KGO at the NTCIR-12 Temporalia Task

Exploring Temporal Information in Search Queries

Xin Kang*, Yunong Wu, Fuji Ren

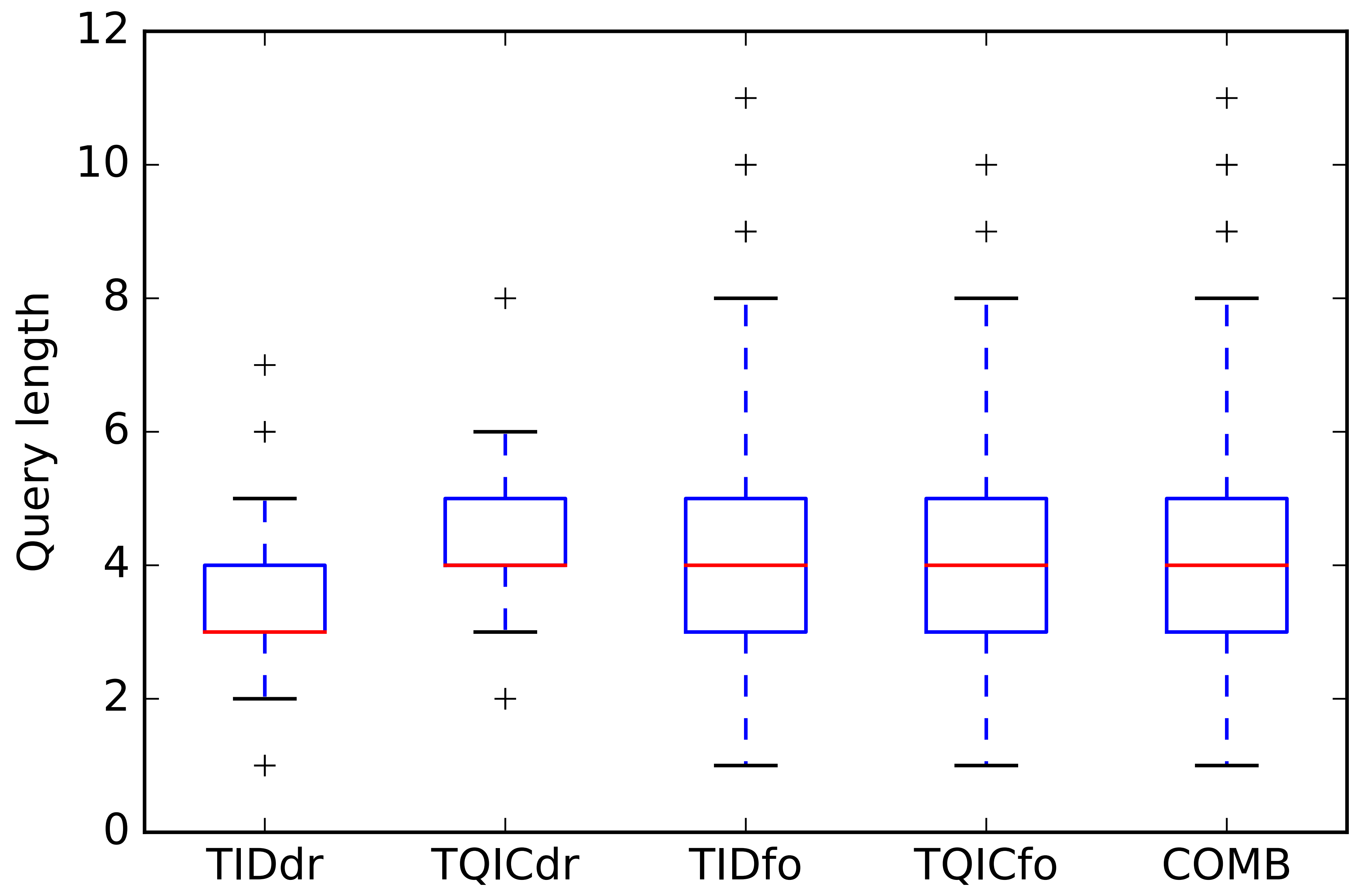
Tokushima University

Introduction

- Temporal Intent Disambiguation (TID)

May 1, 2013

		Pa	Re	Fu	At
when was electricity invented		1.0	0.0	0.0	0.0
beer night 1974		1.0	0.0	0.0	0.0
release date for iPhone 6	Sep. 19, 2014	0.0	0.0	1.0	0.0
memorial day	May. 27, 2013	0.0	0.0	0.6	0.4



Introduction

- Temporal Intent Disambiguation (TID)
  - Multi-label classification with probabilistic interpretation
  - Queries are too short → TID requires extra knowledge.
  - Temporal feature extraction & neural network

Temporal Feature Extraction

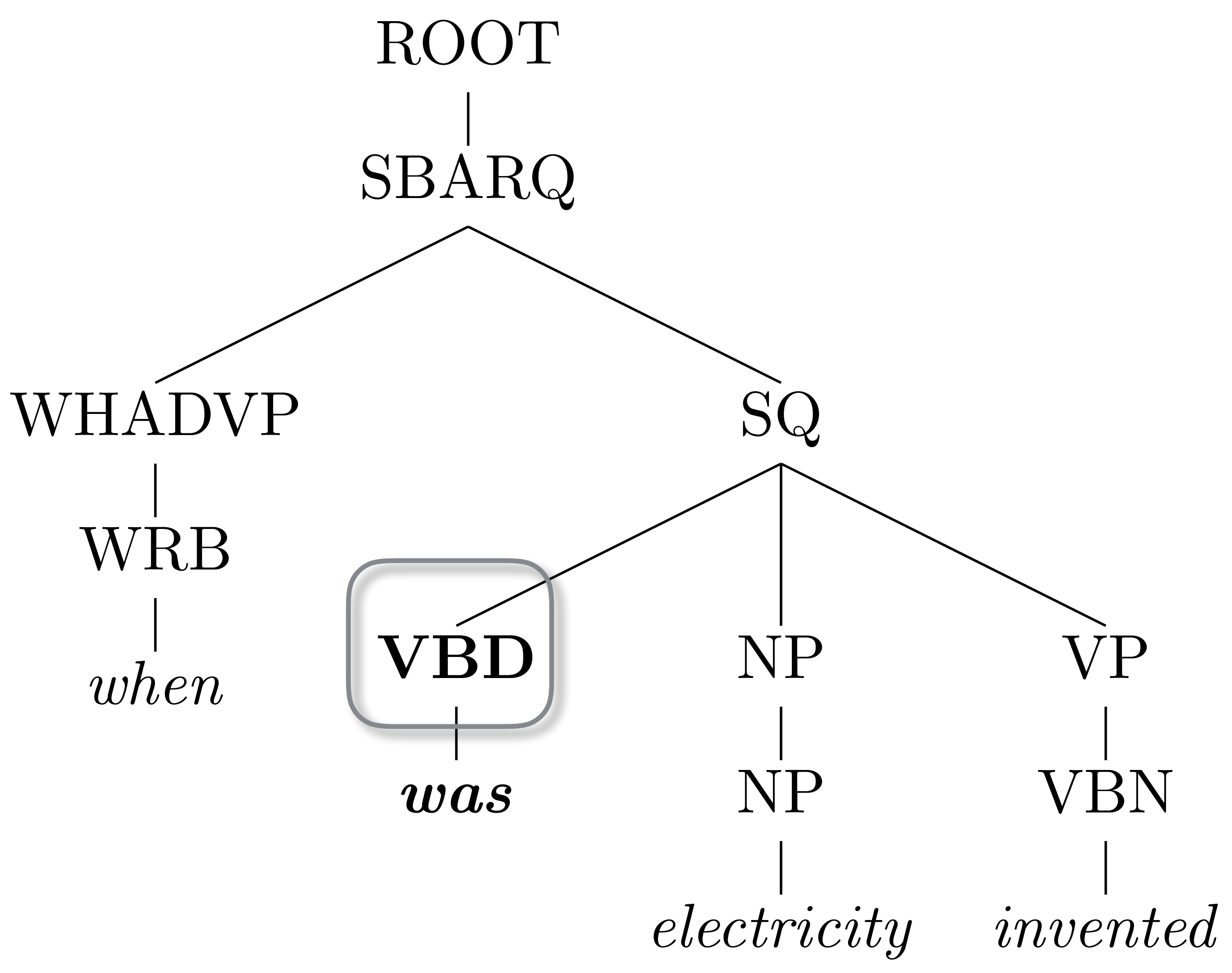
- Explicit Features: Uppermost Verb Tense & Time Gap
- Inexplicit Features: Temporal Named Entity & Holiday
- Other Features: People & Time & Lemma

Pa	Re	Fu	At	Time	Query	Feature
1.0	0.0	0.0	0.0	May 1, 2013	when **was** electricity invented	UVT_VBD
1.0	0.0	0.0	0.0	May 1, 2013	beer night **1974**	DIFF_past
0.0	0.0	1.0	0.0	May 1, 2013	release date for **iPhone 6**	DIFF_future
0.0	0.0	0.6	0.4	May 1, 2013	**memorial day**	DIFF_future

Temporal Feature Extraction

- Explicit Features
  - Upper Verb Tense feature

    indicates the tense of a query, which could be directly obtained by picking tense information (part-of-speech tag) from the main (uppermost) verb in a query.



Pa	Re	Fu	At	Time	Query	Feature
1.0	0.0	0.0	0.0	May 1, 2013	when **was** electricity invented	UVT_VBD

Temporal Feature Extraction

- Explicit Features
  - Time Gap

    represents the difference between an explicit time expression like year 1974 and the query submission time, with year, month, day, season, and period normalized for time differentiation.

	1974	1974-XX-XX
−	May 1, 2013	2013-05-01
		DIFF_past

Pa	Re	Fu	At	Time	Query	Feature
1.0	0.0	0.0	0.0	May 1, 2013	beer night **1974**	DIFF_past

Temporal Feature Extraction

- Inexplicit Features
  - Temporal Named Entity

    extracts temporal information of a named entity in query, by exploring the temporal knowledge from Wikipedia.

Step 1. Get Wikipedia summary of iPhone 6

The iPhone 6 and iPhone 6 Plus are smartphones designed and marketed by Apple Inc. The devices are part of the iPhone series and were unveiled on September 9, 2014, and released on September 19, 2014. The iPhone 6 and iPhone 6 Plus jointly serve as successors to the iPhone 5C and iPhone 5S.

Pa	Re	Fu	At	Time	Query	Feature
0.0	0.0	1.0	0.0	May 1, 2013	release date for **iPhone 6**	DIFF_future

Temporal Feature Extraction

- Inexplicit Features
  - Temporal Named Entity

    extracts temporal information of a named entity in query, by exploring the temporal knowledge from Wikipedia.

Step 2. Extract the Time Gap features:

unveiled	September 9, 2014	2014-09-09
−	May 1, 2013	2013-05-01
		DIFF_future
released	September 19, 2014	2014-09-19
−	May 1, 2013	2013-05-01
		DIFF_future

Pa	Re	Fu	At	Time	Query	Feature
0.0	0.0	1.0	0.0	May 1, 2013	release date for **iPhone 6**	DIFF_future

Temporal Feature Extraction

- Inexplicit Features
  - Temporal Named Entity

    extracts temporal information of a named entity in query, by exploring the temporal knowledge from Wikipedia.

Step 3. Resolve the context correlation:

$$s(t,e) = \cos\left(\frac{1}{|C(t)|}\sum_{w \in C(t)} \mathrm{v}(w), \frac{1}{|C(e)|}\sum_{w \in C(e)} \mathrm{v}(w)\right)$$

s(September 9, 2014, iPhone 6) = 0.2551
s(September 19, 2014, iPhone 6) = 0.3291

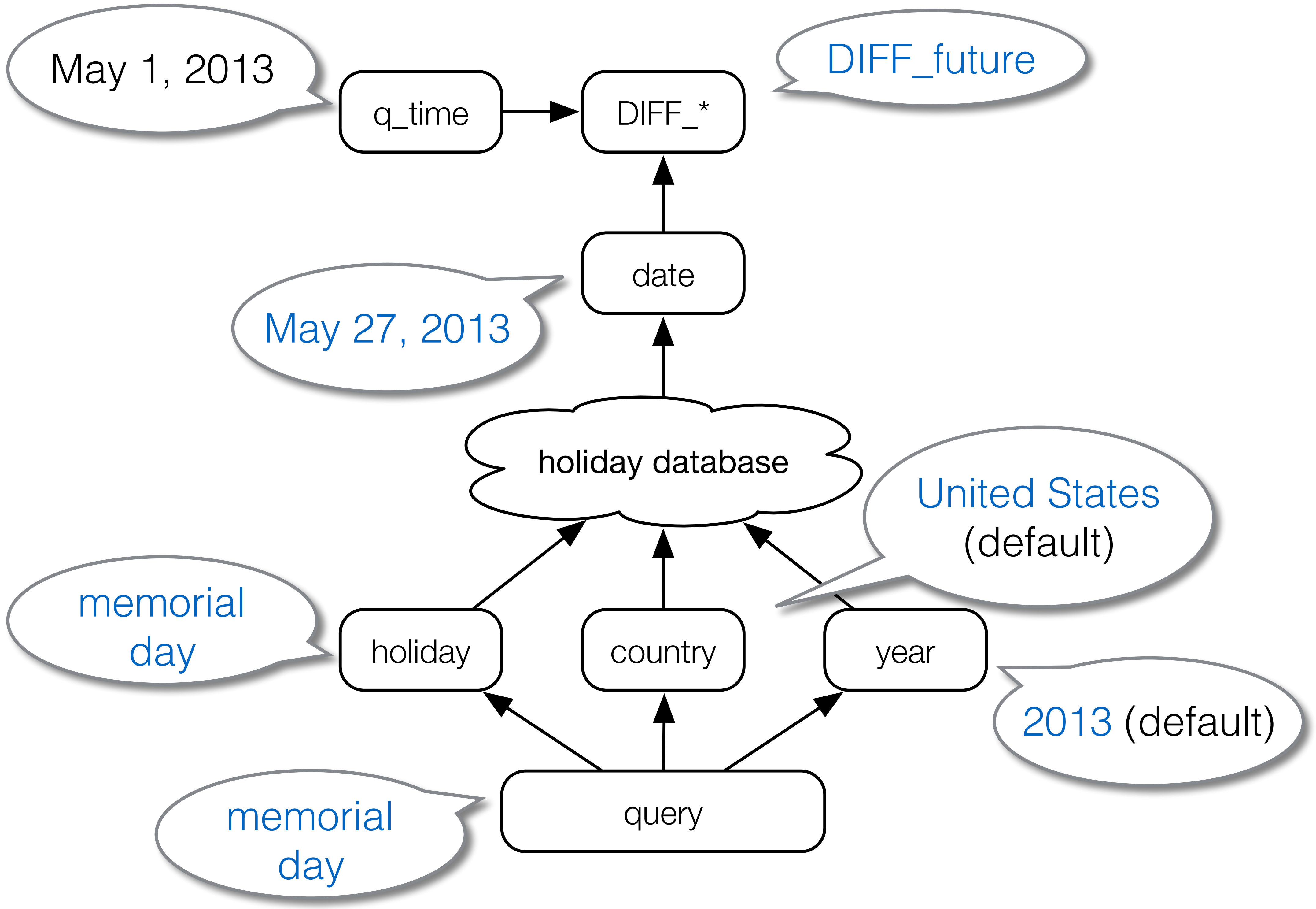
C(September 9, 2014) = {unveiled}
C(September 19, 2014) = {released}
C(iPhone 6) = {release, date, for}

Pa	Re	Fu	At	Time	Query	Feature
0.0	0.0	1.0	0.0	May 1, 2013	release date for **iPhone 6**	DIFF_future

Temporal Feature Extraction

- Inexplicit Features
  - Holiday

    extracts temporal information of holidays in a query from a holiday database (HDB):
    http://www.timeanddate.com/holidays/



Pa	Re	Fu	At	Time	Query	Feature
0.0	0.0	0.6	0.4	May 1, 2013	**memorial day**	DIFF_future

Temporal Feature Extraction

- Other Features
  - People & Time

    clusters people and events based on their semantic similarities in TextRazor Entity Extraction (to decrease the feature space and avoid over-fitting)

Super Bowl: RECURRING_EVENT

Amy Winehouse: DECEASED_PERSON, MEASURED_PERSON, PERSON

Pa	Re	Fu	At	Time	Query
0.2	0.2	0.4	0.0	May 1, 2013	what time does **Super Bowl** start
0.8	0.0	0.0	0.2	May 1, 2013	how did **Amy Winehouse** die

Temporal Feature Extraction

- Other Features
  - Lemma

    normalizes words in a query.
    Part-of-speech tags in verb lemmas
    are kept for their tense information.

when	→	when
was	→	VBD_be
electricity	→	electricity
invented	→	VBN_invent

Pa	Re	Fu	At	Time	Query	Feature
1.0	0.0	0.0	0.0	May 1, 2013	when **was** electricity invented	UVT_VBD

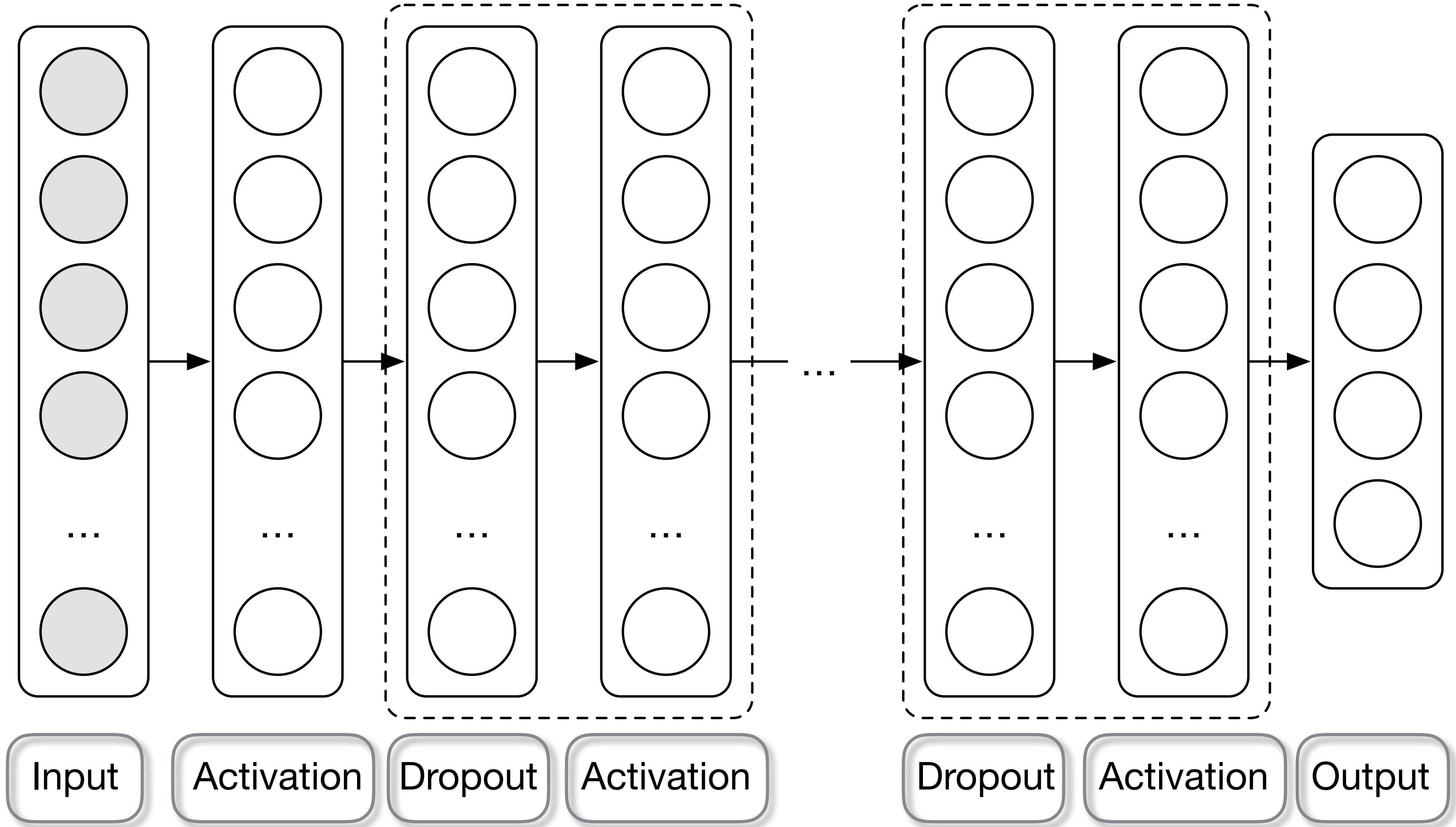
Neural Network Construction

- Generate Probabilistic Temporal Predictions

$$\tilde{p}_j = p(y = j|\boldsymbol{x}; \boldsymbol{W}) = \frac{\exp(\boldsymbol{x}^T \boldsymbol{W}_j)}{\sum_{j'=1}^{J} \exp(\boldsymbol{x}^T \boldsymbol{W}_{j'})}$$

with a cross entropy cost function

$$H(p, \tilde{p}) = -\sum_{j=1}^{J} p_j \log \tilde{p}_j$$



Deep neural network with dropout for TID. Hidden activations are selected from softplus, relu, tanh, sigmoid, hard sigmoid, and linear.

Experiments

- Experiment Setup
  - Training: TID Dry Run (93 samples) & TQIC data (400 samples)
  - Testing: TID Formal Run (300 samples)
  - Parameters are selected based on the averaged per-class absolute loss

$$\mathrm{loss}(p, \tilde{p}) = \frac{1}{J} \sum_{j=1}^{J} |p_j - \tilde{p}_j|$$

and the cosine similarity

$$\mathrm{sim}(p, \tilde{p}) = \frac{\sum_{j=1}^{J} p_j \times \tilde{p}_j}{\sqrt{\left(\sum_{j=1}^{J} p_j \times p_j\right)\left(\sum_{j=1}^{J} \tilde{p}_j \times \tilde{p}_j\right)}}$$

Experiments

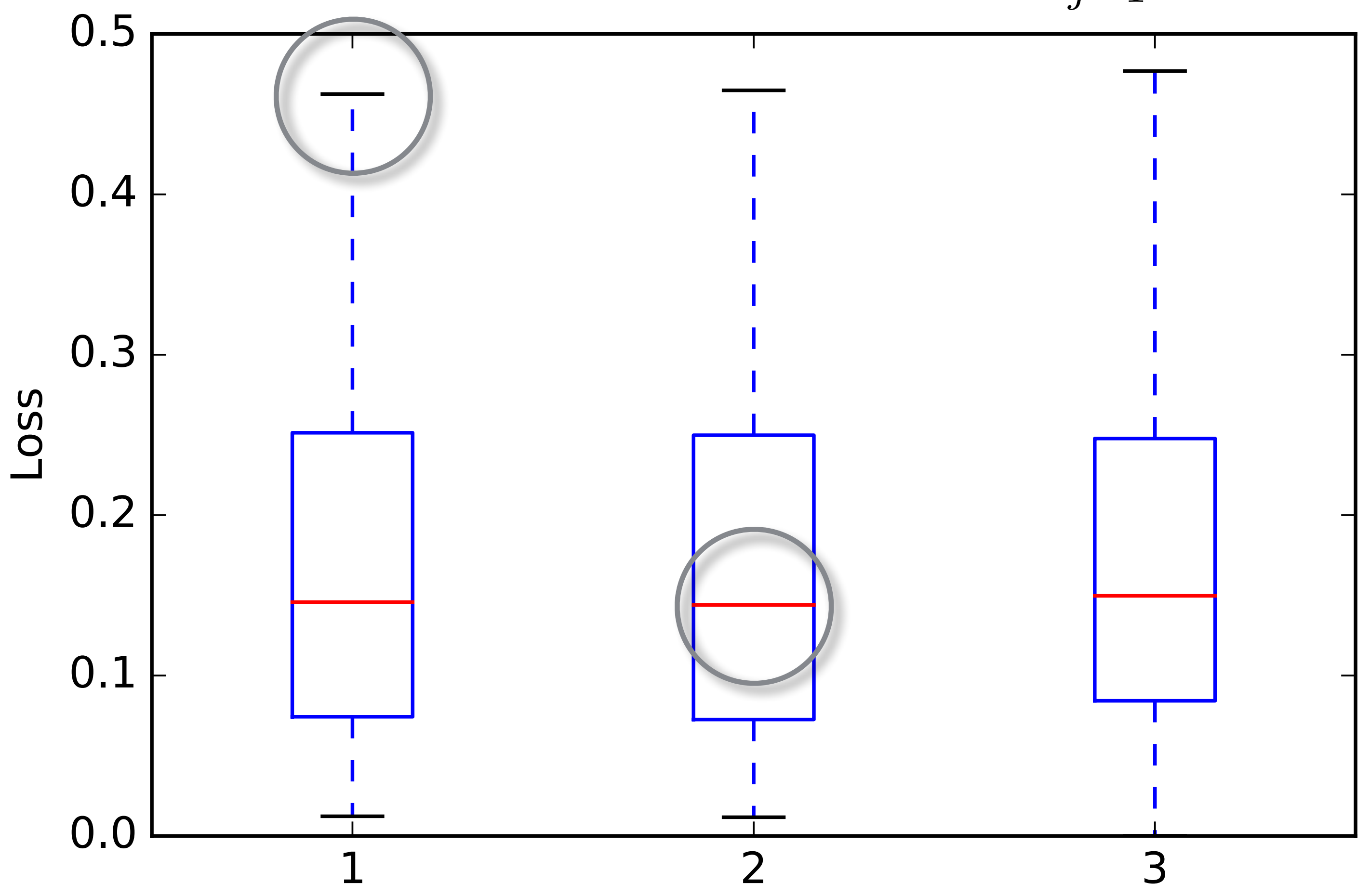
- Experiment Result

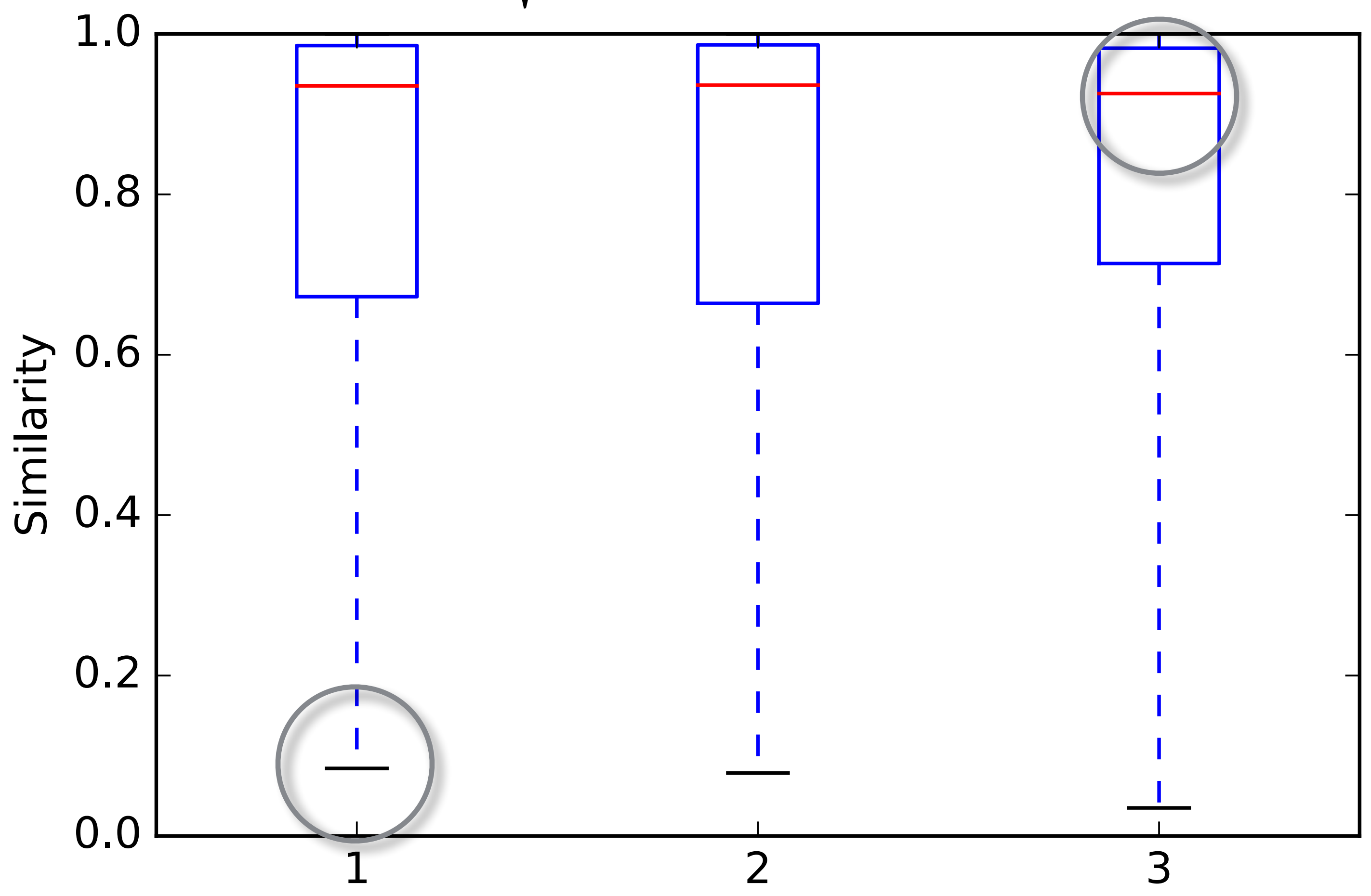
$$\text{loss}(p,\tilde{p}) = \frac{1}{J}\sum_{j=1}^{J}|p_j - \tilde{p}_j|$$

$$\text{sim}(p,\tilde{p}) = \frac{\sum_{j=1}^{J} p_j \times \tilde{p}_j}{\sqrt{\left(\sum_{j=1}^{J} p_j \times p_j\right)\left(\sum_{j=1}^{J} \tilde{p}_j \times \tilde{p}_j\right)}}$$

Run-1: the best errors

Run-2: the best mean loss (0.1676)





Run-3: the best mean similarity (0.8136)

Run	Criteria	Neural Network Parameters				
		L	$n^{(*)}$	$a^{(*)}$	b	N
1	sim-L	2	32, 16	relu, hard sigmoid	256	385
2	loss-L	2	32, 16	relu, hard sigmoid	256	392
3	sim+L	3	64, 32, 16	softplus, hard sigmoid, linear	256	578

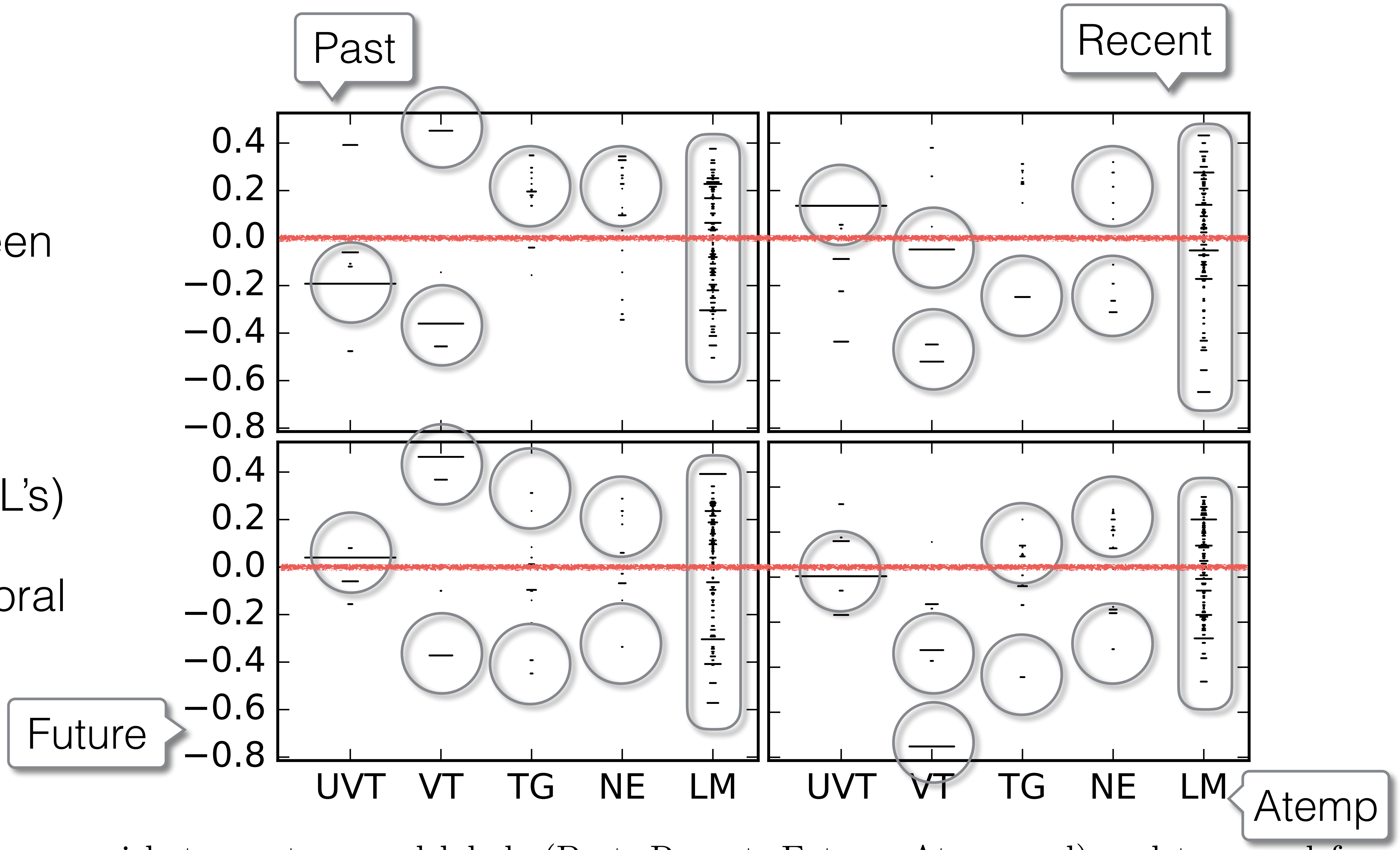
Configurations of L (the number of hidden layers), $n^{(l)}$ (the number of neurons in layer l), $a^{(l)}$ (the activation function from in layer l), b (the batch size), and N (the number of training epochs) in 3 Runs.

Experiments

- Feature Analysis
  - Normalised point-wise mutual information (npmi) for evaluating the association between temporal features and temporal labels

$$\mathrm{npmi}(x_i; y_j) = -\log \frac{p(x_i, y_j)}{p(x_i)p(y_j)} / \log p(x_i, y_j)$$

  - UVT centres around 0 (too many UVT_NULL's)
  - VT diverges from 0 for Past, Future, Atemporal
  - TG diverges from 0 (sensitive feature)
  - NE diverges from 0 (sensitive feature)
  - LM spreads in a wide range (need finer investigation)



npmi between temporal labels (Past, Recent, Future, Atemporal) and temporal features (Uppermost Verb Tense, Verb Tense, Time Gap, Named Entity, LeMma).

Conclusions

- Exploring the temporal information from Wikipedia and resolving the temporal correlation with vectorized context-similarity.
- Deriving the abstract temporal features and generating the probabilistic temporal predictions with a deep neural network.
- Run-2 with fewer layers achieved the best loss score and Run-3 with more layers rendered the highest similarity score.
- Examining the association between temporal features and temporal labels suggested some directions for improvement.

Thank you for your attention!