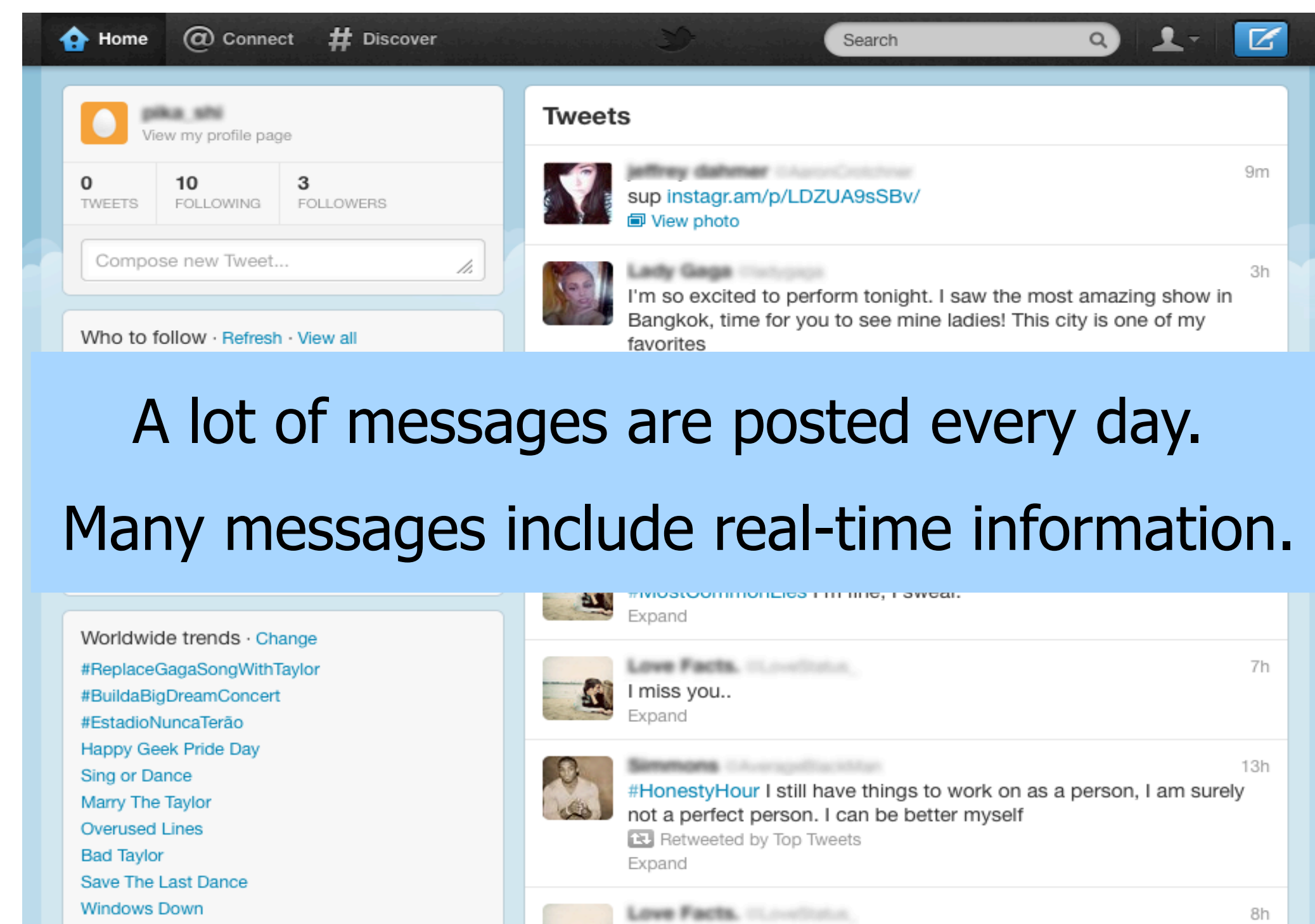
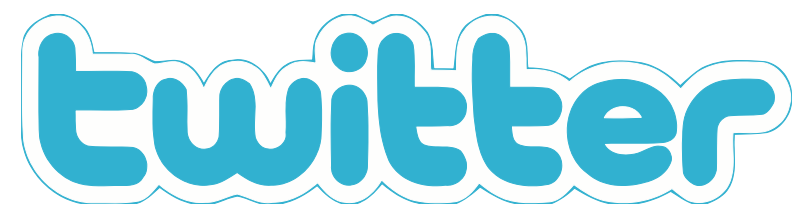


Tweet Classification Based on Their Lifetime Duration

Hikaru Takemura, Keishi Tajima (Kyoto University)

Motivation



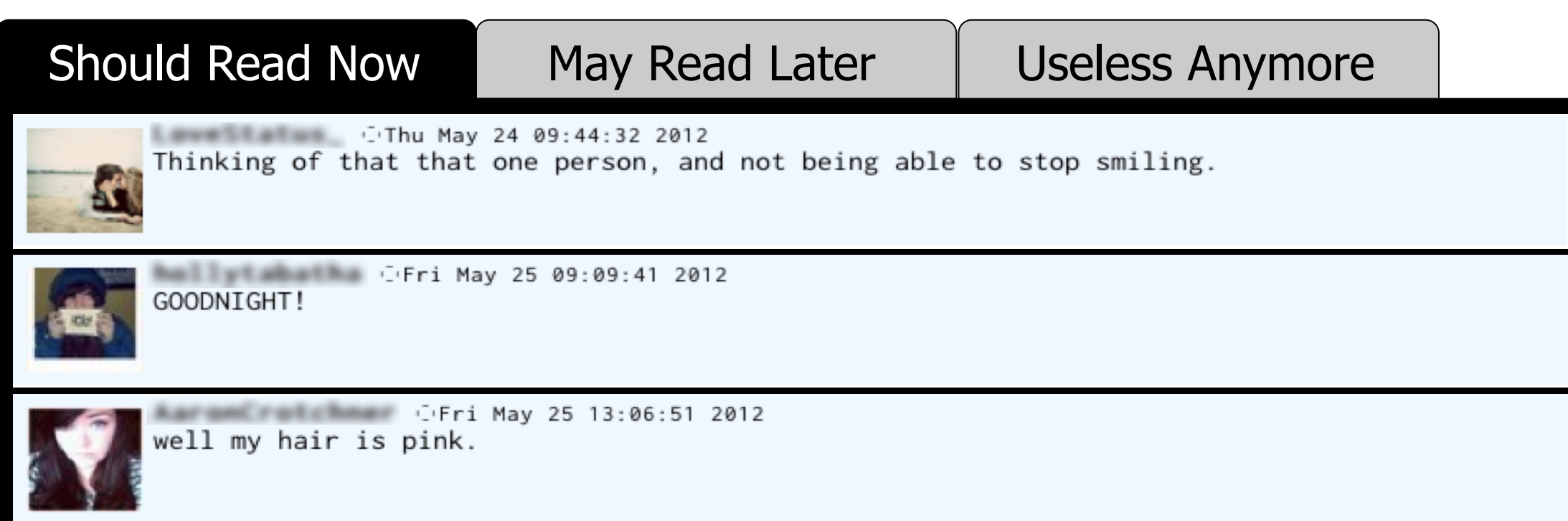
A lot of messages are posted every day.
Many messages include real-time information.

Problem

- Users often find a message including some real-time information only after its value has decreased.
- Some old but useful messages are buried in many old messages whose value has decreased.

Classify into three categories

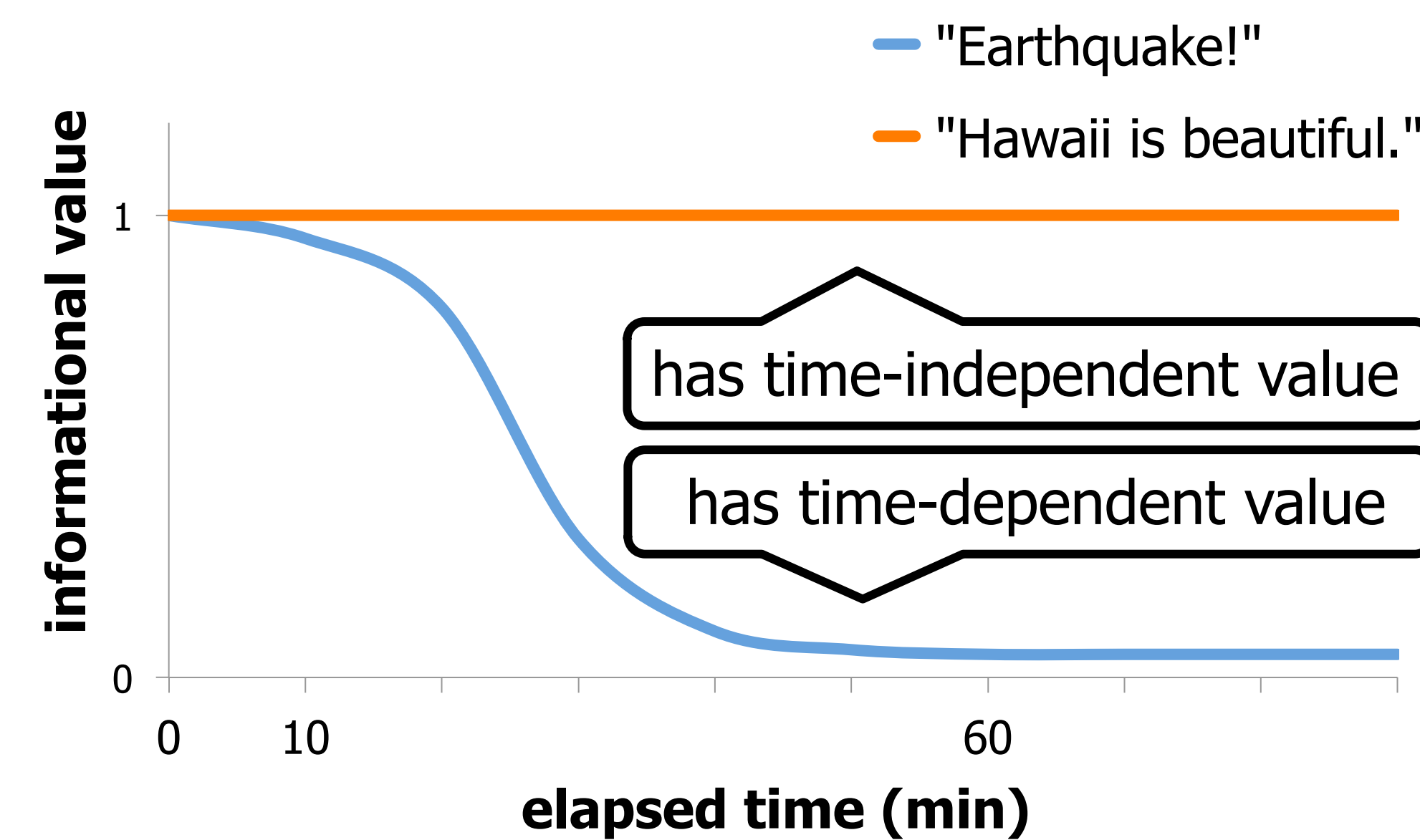
- messages that users should read now
- messages that users may read later
- messages that are not useful anymore



Lifetime Duration

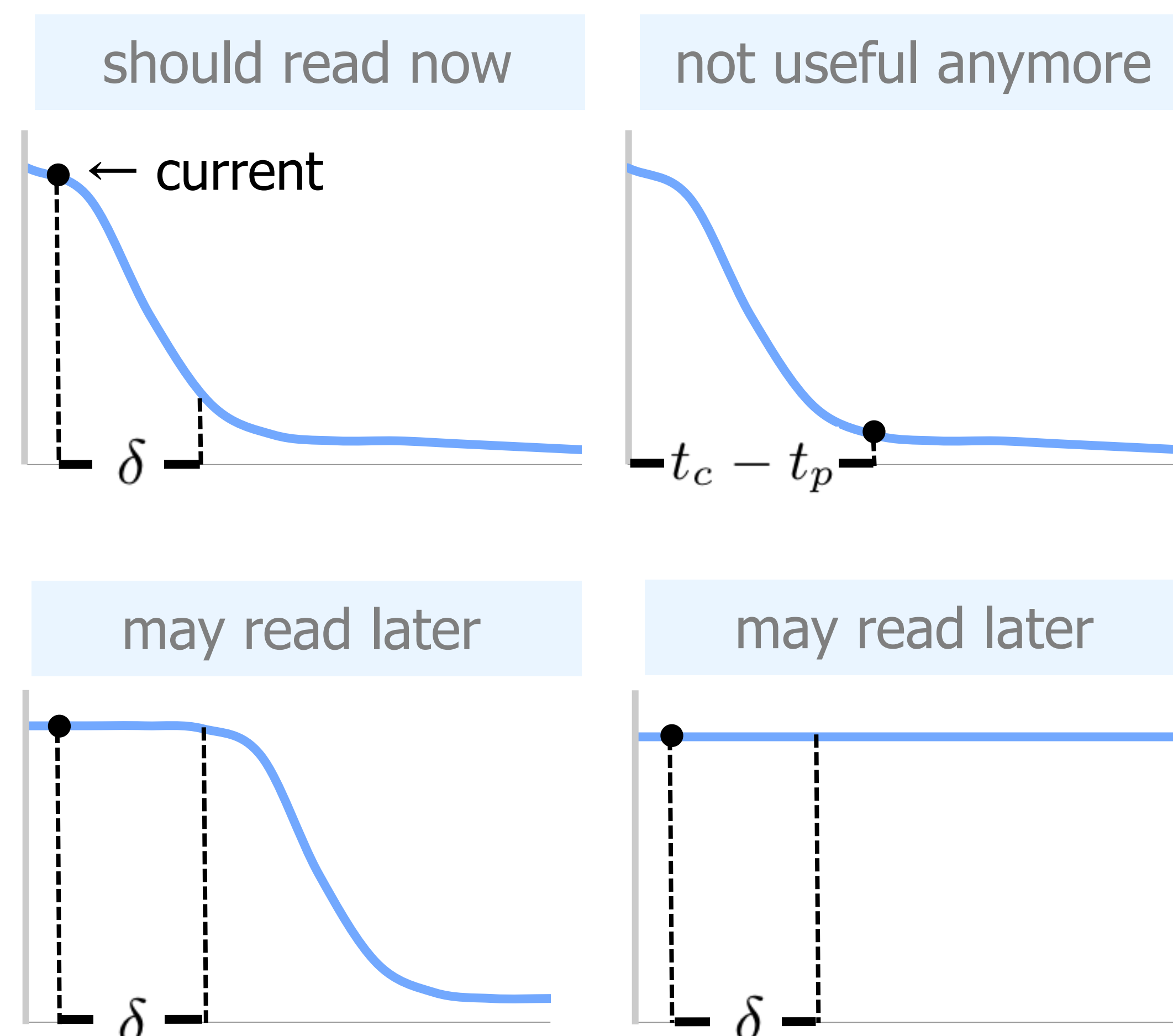
Time-Dependency of Values

- We consider two types of informational value
 - time-dependent value
 - time-independent value



Classification of Messages

t_c : current time t_p : posting time
 δ : the login interval of the user



Approaches

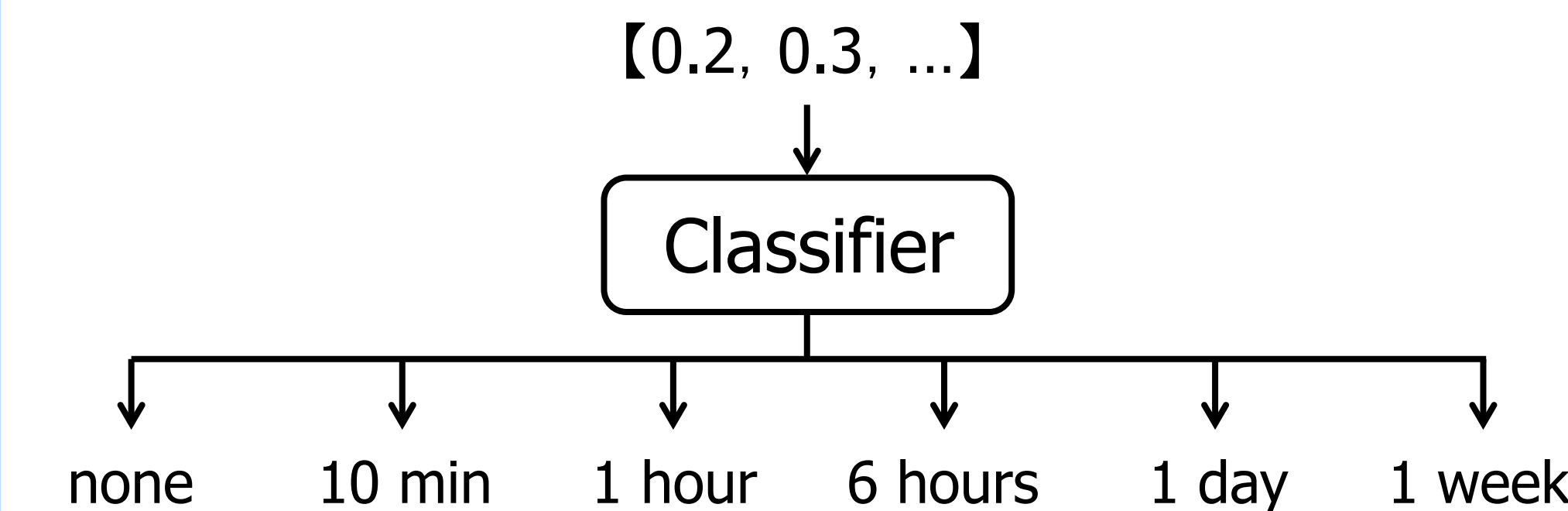
Assumption & Simplification

- Lifetime durations of time-dependent values are fixed to either of :
 - less than 10 minutes or no time-dependent value
 - 10 minutes or longer
 - 1 hour or longer
 - 6 hours or longer
 - 1 day or longer
 - 1 week or longer

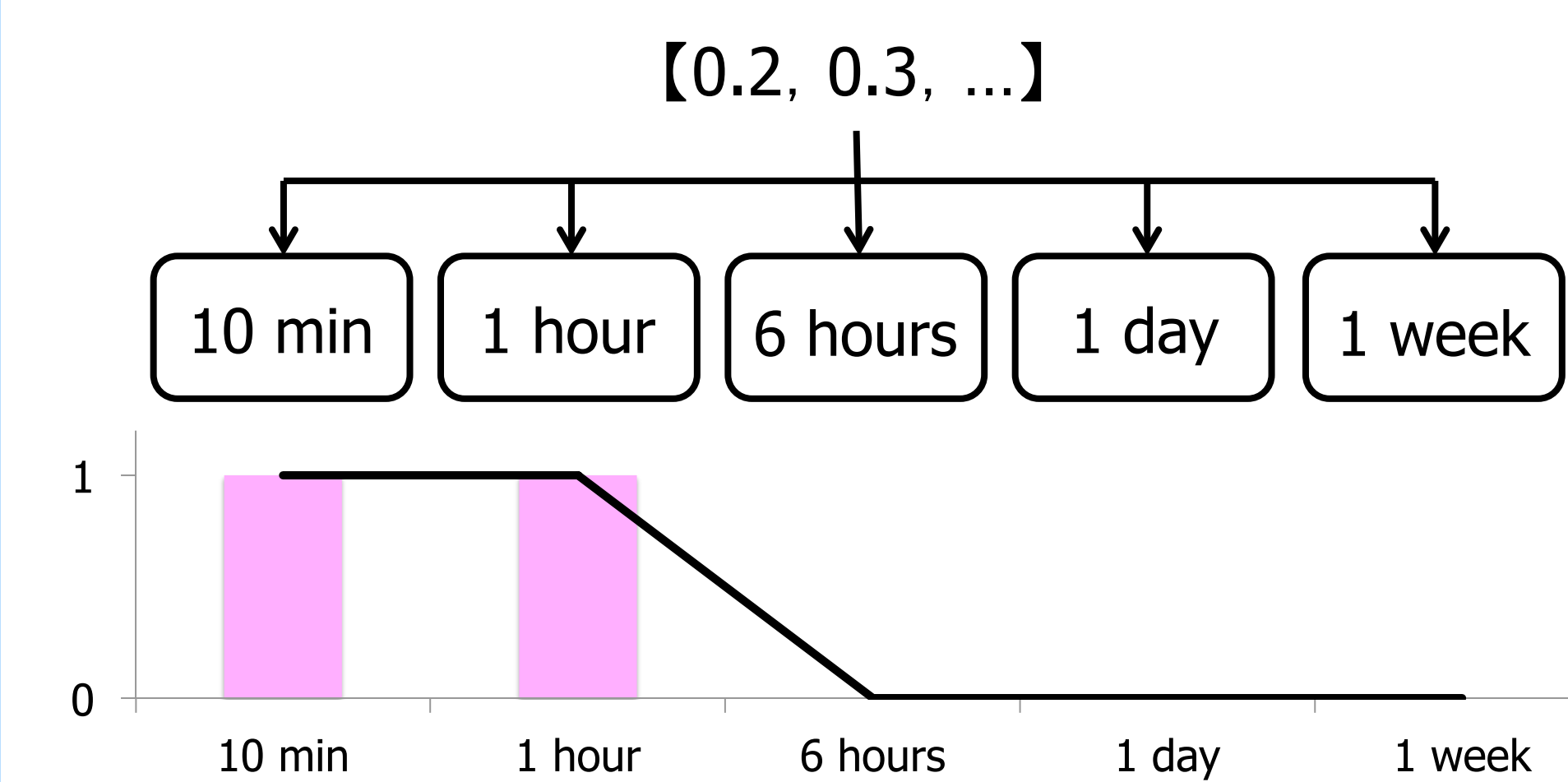
Four Methods

- Determining time-dependent value

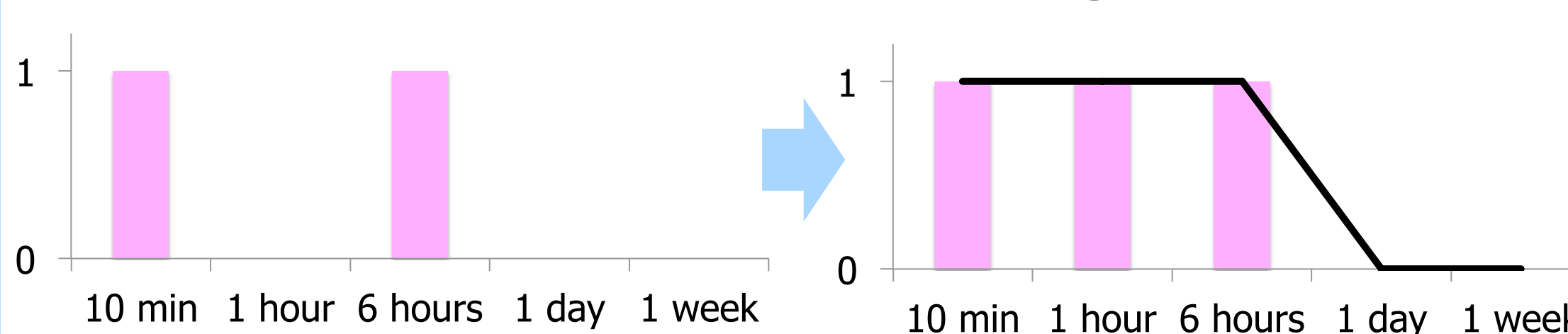
(1) Single Multiclass Classifier



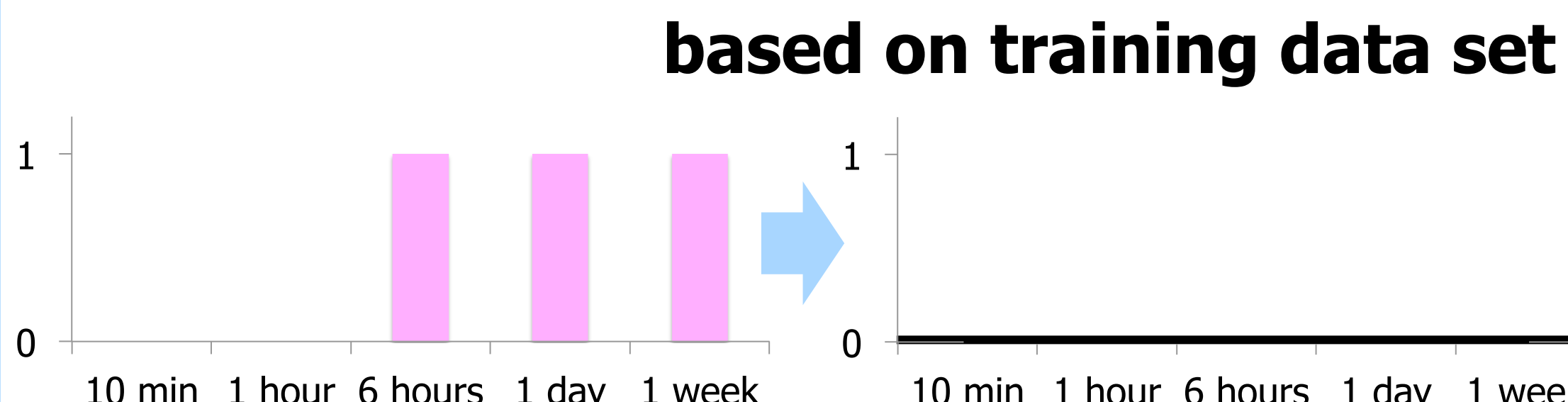
(2) Five Binary Classifiers



(2-a) Monotonically Decreasing Assumption



(2-b) Probabilistic Correction



Features used for the Classifiers

- Number of Characters
- Type of the user who posted the message
- Inclusion of bursty words
- Interval of Reply
- Time expressions in the message
- Inclusion of URL or photos
- Time dependency of each term

Experimental Results

Classifier: SVM and Decision Tree

Accuracy of Determining Lifetime Duration

accuracy(%)	exact	exact or -1	exact or ± 1
(1) multiclass decision tree	34.55	47.82	65.81
(2) five binary SVMs	26.32	38.94	60.74
(2-a) SVMs w/ monotonic	28.54	43.78	65.62
(2-b) SVMs w/ probabilistic	38.59	53.97	74.40

Accuracy of SVMs with Each Feature Removed

removed feature	SVM				
	10 min	1 hour	6 hours	1 day	1 week
with all	55.27	57.49	70.00	74.35	75.45
bursty words	54.79	48.03	72.09	74.37	75.44
URL & photos	55.45	59.26	55.35	58.01	58.81

Accuracy of final classification

into three categories

