3. Ambiguous Request	"I need some information on *auto loans."* *or* "I want to *transfer some money*."

Example of Disambiguation Probes

Ambiguity	Triggers	Department
Balance	CD, checking, savings, IRA	Deposit Services
-	Visa, Mastercard, credit card	Credit Card Services
-	Loan	Loan Servicing

Disambiguation queries are needed to resolve the request

Disambiguation Dialogue Generation (Automatic Search Refinement)



A Disambiguation Dialogue Example

- User Request: "*loan information, please*."
  - Two candidates - *Consumer Lending* or *Loan Servicing*
- Closeness Selection: gives 60 terms
  - For each candidate destination, compare term vectors with difference vectors and select 30 "close" terms
- Relevance Selection: reduces to 27 terms
  - Select "relevant" terms that form a valid n-gram when combining with terms in the original request (e.g. if "*car,loan"* is in the original query vector, then "*new"* is a *relevant* term to form the valid term "*new,car,loan*"

Disambiguation Dialogue (Cont.)

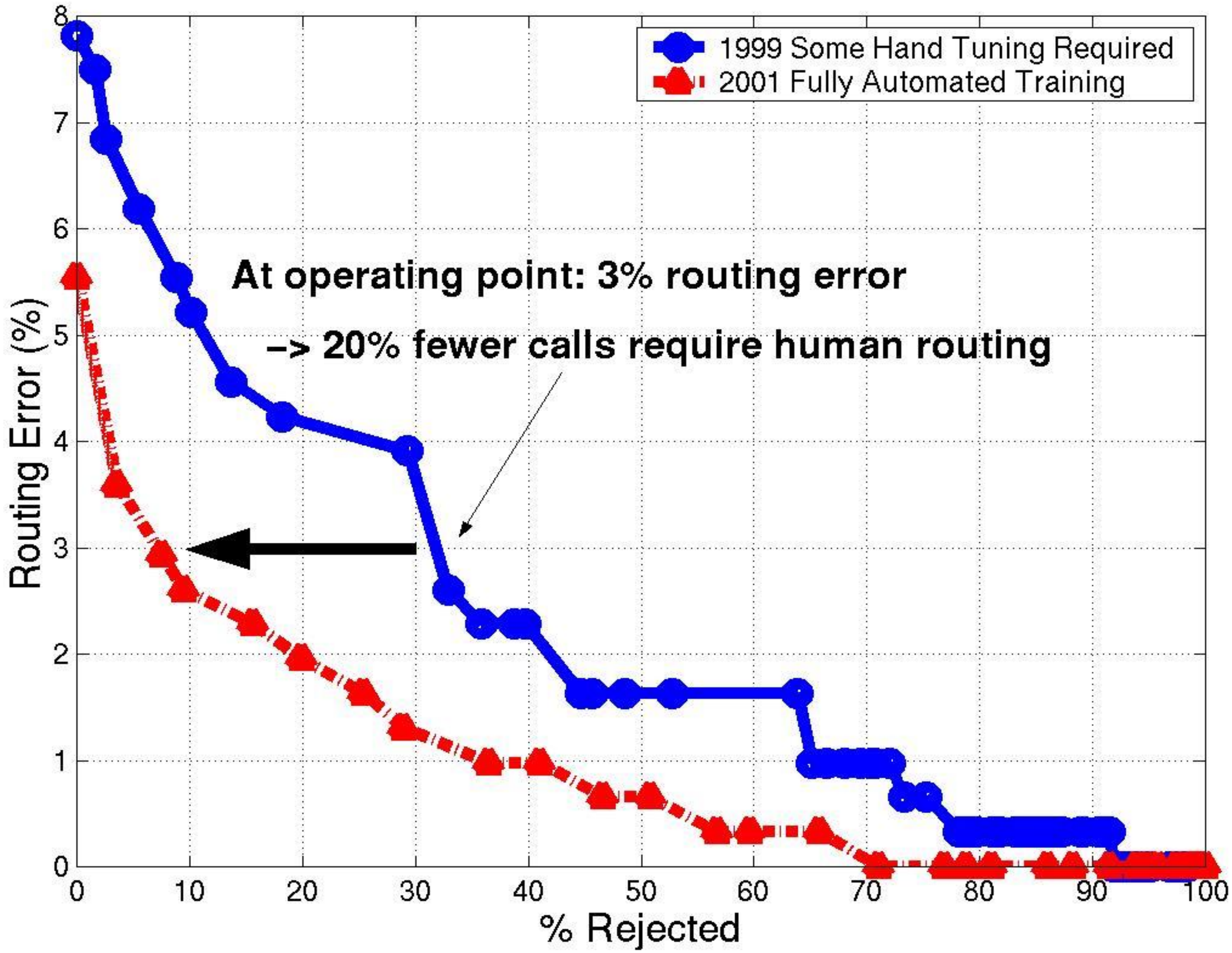
- Disambiguation power selection: gives 18
  - Select terms that will form an unambiguous query
- Select terms with shared head (key indexing) words:
  - Give 11 terms with the head word "loan"
  - Generate a *WH* question: "for what type of loan?"
  - User Response: "I'd like a car loan."
- System generates a *YN* Question:
  - System: "is it about an existing loan?"
  - User Answer: "no, it is a new car loan."
- Ambiguity resolution: usually in three turns
- Generalization:
  - Search as multi-turn collaborative entropy minimization

Performance: Fully Automatic Training

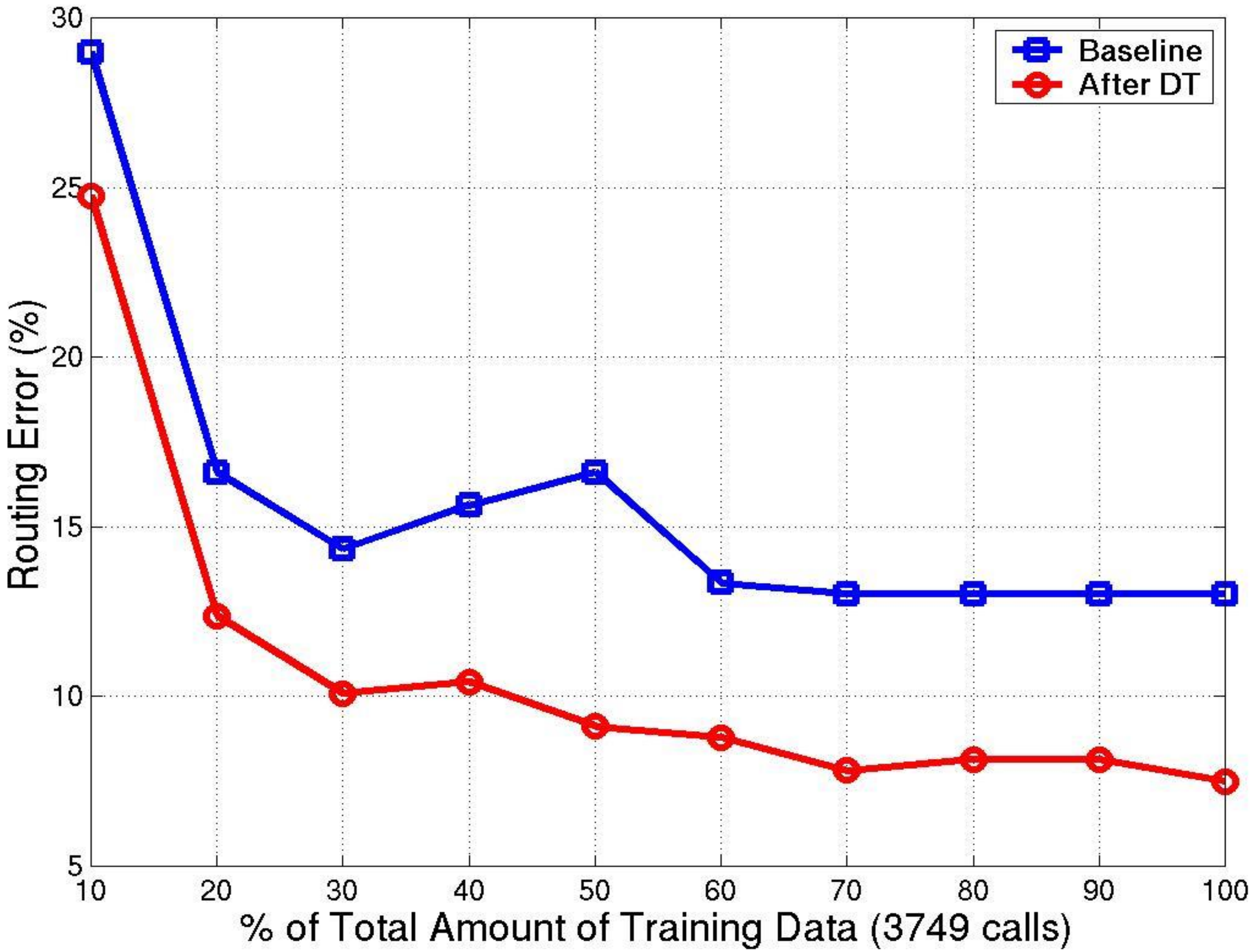
- Term extraction gives 7434 term features (2756 trigrams, 3442 bigrams and 1236 unigrams)
- LDF with only 1236 unigram-based LSA features
- Weights trained with discriminative training, or DT (Gao. *et al*, SIGIR2003, ICML2004, ACM T-IS, 2006)

Category	Text Error	Speech Error
Baseline	9.12%	12.7%
After DT	5.54%	7.82%
Improvement	39%	38%

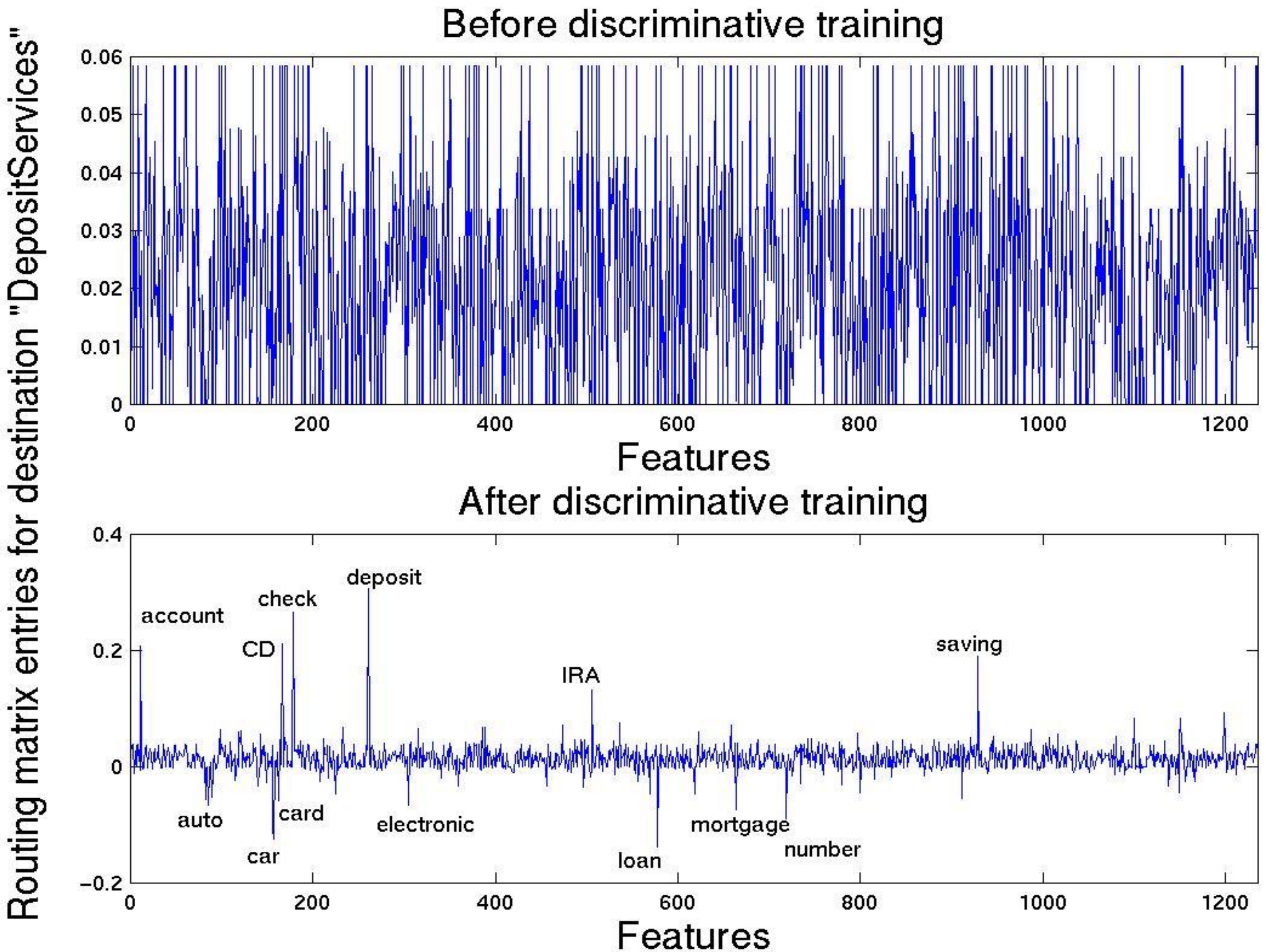
DT Improves Robustness (Kuo, *et al*, T-SAP, 2003)



DT Reduces Training Requirements



DT with Features & Anti-Features

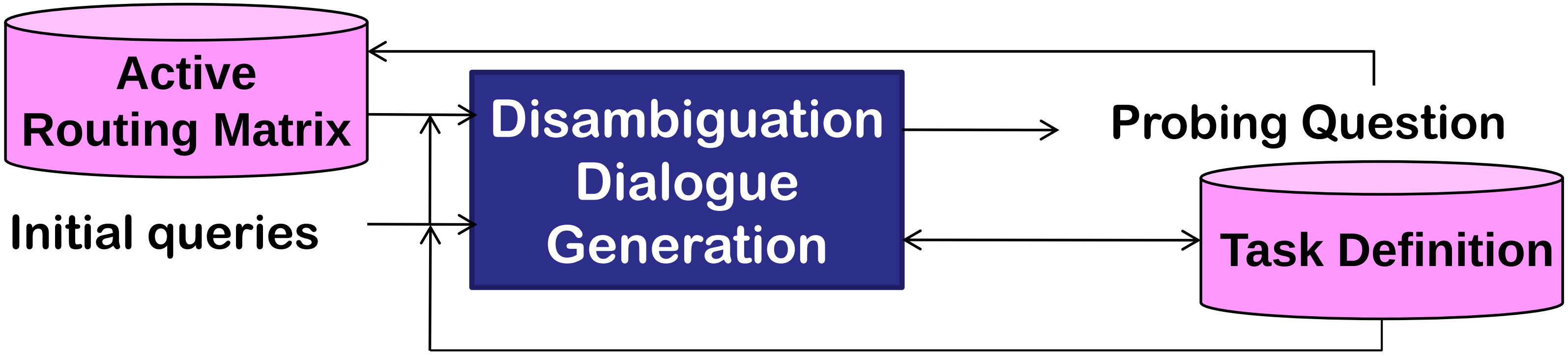


USAA Field Trial Results

- "Trained" human agents - 87% correct routing
- 1309 calls with USAA customers for 2 weeks
  - 96.2% accuracy (8.5% "rejected" to human agents)
  - 93% of customers surveyed show non-negative preference
- NLCR: exceeding USAA expectation
  - Perform better than the human agents
  - Cut down connection time greatly (from 80 to 20 seconds)
- Why USAA went to Nuance eventually?
  - Lucent did not know how to price solutions !!
- *Newsweek* issue on speech business (2001)

Search with Dialogue Disambiguation

- Search as a goal-driven, system-initiated dialog process
  -- Why generating a list not giving specific answers?
  -- Accommodating both novice and expert users
- Search as a collaborative ambiguity minimization problem
  -- Focusing on document and term after each turn taking
  -- Probing actively seeking efficient and effective results
- Progressive task information integration and refinement
  -- Adjustable term/term & document/document distances
  -- Usable dialog history in the current and past sessions



From Single- to Multi-Turn Dialogues

- Technology dimensions
  - Goal: user's intents from the system (task-defined)
  - Attribute: properties used to identify a goal
  - State: current dialogue situation
  - Action: system questions and user responses
  - Policy: system strategies about what actions to take
  - History: sequences of system questions & user responses
- Dialog management (DM)
  - Maintaining the states of the dialogue process
  - Acting according to system policies and user responses
- Search: as a collaborative multi-turn goal-driven dialogue

Dialogue Management Approaches

- Conventional techniques
  - Rule-based with semantic frame-filling or graph-directed
  - Deterministic and pre-defined states
- Statistical techniques: data not easy to collect
  - Markov decision process (MDP): states are discrete, and described by a few simple components, often not scalable
  - Partially observable Markov decision process (POMDP): addressing ASR and NLU errors, a “belief state” is used for state probability distributions at a specific time
- Our proposal: entropy minimization DM (EMDM)
  - Collaborative, goal-driven, task-based, no training
  - DS-states: constructed dynamically and stochastically

An MDP-based State Sequence

- Album Disney's "Frozen" was first given by a user, but the goal "Let It Go" could only be reached after all the discrete components in each state are filled

Singer	=	?
Gender	=	?
Region	=	?
Album	=	?
Company	=	?
Language	=	?
Lyricist	=	?
Composer	=	?
Live	=	?
Time	=	?
Style	=	?
Emotion	=	?

Initial State: S_0

Singer	=	?
Gender	=	?
Region	=	?
Album	=	*Frozen*
Company	=	?
Language	=	?
Lyricist	=	?
Composer	=	?
Live	=	?
Time	=	?
Style	=	?
Emotion	=	?

S_1

……

Singer	=	*Idina Menzel*
Gender	=	*female*
Region	=	*USA*
Album	=	*Frozen*
Company	=	*Disney*
Language	=	*English*
Lyricist	=	*Lopez*
Composer	=	*Lopez*
Live	=	*no*
Time	=	*2013*
Style	=	*popular*
Emotion	=	*exciting*

Final State: S_f

Dynamic Stochastic (DS-)States

- States are dynamically and stochastically definedon the current dialogue situation, but not pre-fixed
- With additional information in the dialogue process, the search space is usually reduced gradually
- The system examines the remaining search space and formulates disambiguation questions related to the attributes with the maximum entropy in order to reduce the overall uncertainty in follow-up dialogues
- Number of turns can be minimized accordingly
- ASR and NLU errors can also be handled (later)

A Song-on-Demand (SoD) Task

- 38117 songs (goals), 10322 albums, 3020 singers
- 12 key attributes of a song and their statistics
  - Most representative: singer, album, time
  - Missing: Style (54%), Composer (50%), Emotion (20%)

ID	Attributes	Description	Value Numbers
1	*Singer*	The name of the singer	3021
2	*Gender*	The gender of the singer	2
3	*Region*	The region of the singer	19
4	*Album*	The album on which the song appears	10322
5	*Company*	The publisher of the song	1193
6	*Language*	The language of the song	10
7	*Lyricist*	The lyricist of the song	5603
8	*Composer*	The composer of the song	5642
9	*Live*	Live version or not	2
10	*Time*	The release date of the song	413
11	*Style*	The style of the song	346
12	*Emotion*	The emotion of the song	59

A Probabilistic Dialog Representation

For a task *D*, the probability of the entire dialog is:

- Overall *J*-turn prob.: $P(\mathbf{S},\mathbf{Q},\mathbf{R} \mid D) = P(\mathbf{S}_1^J, \mathbf{H}_1^J \mid D)$
- Prior prob. for each goal *i* : $P^{(0)}(g_i \mid D), g_i \in \mathbf{G}$
- Prob. of goal *i* at state *j* : $P^{(j)}(g_i \mid S^{(j)}, D) = P_i^{(j)}$
- Prob. of reaching state *j* : $P^{(j)}(S^{(j)} \mid [q^{(l)}, r^{(l)}]_{l=1}^{j}, D)$

1. Prob. of state evolution: $P_s^{(j)} = P(S^{(j)} \mid S^{(j-1)}, \mathbf{H}_1^j, D)$
2. Prob. of next system question: $P_q^{(j)} = P(q^{(j)} \mid S^{(j-1)}, D)$
3. Prob. of next user response: $P_r^{(j)} = P(r^{(j)} \mid q^{(j)}, S^{(j-1)}, D)$

➢ Prob. of current dialog situation:

$$P(q^{(j)}, r^{(j)}, S^{(j)} \mid \mathbf{S}_1^{j-1}, \mathbf{H}_1^{j-1}, D) = P_s^{(j)} * P_q^{(j)} * P_r^{(j)}$$

Goal-Oriented Entropy Characterization

- For a multi-turn dialogue, to reach a particular goal in a set of goals, $\{g_i |_{i=1}^{J}\}$, we have the following:
- Initial entropy: $E^{(0)} = \sum_{i=1}^{I} -P^{(0)}(g_i | D) \log P^{(0)}(g_i | D)$
- Entropy at state *j*: $E^{(j)} = \sum_{i=1}^{I} -P^{(j)}(g_i | D) \log P^{(j)}(g_i | D)$
- Entropy evolution through multi-turn dialogue:
  - Entropy minimization dialogue management (EMDM)

$$E^{(0)} > E^{(1)} > \ldots > E^{(J)} \geq 0$$

DS-State Goal Set and Entropy

For a multi-turn dialogue, we have the following:

(Wu, Li & Lee, T-ASLP, 2015)

- Goal set at DS-state $S^{(j)}$: $\mathbf{G}^{(j)}$
- Entropy of $S^{(j)}$: $$E^{(j)}(\mathbf{G}^{(j)}) = \sum_{i \in \mathbf{G}^{(j)}}^{I} -P_i^{(j)} \log P_i^{(j)}$$
- Prob. of *m*-th answer for *k*-th attribute in $S^{(j)}$ is $P_{k,m}^{(j)}$
- Remaining goal entropy is: $$E_{k,m}^{(j)}(\mathbf{G}^{(j)}) = -\sum_{i \in \mathbf{G}_{k,m}^{(j)}} \frac{P_i^{(j)}}{P_{k,m}^{(j)}} \log \frac{P_i^{(j)}}{P_{k,m}^{(j)}}$$
- Expected entropy reduction: $E_k\{E^{(j)} - E_{k,m}^{(j)}(\mathbf{G}^{(j)})\}$
- In [8], entropy reduction is equal to $E_k^{(j)}$, i.e., asking questions related to maximum-entropy attribute, a_k

Dialogue Example 1: Three Turns

System: What can I do for you?
User: I'd like to listen to a song.

- Attribute Album has the highest entropy (disambiguation)

System: On which album does the song appear?
User: The song is on the album "My Room"

- 9 songs left: Attribute Lyricist has now the highest entropy

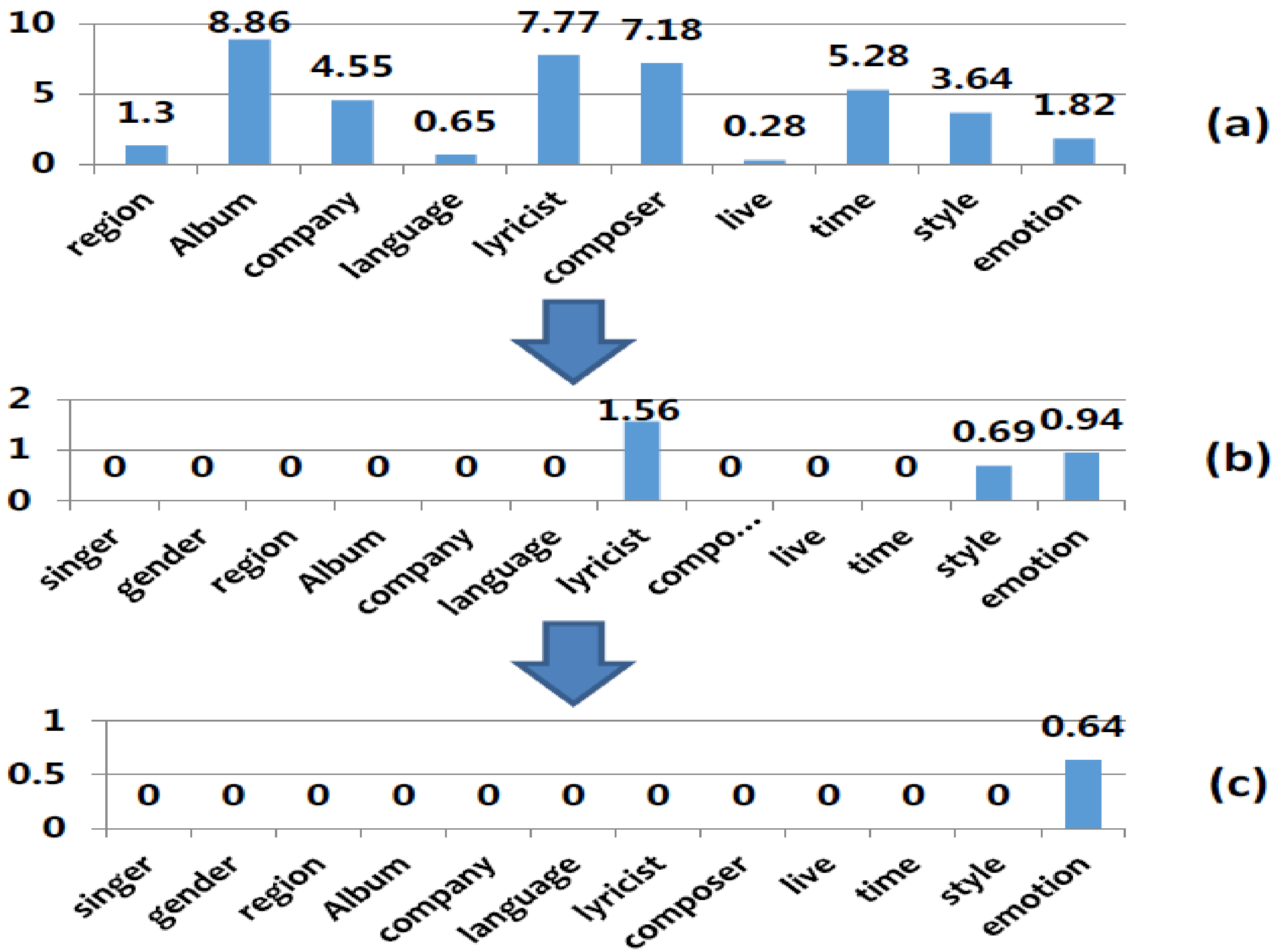
System: Who is the lyricist for that song?
User: Peggy Hsu

- 3 songs left: Attribute Emotion has now the highest entropy

System: What about the emotion?
User: It's a little bit sad.

Goal: "Under The Moonlight"

Entropy Evolution Example with EMDM



Dialogue Example 2: Four Turns

System: What can I do for you?

User: I would like a song by Maggie Chiang

System: What kind of song?

User: A popular one

System: Who is the lyricist of that song?

User: Peggy Hsu

System: On what album does that song appear?

User: It is on My Room.

Goal: "Under The Moonlight"

Experiment 1: Simulation, No Errors

- *Single knowledgeable and cooperative user*
- *System begins asking a question about a particular attribute, then updates the goal set based on the user's response. This process continues until:*
    1. *Only one song remains in the candidate set, or*
    2. *Entropy of all 12 attributes drops to zero, or*
    3. *All 12 attributes have been inquired by the system*
- *Four DM strategies are compared: sequential, random*, database summary DM (entropy-like), and *EMDM*, with the former three discussed in the *DSDM* paper (MDP/POMDP is hard to compare)

A Comparison of Average Dialog Turns

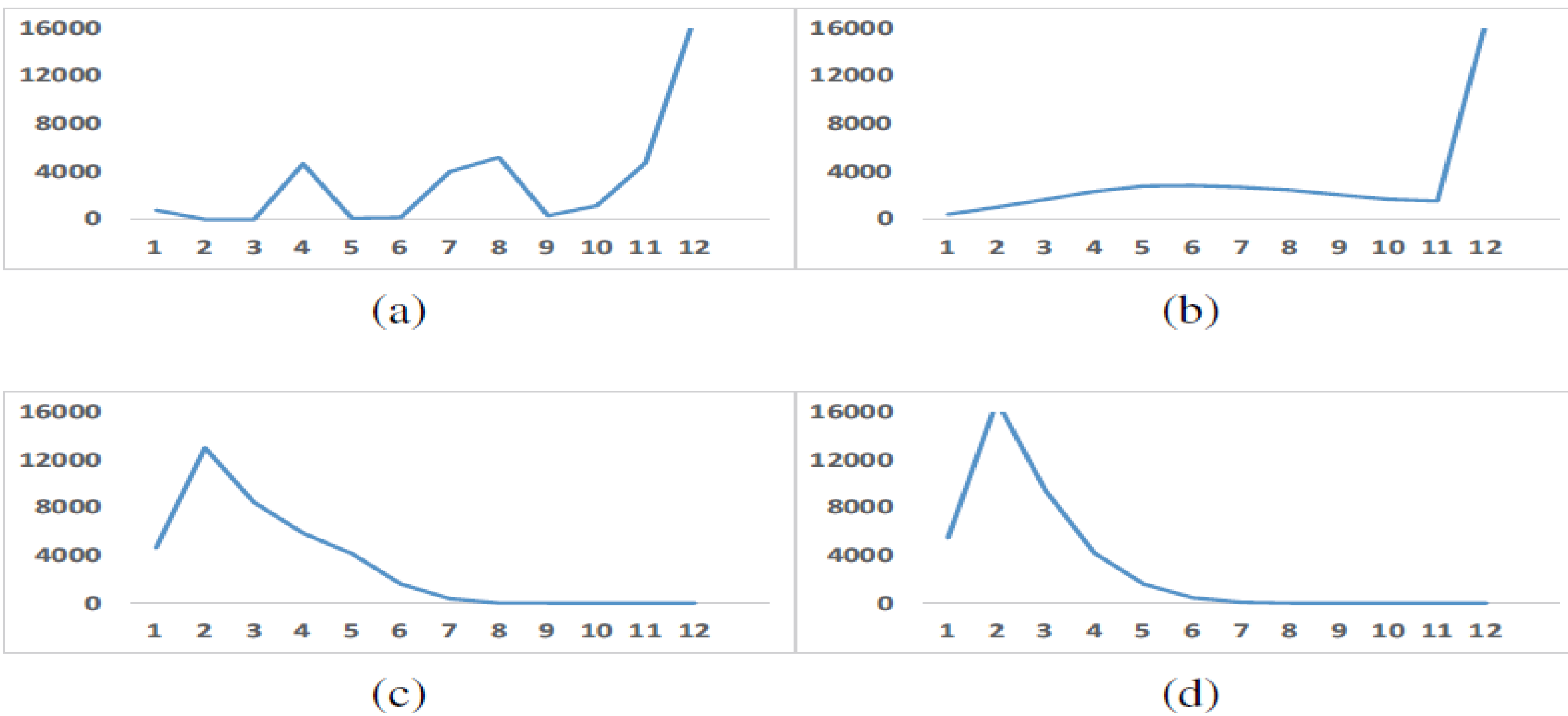
1. *Sequential*: choosing questions in a fixed order
2. *Random*: choosing attributes in a random order
3. *DSDM*: database summary DM (entropy-like)
4. *EMDM*: entropy minimization

- *The first three were discussed in Polifroni/Walker)*
- *Uniform*: no prior knowledge, uniform song density
- *Sampling*: density from dialog history, 500K times

Strategy	Sequential	Random	DSDM	EMDM
Uniform setting	9.30	8.30	3.33	3.31
Sampling setting	8.31	7.16	3.22	3.07

Histogram Comparison of Dialog Turns

- *(a) Sequential, (b) Random*, (c) *DSDM*, (d) *EMDM*
- (a) and (b) often require all 12 attributes to be asked
- (c) and (d) give less turns knowing some DB content



(a) (b) (c) (d)

Detailed Comparison of EMDM & DSDM

- *#E: number of EMDM dialogue turns*
- *#D*: *number of DSDM dialogue turns*
- Both "probabilistic" strategies perform similarly in the uniform attribute selection setting
- *EMDM* works much better than DSDM when they perform differently (about 17%) in sampling setting

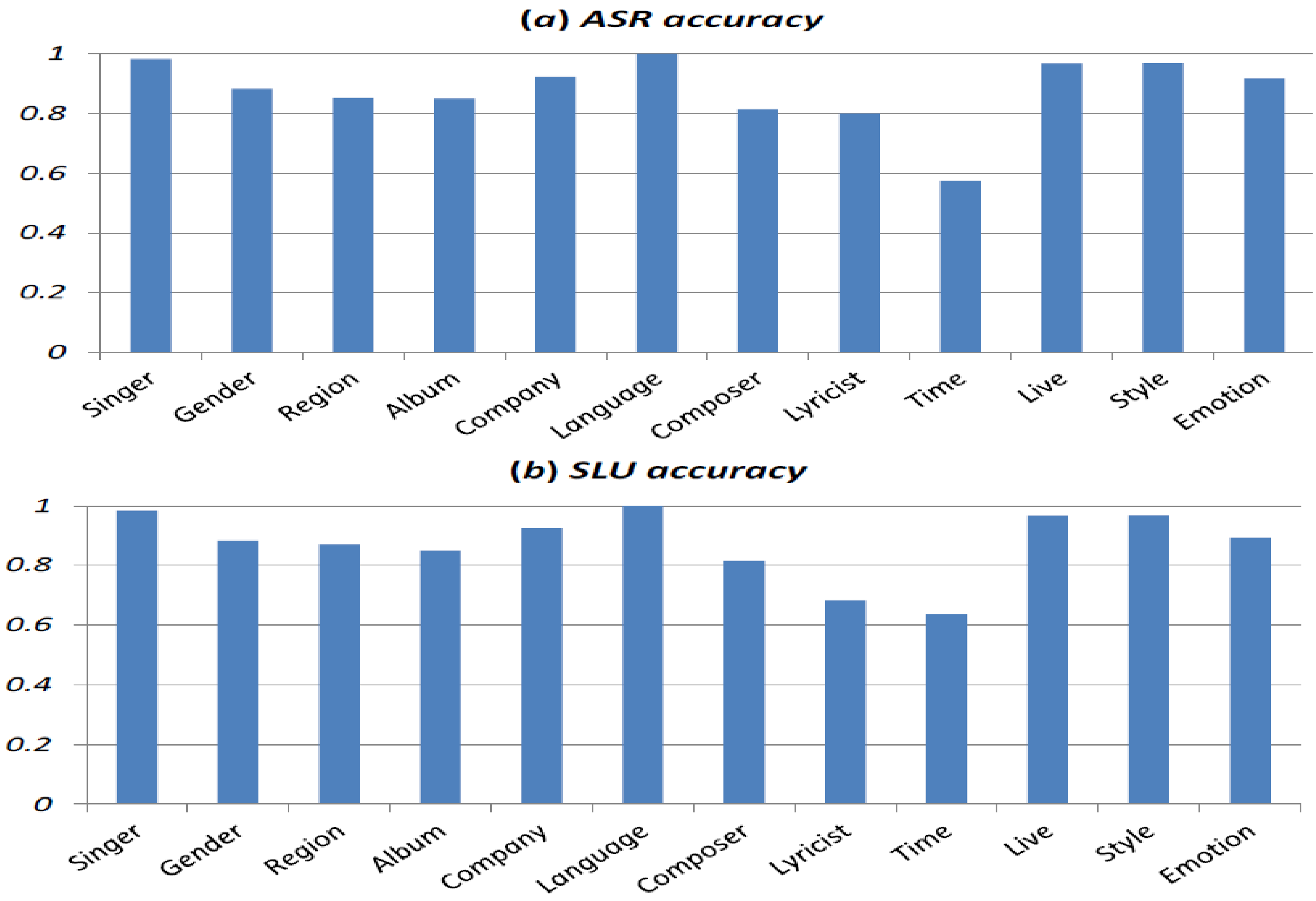
Strategy	**#E<#D**	**#E=#D**	**#E>#D**	**total**
Uniform	4.09%	93.68%	2.23%	38117
Sampling	15.38%	82.75%	1.87%	500,000

Experiment 2: with ASR & SLU Errors

- *Online with 6 users, 10 songs each for 60 requests*
- *For DSDM and EMDM, top SLU candidates can be used to update DS-state to get follow-up questions*

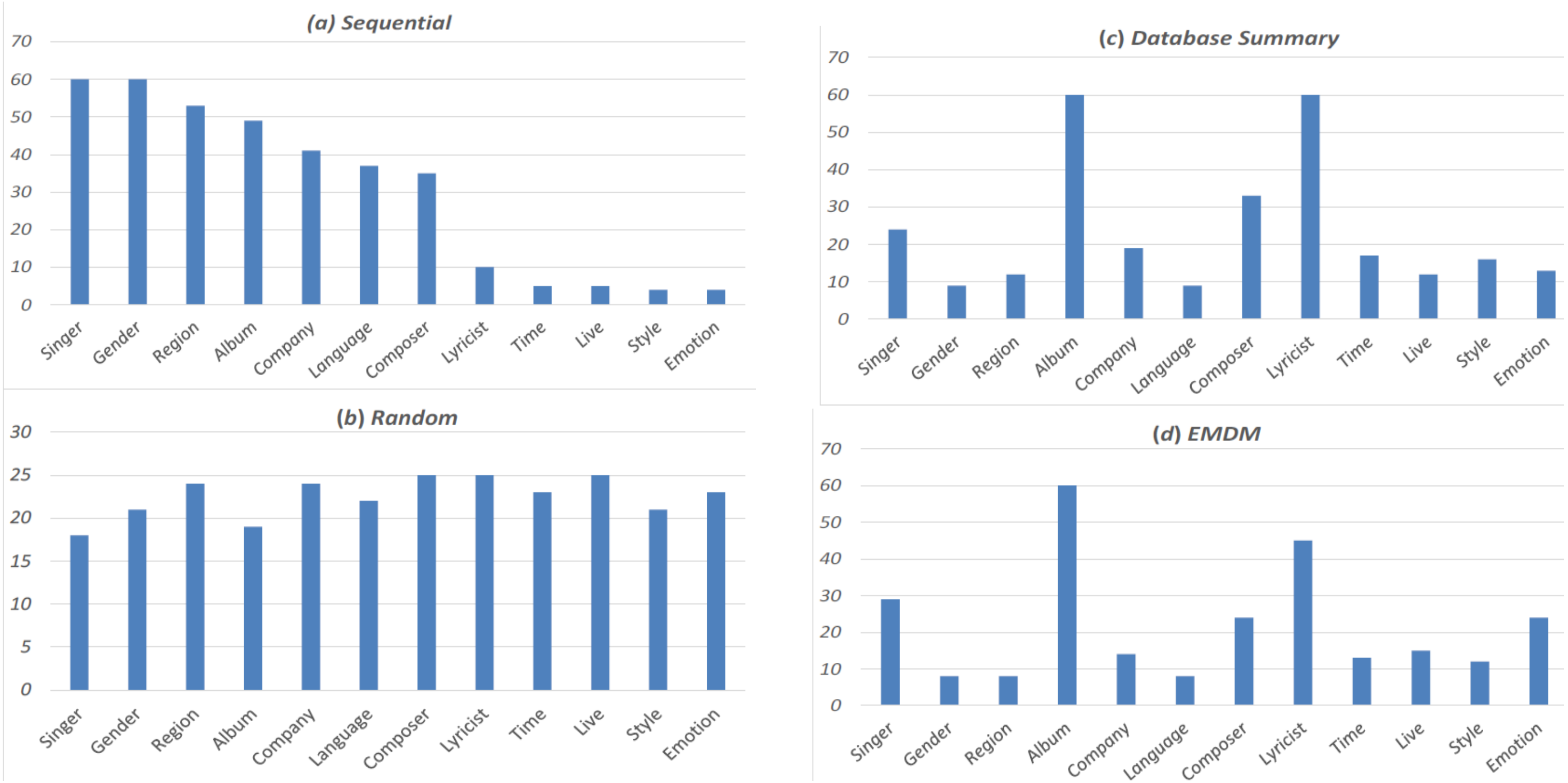
Strategy	**Sequential**	**Random**	**DSDM (Top 5)**	**EMDM (Top 5)**
ASR accuracy	90.9%	89.3%	84.5% (88.7%)	85.4% (89.2%)
SLU accuracy	90.6%	88.5%	82.7% (88.4%)	83.5% (88.8%)
Dialog success rate	50.0%	61.7%	80.0%	86.7%
# of dialog turns	8.75	6.23	5.63	5.17

Accuracies with ASR/SLU Errors



Distribution of the Attribute Questions

- *Sequential*: later attributes were less inquired
- *Random*: uniform distributions
- *DSDM* and EMDM: similar distribution



Summary

- Text categorization: a unifying theme for multimedia document search and retrieval
- Call routing: multi-turn IR dialogue for search
- Stochastic representation of dialogs
- Dynamic stochastic (DS-)state and entropy
- EMDM outperforms competing dialog strategies
  - A new system-initiated DM strategy with no training
- Tunable DM: a simulation tool for data collection?
- JDAI's recent goal-driven competition: new interest?

Key References

1) C.-H. Lee, R. Carpenter, W. Chou, J. Chu-Carroll, W. Reichl, A. Saad, and Q. Zhou, "On Natural Language Call Routing," *Speech Communication*, Vol. 31, pp. 309-320, 2000.
2) H.-K. J. Kuo and C.-H. Lee, "Discriminative Training for Robust Natural Language Call Routing," *IEEE Trans. on Speech and Audio Proc.*, Vol. 11, No.1, pp. 24-35, Jan. 2003.
3) S. Gao, W. Wu, C.-H. Lee and T.-S. Chua, "A Maximal Figure-of-Merit Learning Approach to Text Categorization," *ACM SIGIR*, pp. 174-181, Toronto, Canada, July 2003.
4) S. Gao, W. Wu, C.-H. Lee and T.-S. Chua, "A Maximal Figure-of-Merit (MFoM) Learning Approach to Robust Classifier Design for Text Categorization," *ACM Trans. on Information Systems*, Vol. 2, No. 4, Issue 2, pp. 190-216, April 2006. (text)
5) H. Li, B. Ma and C.-H. Lee, "An Acoustic Segment Modeling Approach to Universal Acoustic Characterization and Spoken Language Identification," *IEEE Trans. Audio, Speech and Language Proc.*, Vol. 15, No. 1, pp. 271-284, January 2007. (spoken language ID)
6) J. Reed and C.-H. Lee, “Preference Music Ratings Prediction Using Tokenization and Minimum Classification Error Training,” *IEEE Trans. Audio, Speech and Language Proc.*, Vol. 19, No. 8, pp. 2394-2303, Nov. 2011. (music and audio retrieval)
7) I. Kim and C.-H. Lee, “An Efficient Gradient-based Approach to Optimizing Average Precision through Maximal Figure-of-Merit Learning,” *Journal of Signal Processing Systems,* Vol. 64, No. 9, pp. 1-11, Sept. 2013. (image/video retrieval)
8) J. Wu, M. Li and C.-H. Lee, “A Probabilistic Framework for Representing Dialog Systems and Entropy-Based Dialog Management through Dynamic Stochastic State Evolution,” *IEEE/ACM Trans. Audio, Speech and Language Proc.* Vol. 23, No. 11, pp. 2026-2035, 2015*. (speech)*

Acknowledgment

Thank You.