Unsupervised Modeling of Dialog Acts in Asynchronous Conversations

Shafiq Joty, Giuseppe Carenini
University of British Columbia
Vancouver, Canada.

Chin-Yew Lin
MSRA
Beijing, China.

Asynchronous conversations and their dialog structure

- Participants collaborate at different times.
  - Emails, Blogs, Fora.
- Interaction is conversational.
  - Take turns.
  - A turn: joint action of writing and reading.
- Perform comm. acts (dialog acts) like question, answer, request, accept.
- Dialog structure: dialog acts, dialogic structure (adjacency pairs).

Example

From: Charles ***To:*** WAI AU Guidelines ***Date:*** Thu May ***Subj:*** Phone connection to ftof meeting.

It is probable that we can arrange a telephone connection, to call in via a US bridge. <Statement>

Are there people who are unable to make the face to face meeting, but would like us to have this facility? <Yes-No Question>

From: William ***To:*** Charles ***Date:***Thu May ***Subj:*** Re: Phone connection to ftof meeting.

>Are there people who are unable to make the face to face meeting, but would like us to have this facility?

At least one "people" would. <Accept Response>

……………………...

From: Charles ***To:*** WAI AU Guidelines ***Date:*** Mon Jun *Subj:* RE: Phone connection to ftof meeting.

Please note the time zone difference, and if you intend to only be there for part of the time let us know

which part of the time. <Action Motivator>

9am - 5pm Amsterdam time is 3am - 11am US Eastern time which is midnight to 8am pacific time. <Statement>

Until now we have got 12 people who want to have a ptop connection. <Statement>

Cheers, <Polite>

Motivation of DA modeling

- Important step towards conversation analysis.
- Useful applications (spoken dialog):
  - Artificial companions [Wilks 2006].
  - Task learning agents [Allen et al. 2007].
  - Meeting summarization [Murray et al. 2006].
  - Flirtation detection [Ranganath et al. 2009].
- We believe similar benefits will also hold for written asynchronous conversation.
- Abstractive summarization and visualization.

Challenges

- Very little work in asynchronous domain.

- In synchronous spoken domains (meetings, phone)
  - Conversational flow is sequential.
  - Supervised sequence labeler (HMM, MEMM, CRF).
  - Applied to the temporal order.
- But, in asynchronous domains
  - Conversational flow is **not** sequential.
  - Should a model consider the sequence dependencies?
  - If yes, then how?
  - Two options: (a) temporal order, (b) graph-structural order.

- Supervised setting becomes unrealistic.
  - Number of new media grows
  - New ways of communication.

Our approach

- Unsupervised DA modeling:
  - Find the DA clusters.
  - Assign label to each cluster.
- First application to emails and fora.
- Generalize across domains.

Contributions
Outline of the rest of the talk

- Unsupervised DA Models
  - Deterministic graph-theoretic.
    - Evaluation of graph-theoretic.
  - Probabilistic conversational
    - HMM.
    - HMM+Mix.
    - Evaluation of conv. models.

- Data preparation
  - Datasets
  - Agreement
  - Graph structural data

Datasets

- 12 DA tagset (MRDA)
- Test datasets:
  - Email: 40 threads (BC3) (W3C)
  - Forum: 200 threads (TripAdvisor).
- Kappa (2 annotators):
  - 0.79 (email).
  - 0.73 (forum).
- Train datasets:
  - Email: 23,957 threads (W3C).
  - Forum: 25,000 threads (TripAdvisor).
- Thread structure:
  - Email: Yes.
  - Forum: No.

Tag	Email	Forum
Statement	69.56%	69.56%
Polite	6.97%	6.97%
Yes-no ques.	6.75%	6.75%
Action motiv.	6.09%	6.09%
Wh-question	2.29%	2.29%
Accept resp.	2.07%	2.07%
Open-end ques.	1.32%	1.32%
Ack & appre	1.24%	1.24%
Or-clause ques	1.10%	1.10%
Reject response	1.06%	1.06%
Uncert. response	0.79%	0.79%
Rhet. Question	0.75%	0.75%

Extracting conversational structure

- Sequence dependencies in the conversational models:
  - Temporal order.
  - Graph-structural order.
- Temporal order:
  - Arranged based on the arrival time.
- Graph-structural order:
  - Find graph structure of the conversation.
  - Derive the data from the structure.

Graph structure of emails

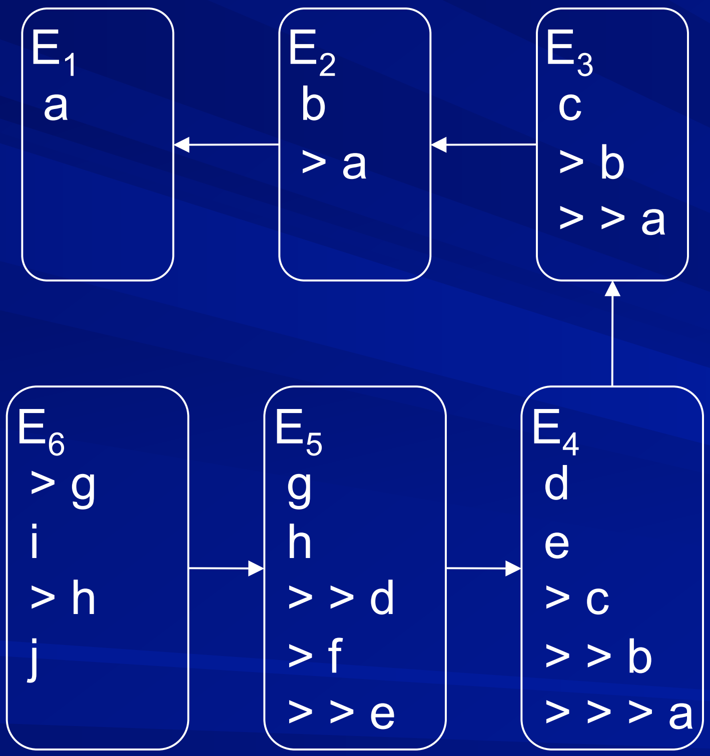
- We analyze the actual body of the emails.
- We find two kinds of fragments:
  - New fragment (depth level 0)
  - Quoted fragment (depth level > 0)
  - Example:

>Are there people who are unable to make the face to face meeting, but would like us to have this facility? (Quoted Fragment depth level 1)

At least one “people” would. (New Fragment)

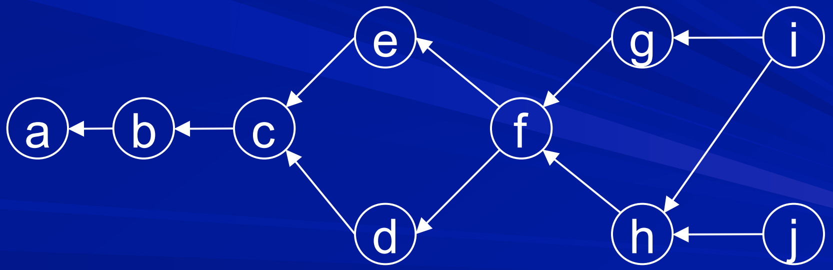
- We form a fragment quotation graph (FQG):
  - Nodes represent fragments.
  - Edges represent referential relations.

Graph structure for emails (FQG)



An email conversation with 6 emails.

- Nodes
  - Identify quoted and new fragments
- Edges
  - Neighbouring quotations



- a-b-c-e-f-g-i
- a-b-c-d-f-h-j

Graph structure for TripAdvisor

- Graph structure for TripAdvisor:
  - No thread structure.
  - Hardly quote.
  - Almost always respond to the initial post.
  - Mention names to respond to their post.

- A post usually responds to the initial post unless it mentions other participants' names.

Contributions
Outline of the rest of the talk

- Data preparation
  - Datasets
  - Agreement
  - Graph structural data

- Unsuper. DA Models
  - Deterministic graph-theoretic.
    - Evaluation of graph-theoretic.
  - Probabilistic conversational
    - HMM
    - HMM+Mix
    - Evaluation of conv. models.

Graph-based clustering

- Similar sentences should receive same DA tag.
- Form a complete similarity graph G= (V, E)
  - Nodes V represent the sentences.
  - Edge weights w (a,b) represent similarity.
- Formulate the clustering problem as a N-mincut graph-partitioning problem.
- Find optimal clusters using '*normalized cut'* criteria (Shi & Malik, 2000).
- The edge weights can be assigned in various ways.

Similarity metrics as edge weights

- TF.IDF-based cosine similarity.
- TF.IDF-based cosine similarity with nouns masked.
- Word Subsequence Kernel (WSK).
- Extended WSK with POS (ESK-P).
- BE-based dependency similarity.
- Syntactic tree kernel (TK).

Evaluation of graph-theoretic clustering

- Mean 1-to-1 accuracy:

Corpus	BOW	BOW-M	WSK	ESK-P	BE	TK	All	Baseline
Email	62.6	34.3	64.7	24.8	39.1	22.5	26.0	70.0
Forum	65.0	38.2	65.8	36.3	46.0	30.1	32.2	66.0

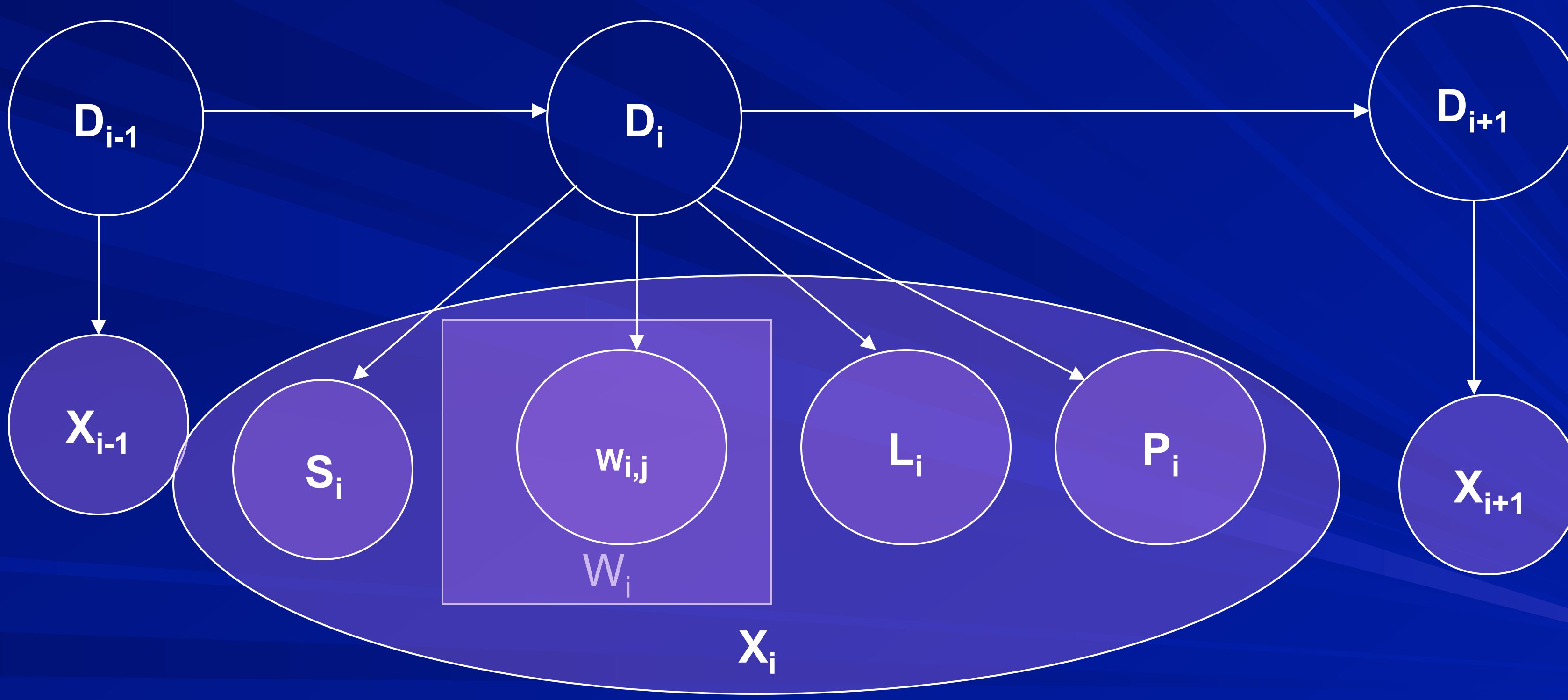
- None can beat the baseline (majority class).

Limitations graph-theoretic model

- Doesn't model sequential structure.
  - e.g., "question" followed by "answer"
- Confused by topical clusters.
- Doesn't allow to incorporate other crucial conversational features (e.g., speaker, length, relative position) in a principled way.

HMM conversational model

- D := Dialog Act, X = Feature vector
- W_{ij} := word, S := Speaker, L := Length, P := Relative Position

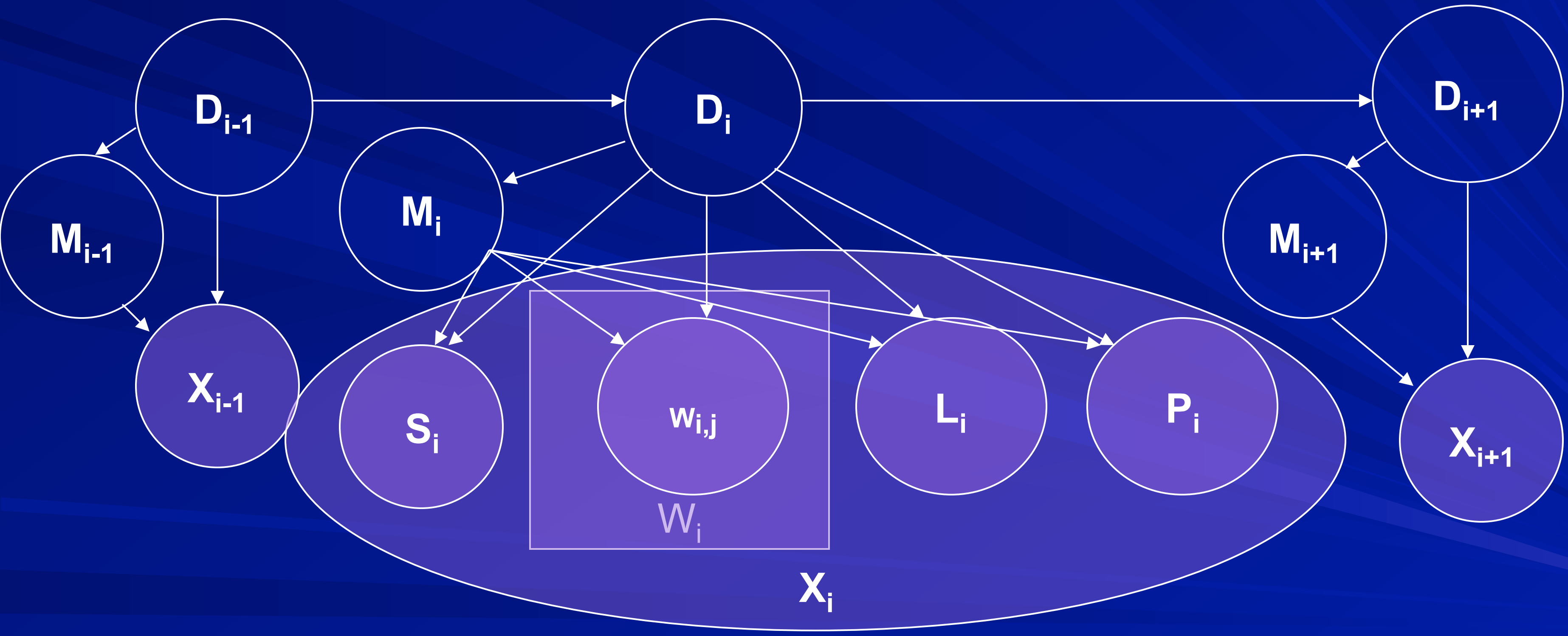


Limitation of HMM conv. model

- Basically a content model. (Regina, 2004).
- Conversational features are also important to find topical clusters. (Joty et al, 2011).
- Without additional guidance it tends to find topical clusters in addition to DA cluster.
- Changing the data in an attempt to abstract away the topic words didn't work.
- We need the model to account for this.

HMM+Mix conversational model

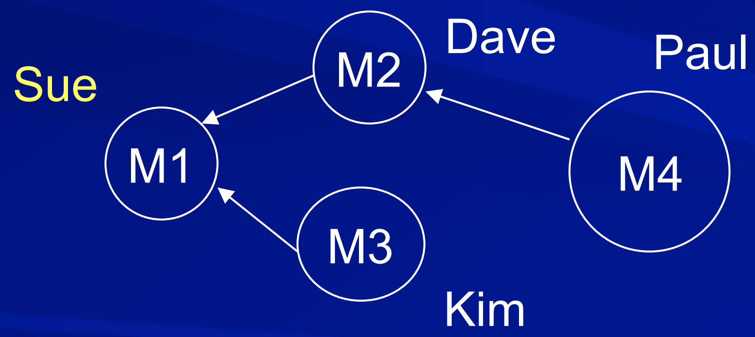
- D := Dialog Act, X = Feature vector, M = Mixture component
- W_{ij} := word, S := Speaker, L := Length, P := Relative Position



- Emission distribution is defined as a mixture model.
- Not only explain away the topics but also enrich the emission

Learning & Inference in the conversational models

- Symmetric Dirichlet prior (alpha=2) over all multinomials.
- Baum-Welch (EM) with forward backwards. [See paper]
- EM initialization: Multiple (10) restarts.
- Viterbi decoding to infer the most probable sequence.



- M1-M3
- M1-M2-M4

- Use maximum vote for the duplicated sentences in the graph-structural order.

Experimental setup

- Train on randomly selected 12,000 conversations (having at least two posts in each of them) for each corpus.
- Repeat this 50 times.
- Number of DAs available was set to 12.
- Number of mixture component M in HMM +Mix was empirically set to 3.

Evaluation of conversational models

	Email		Forum	
	Temporal	Graph	Temporal	Graph
Baseline	70.00	70.00	66.00	66.00
HMM	73.45	76.81	69.67	74.41
HMM+Mix	76.73	**79.66**	75.61	**78.35**

- Models learn better sequential dependencies with the graph-structural order ($p<0.05$).
- HMM+Mix is a better conversational model ($p<0.05$).

Future work

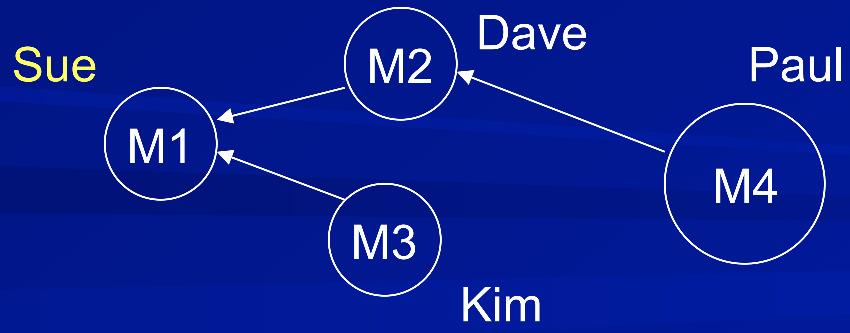
- Bayesian versions of the conversational models.
- Apply to other conversational modalities.
- Try domain adaptation.

Questions?

Thanks

Graph structure for TripAdvisor

- Graph structure for TripAdvisor (MSRA):
  - No thread structure.
  - Hardly quote.
  - Almost always respond to the initial post.
  - Mention names to respond to their post.
- A post usually responds to the initial post unless it mentions other participants' names.

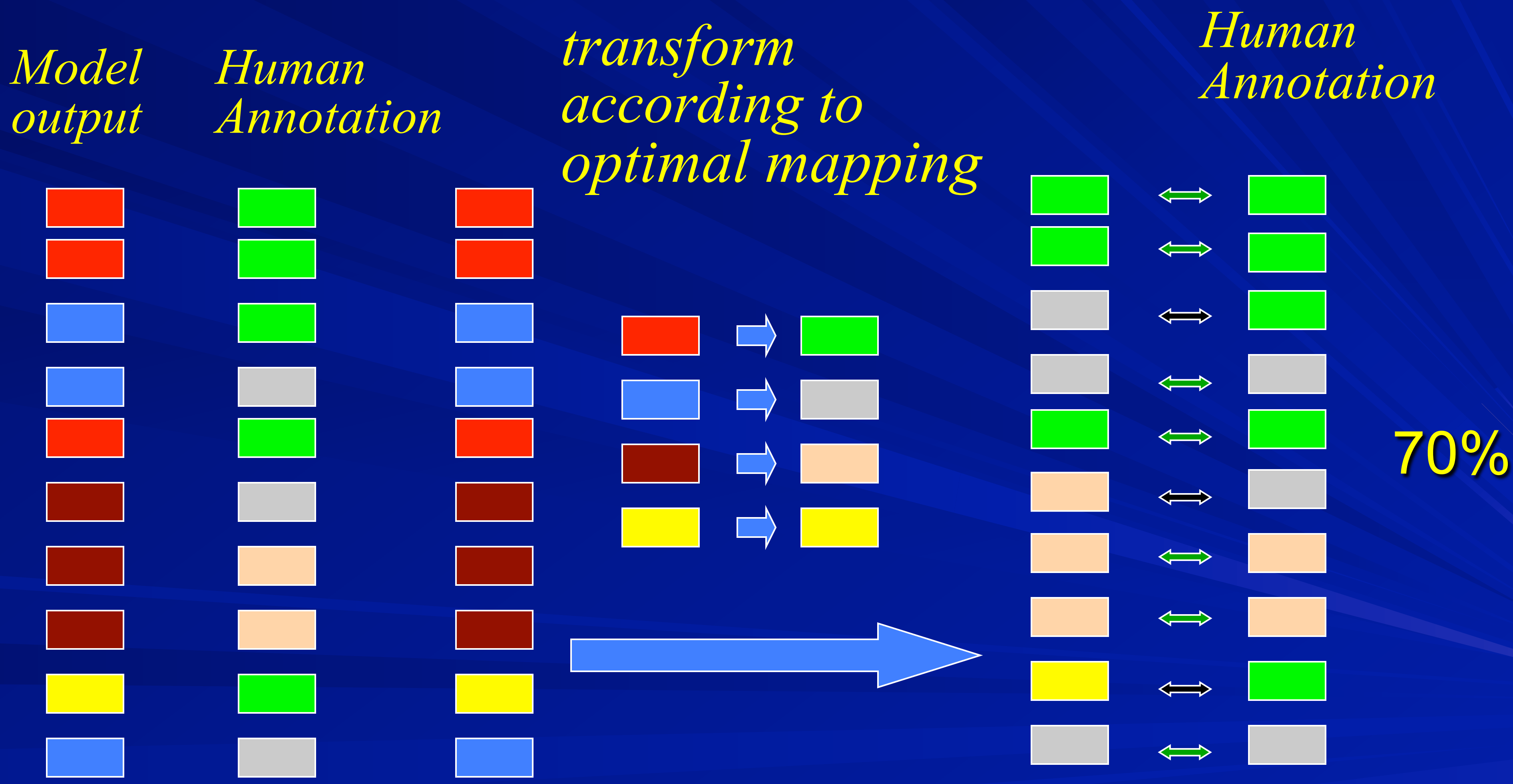


- M1-M3
- M1-M2-M4

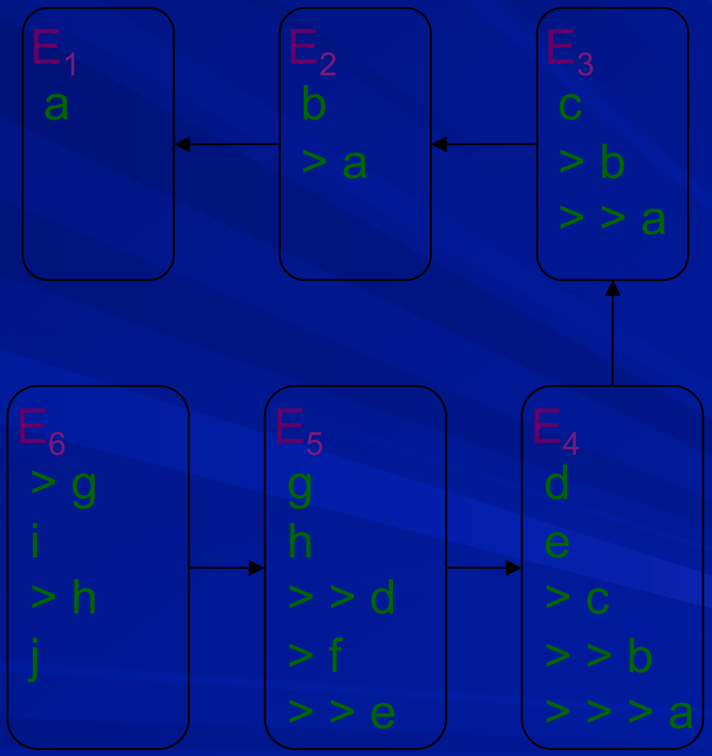
Evaluation Metric

- Compare model output with human DA annotation.
- Unsup. clustering doesn't assign any DA label.
- Metrics like kappa, F1 score are not applicable.
- We use 1-to-1 metric (Elsner & Charniak, ACL08).
- *1-to-1 measures the global similarity by pairing up the clusters of 2 annotations to maximize the total overlap.*

1-to-1



Model output
Human Annotation
transform according to optimal mapping
Human Annotation
70%


E_1
a
E_2
b
> a
E_3
c
> b
> > a
E_6
> g
i
> h
j
E_5
g
h
> > d
> f
> > e
E_4
d
e
> c
> > b
> > > a