

A Probabilistic Model for Personalized Tag Prediction

Dawei Yin, Zhenzhen Xue, Liangjie Hong
and Brian D. Davison
Lehigh University

The Tag Prediction Problem

The screenshot shows the Delicious website interface. At the top, there are navigation tabs for Home, Bookmarks, People, and Tags. Below this is a section for 'Recent apple Bookmarks' with a sub-section for 'Recent | Popular'. A search bar contains the tag 'apple' and a prompt 'Type another tag'. A 'Display options' dropdown is visible. The main content area lists several bookmarks, each with a title, author, date, and a set of tags. The 'apple' tag is highlighted in each set, and a blue box indicates the count of items with that tag. For example, the bookmark 'EyeTV Hybrid INT' has 22 items with the 'apple' tag, and 'Path Finder 5 by Cocoatech' has 1767 items with the 'apple' tag.

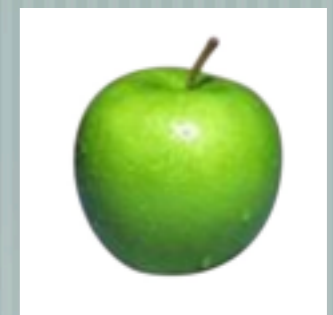
Bookmark Title	Author	Tags	Apple Tag Count
Nelson's Weblog: tech / bad / apple-patent-bully	Jeff Egnaczyk	patent, apple	22
EyeTV Hybrid INT	Ryan S. Yeung	mac, tv, tech, apple	2
Caramel Apple Oatmeal Cookies Recipe - MyRecipes.com	squeakyfromage	oatmeal, cookies, caramel, baking, dessert, lowfat, healthy, cooking, apple	13
How-To: Setup an Airport Extreme in Bridge Mode	delboycarter	networking, how-to, apple	1767
Converting file formats in Mac OS X	wonder_fm	howto, convert, files, mac, music, software, file, conversion, apple	7
Apple sues HTC over iPhone patents	CHARGER_LIFE	iphone, sues, htc, real, cheater, or, cheater, apple	6
Path Finder 5 by Cocoatech	Rogério Sena	mac, software, osx, finder, tools, utilites, explorer, commander, pathfinder, apple	7
The Apple Patents Cockpunching All Smart Phones: An Illustrated Guide ...	atomaton	mobile, ia, digram, mac, apple	6
BlackBerry Tethering via USB on Mac - BlackBerryForums.com : Your Number One BlackBerry Community			6

User

Tag

Item

Time



The Tag Prediction Problem

The screenshot shows the Delicious website interface. At the top, there are navigation tabs for Home, Bookmarks, People, and Tags. Below this is a section titled "Recent apple Bookmarks" with a sub-header "Recent | Popular" and a note "See popular apple bookmarks." A search bar contains the tag "apple" and a prompt "Type another tag". A "Display options" dropdown is visible on the right. The main content area lists several bookmarks, each with a title, author, and a set of tags. The tags are displayed as small grey buttons, and the count of items for each tag is shown in a blue box. The bookmarks listed are:

- Nelson's Weblog: tech / bad / apple-patent-bully** (SAVE) by Jeff Egnaczyk. Tags: patent, apple. Count: 22.
- EyeTV Hybrid INT** (SAVE) by Ryan S. Yeung. Tags: mac, tv, tech, apple. Count: 2.
- Caramel Apple Oatmeal Cookies Recipe - MyRecipes.com** (SAVE) by squeakyfromage. Tags: oatmeal, cookies, caramel, baking, dessert, lowfat, healthy, cooking, apple. Count: 13.
- How-To: Setup an Airport Extreme in Bridge Mode** (SAVE) by delboycarter. Tags: networking, how-to, apple. Count: 2.
- Converting file formats in Mac OS X** (SAVE) by wonder_fm. Tags: howto, convert, files, mac, music, software, file, conversion, apple. Count: 13.
- Apple sues HTC over iPhone patents** (SAVE) by CHARGER_LIFE. Tags: iphone, sues, htc, real, cheater, or, cheater.. Count: 1767.
- Path Finder 5 by Cocoatech** (SAVE) by Rogerio Sena. Tags: mac, software, osx, finder, tools, utilites, explorer, commander, pathfinder, apple. Count: 7.
- The Apple Patents Cockpunching All Smart Phones: An Illustrated Guide ...** (SAVE) by atomaton. Tags: mobile, ia, digram, mac, apple. Count: 6.
- BlackBerry Tethering via USB on Mac - BlackBerryForums.com : Your Number One BlackBerry Community** (SAVE). Count: 6.

User

?

Item

Time



Search for a keyword or phrase...

Have an account? [Sign in](#)

Discover what's happening right now, anywhere in the world

[Wikileaks](#) [Oil Spill](#) [Arcade Fire](#) [TRENDING TOPICS](#) [Valdivia](#) [Gulf](#) [CEO](#) [WikiLeaks](#) [iMac](#) [Edson Celulari](#) [Keon](#)

New to Twitter?



Twitter is a rich source of instant information. Stay updated. Keep others updated. It's a whole thing.



Save a new bookmark

Now add tags and notes

URL Required

TITLE Required

NOTES 1000 characters left

TAGS Space separated, 128 characters per tag

FOR Space separated

Mark as Private

[Save](#)

[Cancel](#)

Tags

[Send](#)

Sort: [Alpha](#) | [Frequency](#)

▼ Popular

[2.0](#) [blog](#) [blogging](#) [blogs](#) [chat](#) [collaboration](#) [communication](#)

Why Tag Prediction?

- [Improve web search

- Query expansion

- [Personalized search

- Capture users' interests

- [Automated web resource classification

- [Improve efficiency of user interface

Related Work

Content-based Methods

- encode users' preferences from textual information
- can predict tags for new users and new items
- Lipczak et al.'s method [ECML PKDD 2009]

Graph-based Methods

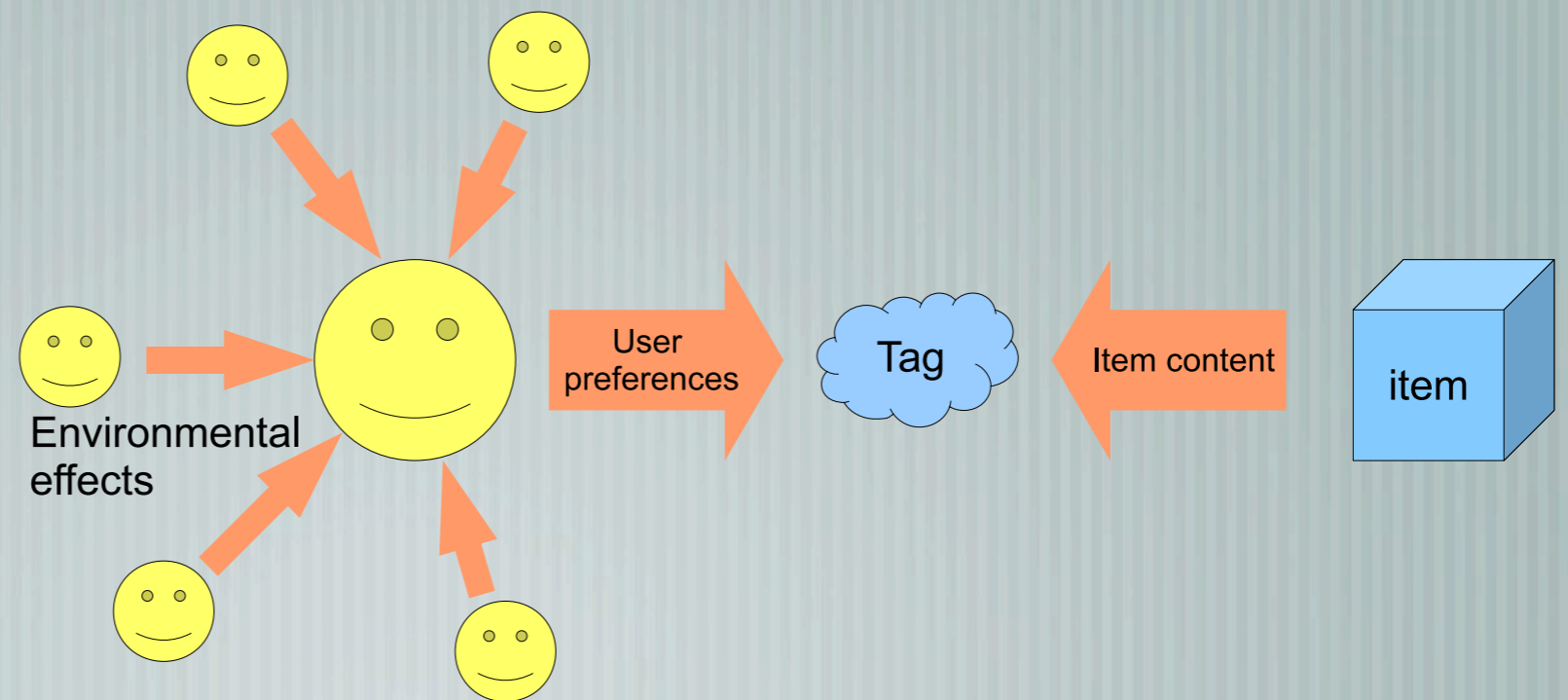
- usually have stronger assumptions than content-based ones
- can provide better performance
- Rendle et al.'s methods [KDD 2009]

Tag Prediction Principles

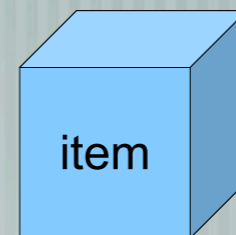
— [The ego-centric effect

— [Environmental effects

— [Item content



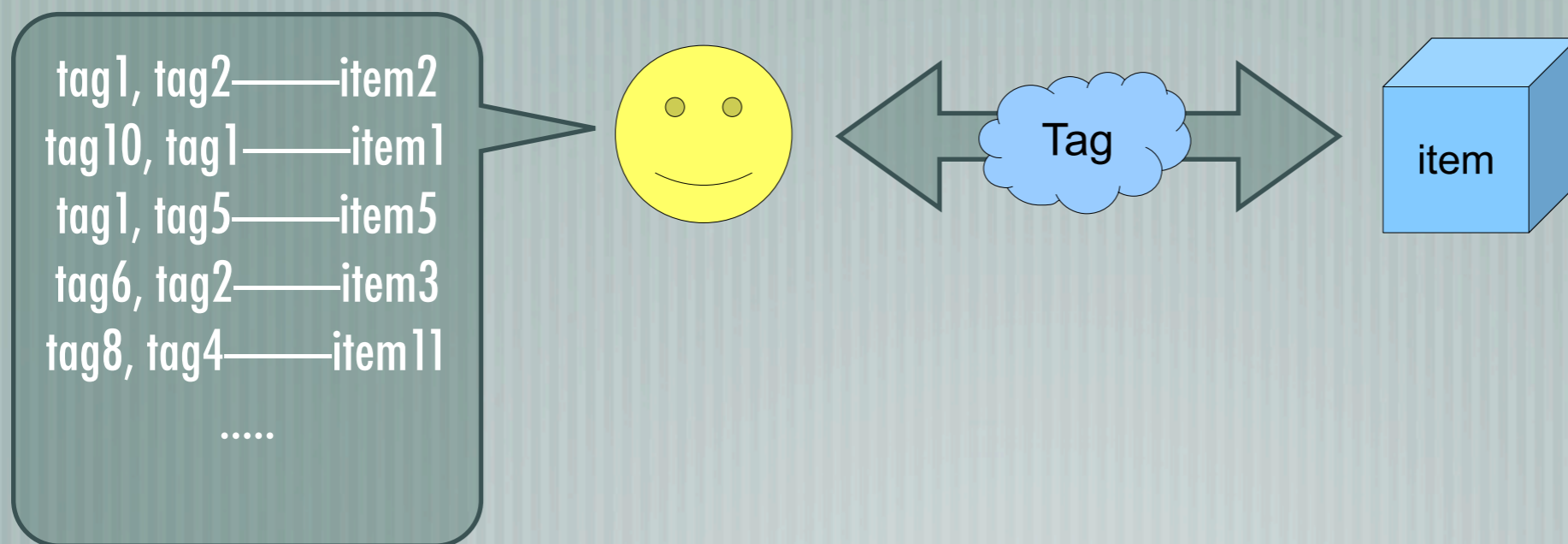
Tag Prediction Principles



Tag Prediction Principles

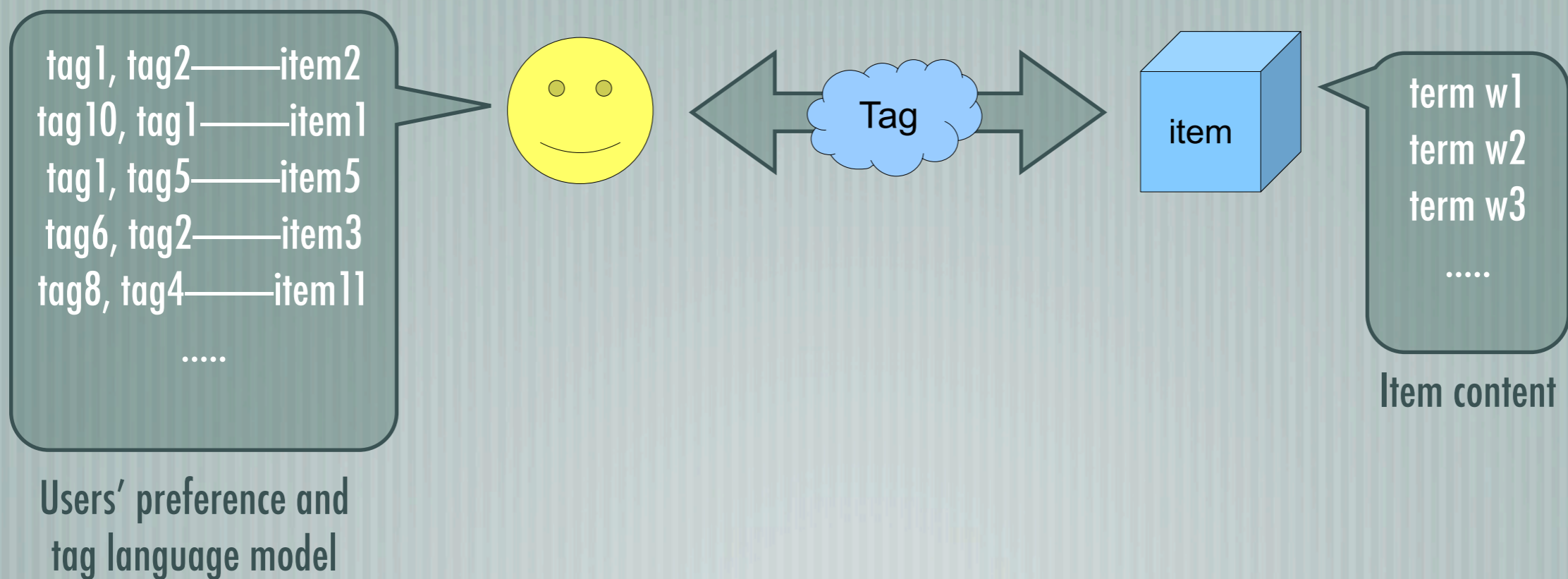


Tag Prediction Principles

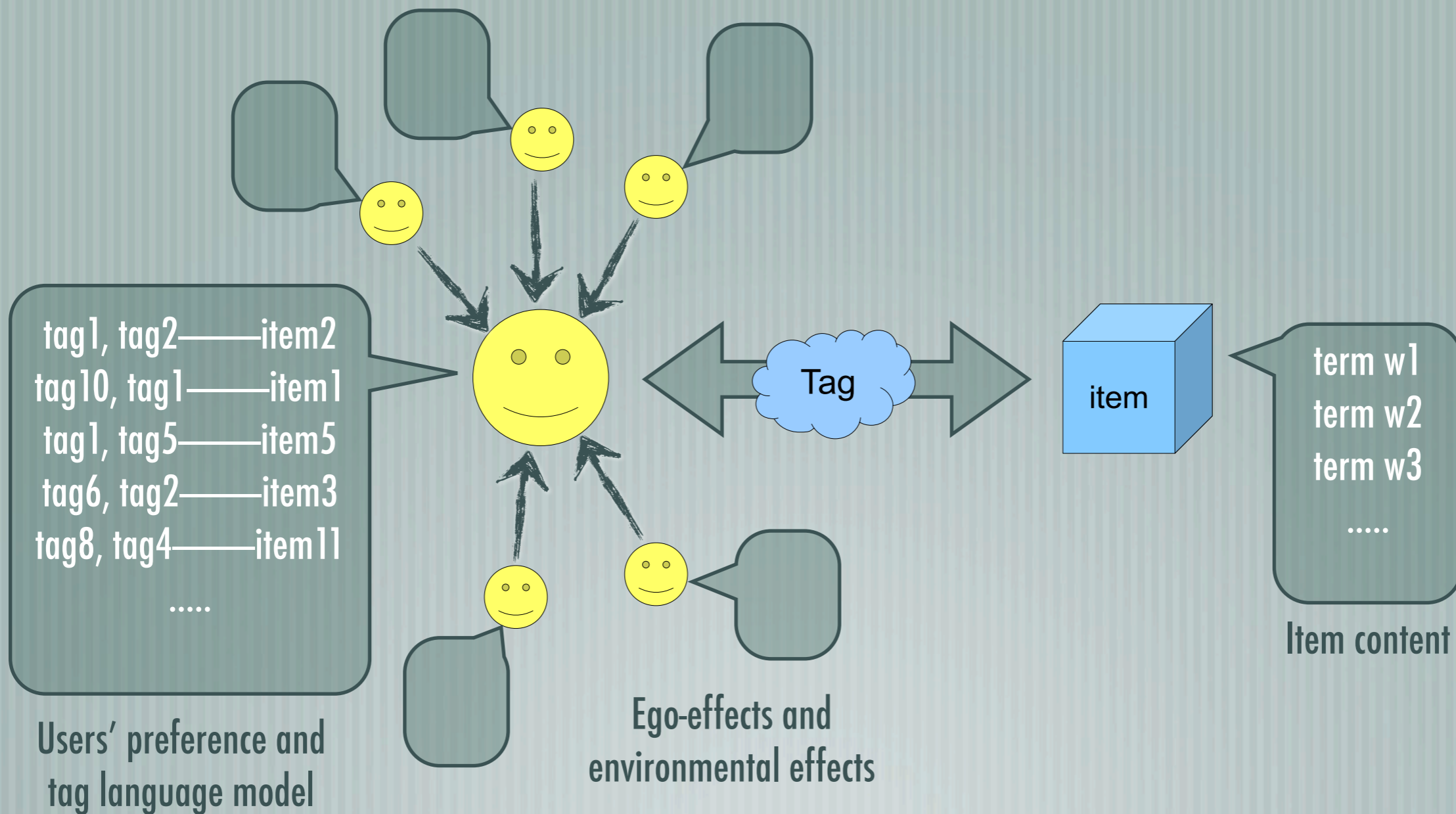


Users' preference and
tag language model

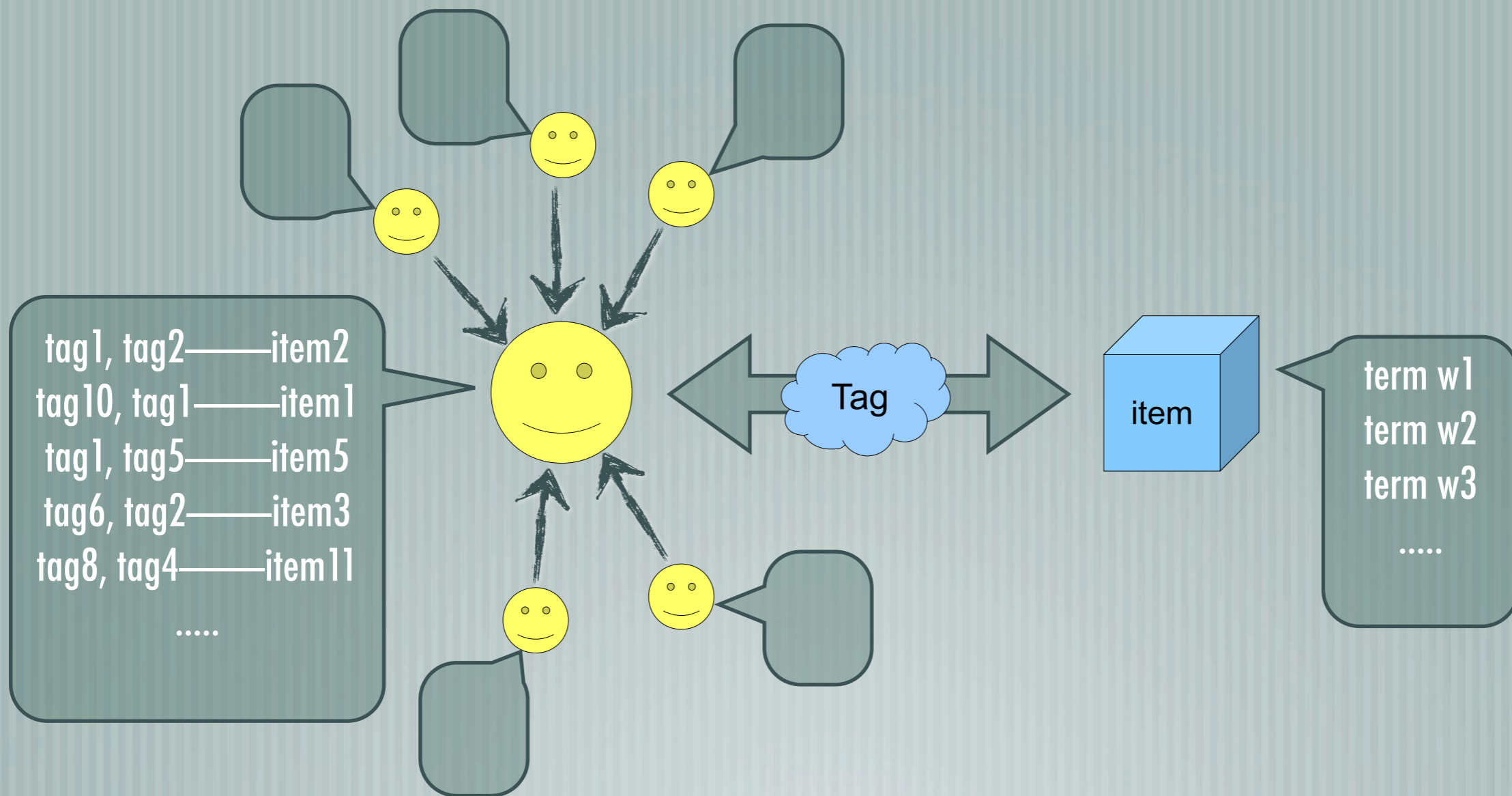
Tag Prediction Principles



Tag Prediction Principles



Tag Prediction Principles



Prediction Model

$$P(t|i, u) \propto \prod_{w \in W_i} \sum_{u_k} P(w|t, u_k) P(u_k|u) \times \sum_{u_k} P(t|u_k) P(u_k|u)$$

Data Set

— [The bookmark dataset of the ECML PKDD 09 Challenge Workshop

— [2,679 users, 263,004 items, 56,424 tags, 262,336 posts and 1,401,104 records.

— [We randomly chose 668 posts as test data

Traditional Evaluation

Randomly Choose Test Posts



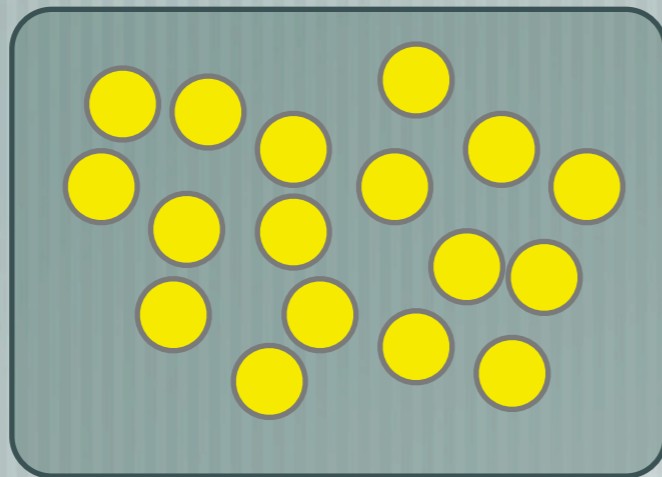
Traditional Evaluation

Randomly Choose Test Posts

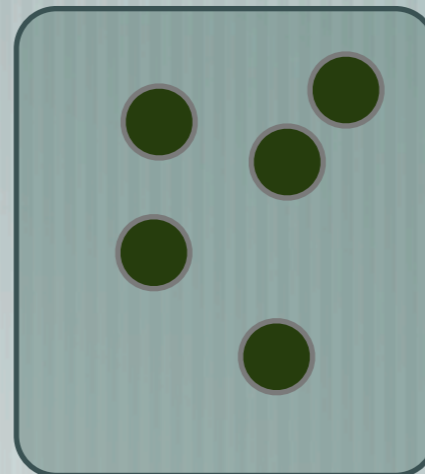


Traditional Evaluation

Randomly Choose Test Posts



Training posts



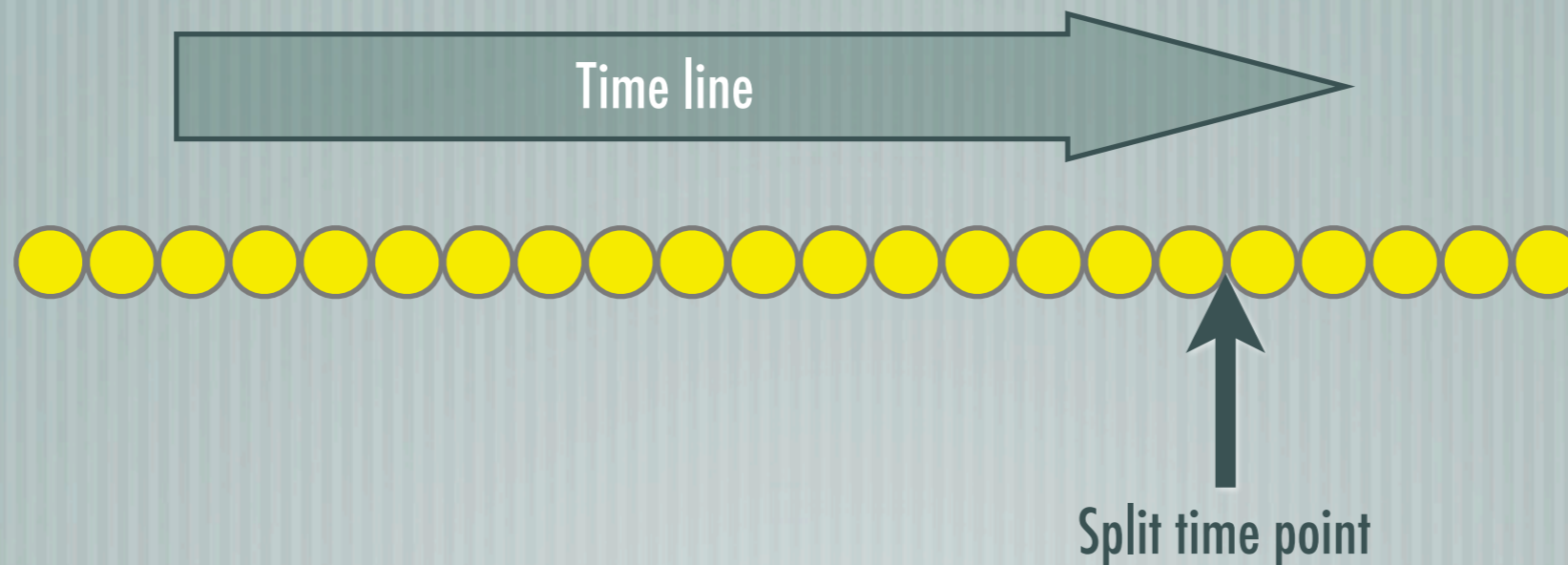
Test posts

Traditional Evaluation

Temporally Generate Test Posts

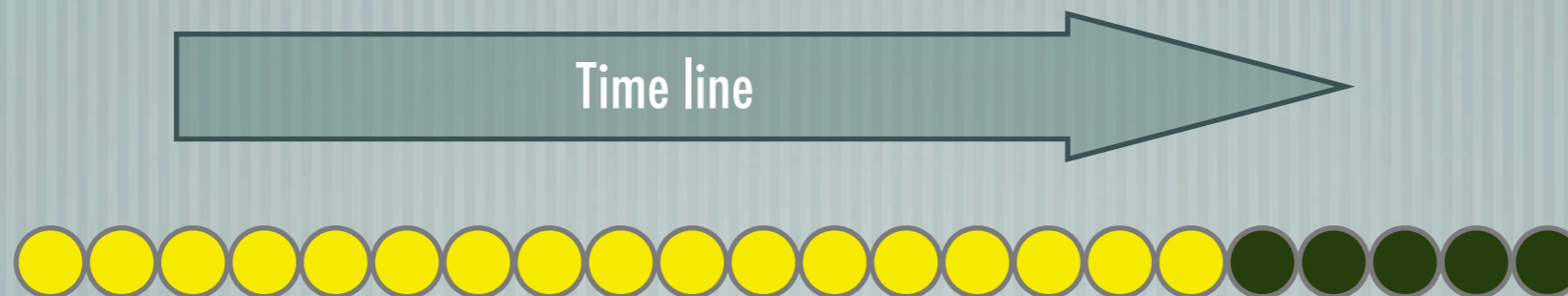
Traditional Evaluation

Temporally Generate Test Posts



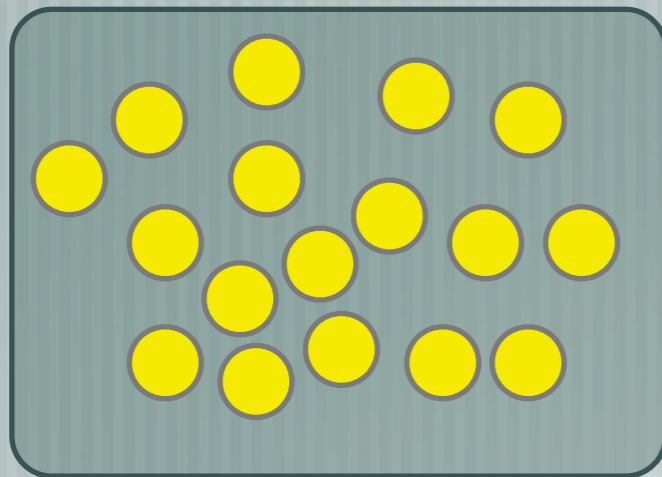
Traditional Evaluation

Temporally Generate Test Posts

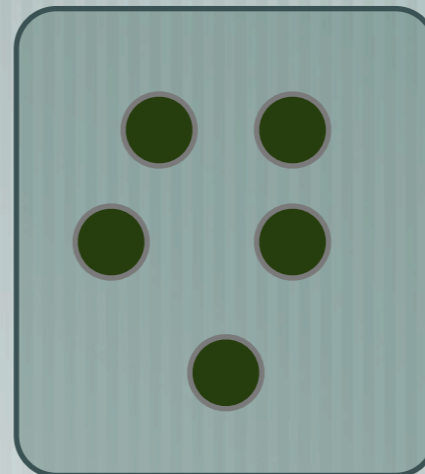


Traditional Evaluation

Temporally Generate Test Posts



Training posts



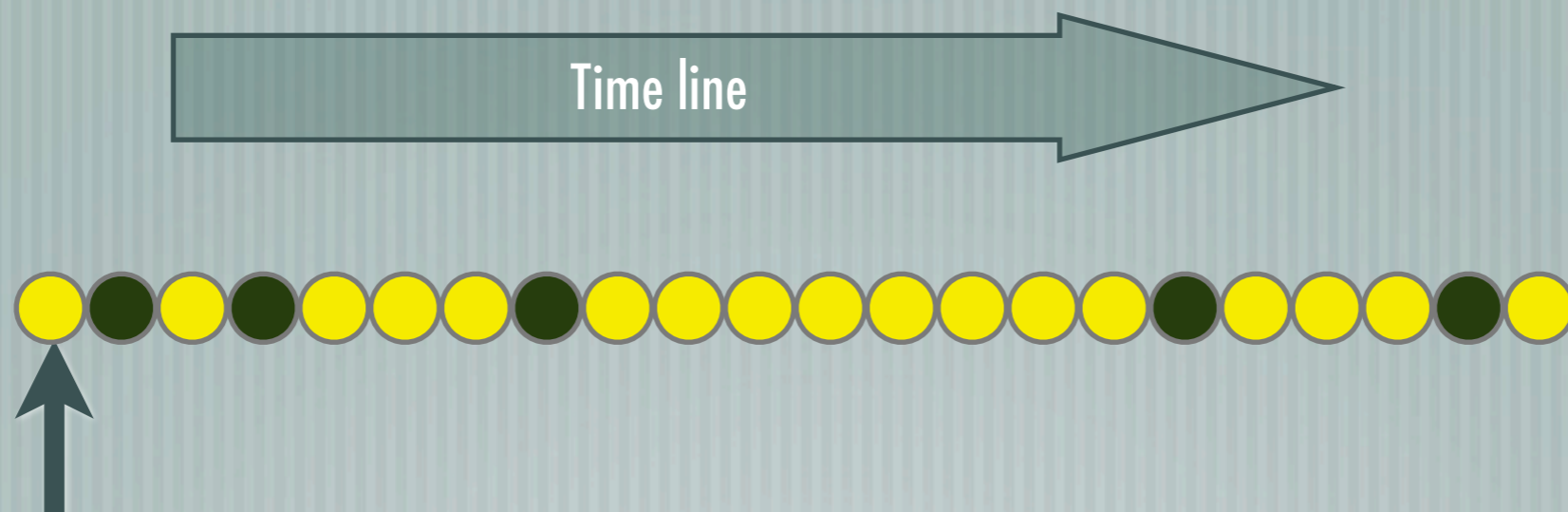
Test posts

Online Evaluation

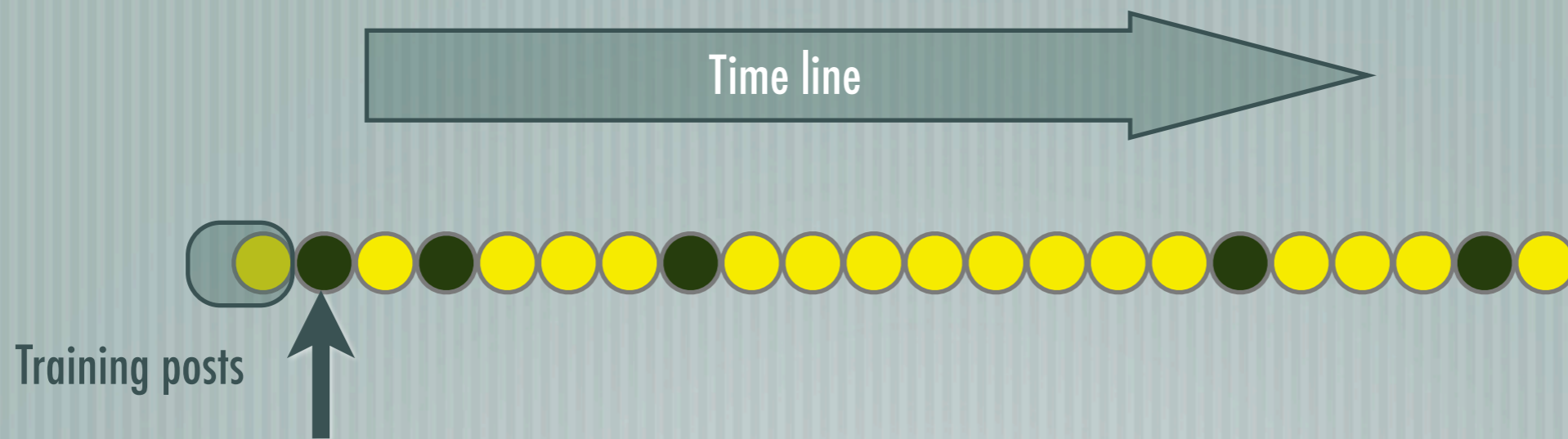
Online Evaluation



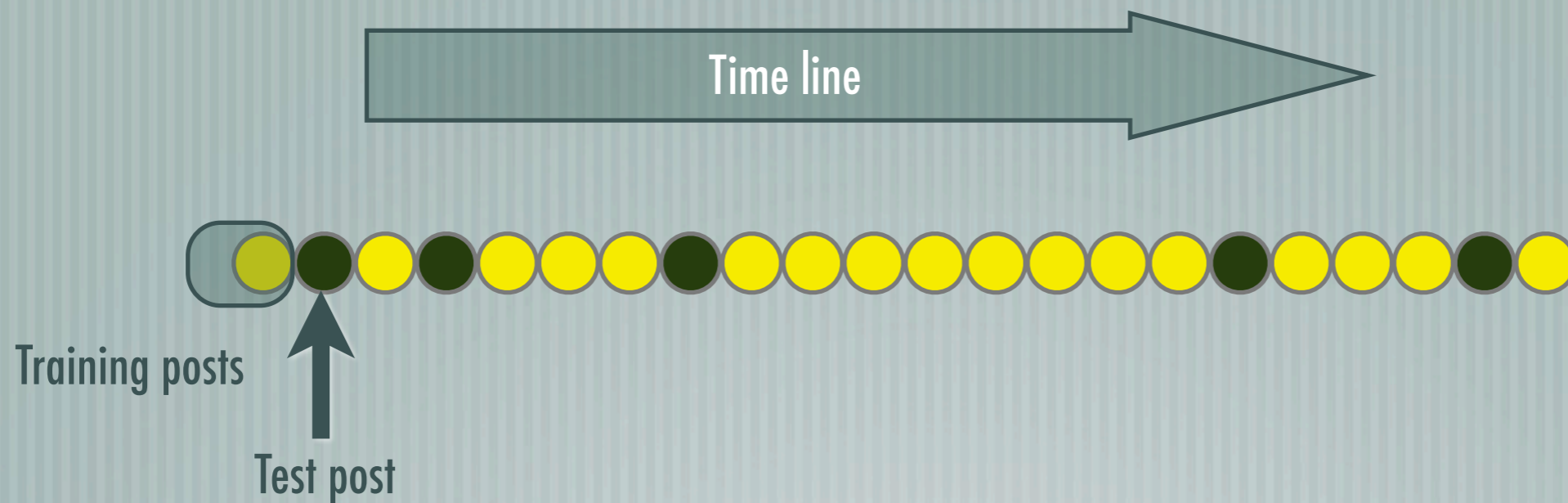
Online Evaluation



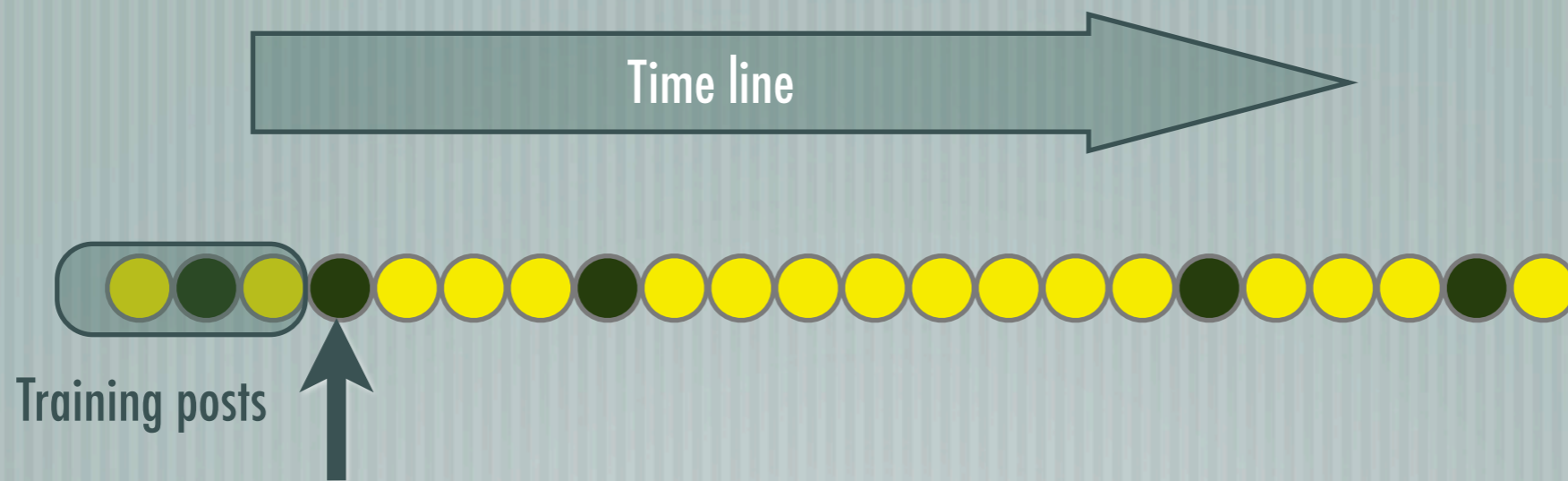
Online Evaluation



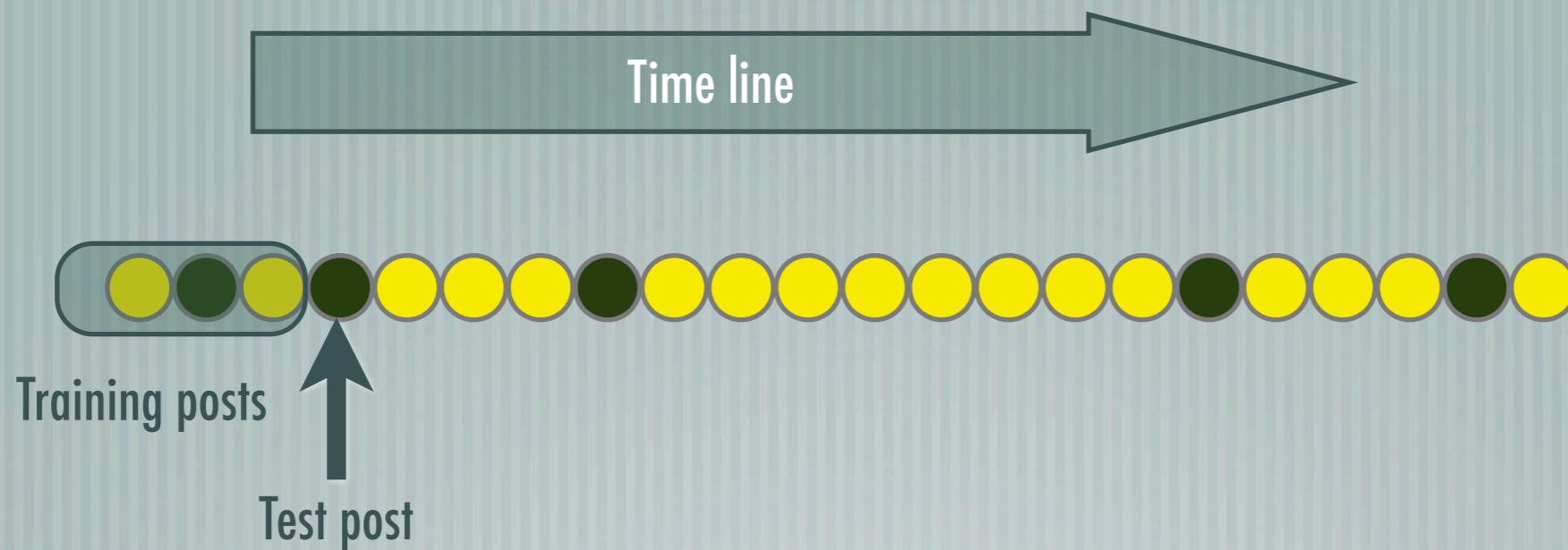
Online Evaluation



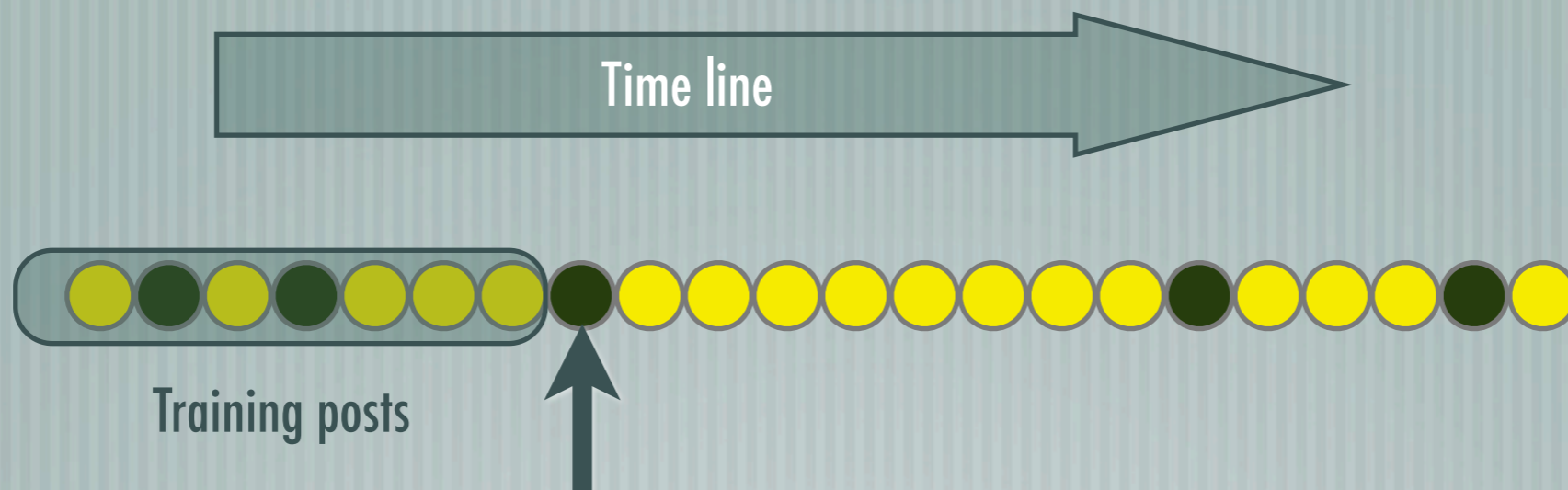
Online Evaluation



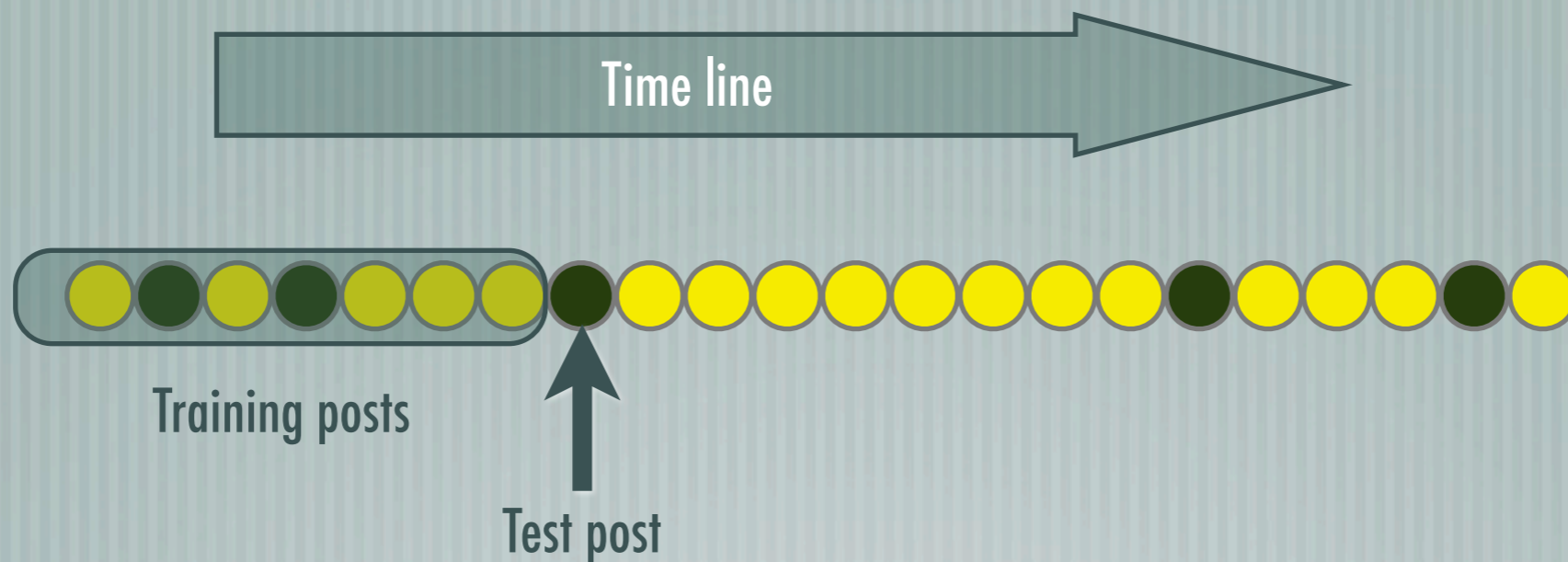
Online Evaluation



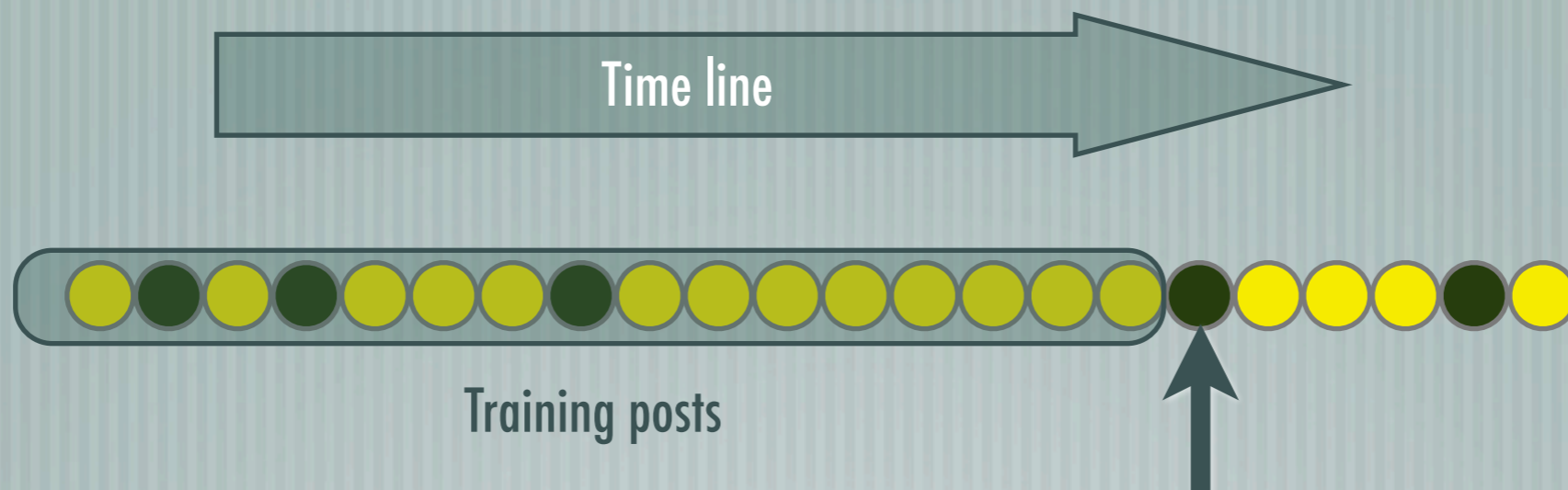
Online Evaluation



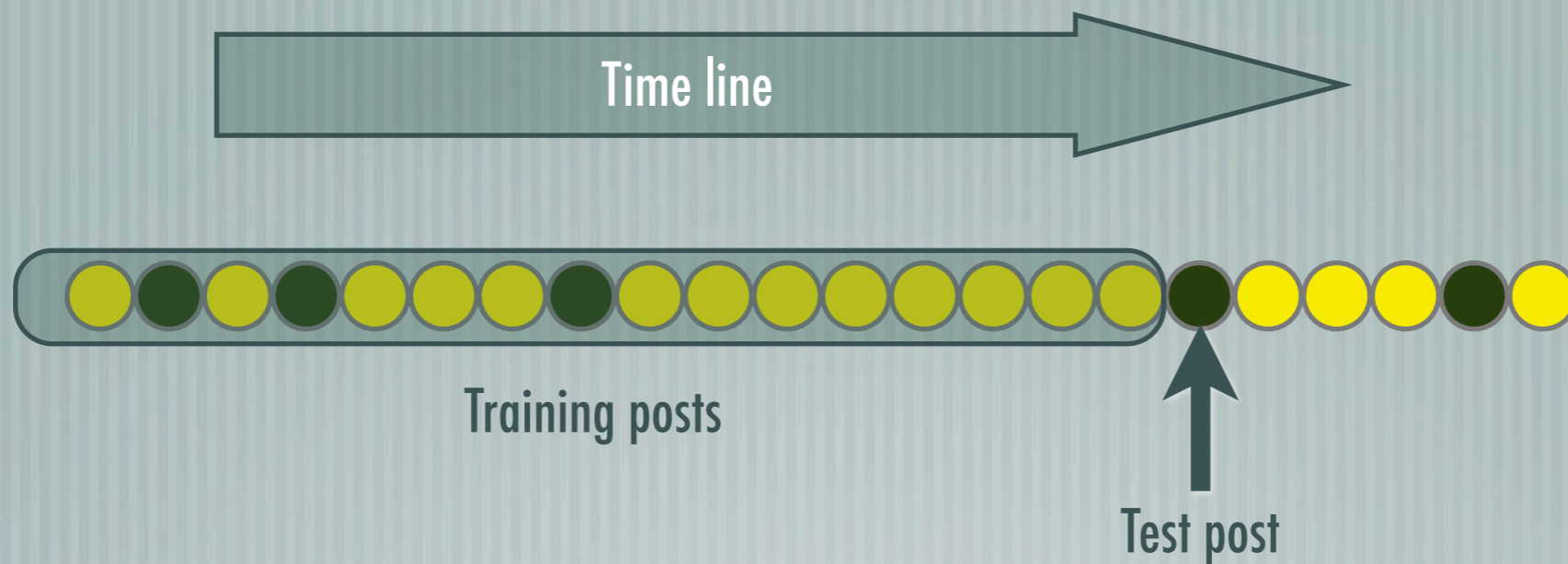
Online Evaluation



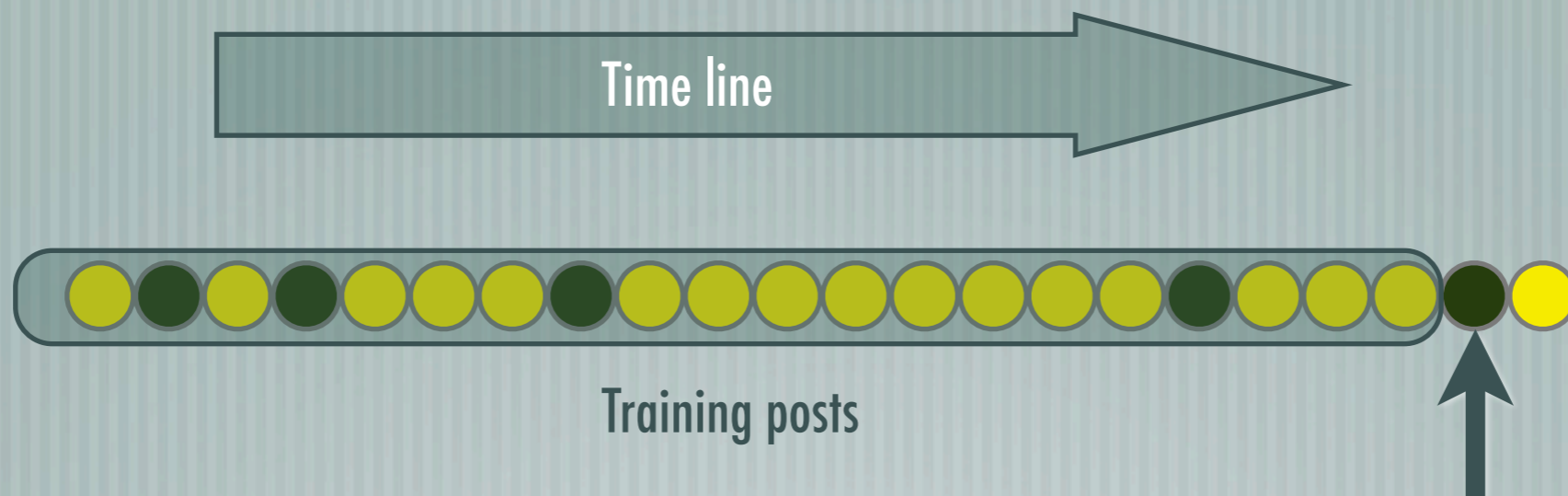
Online Evaluation



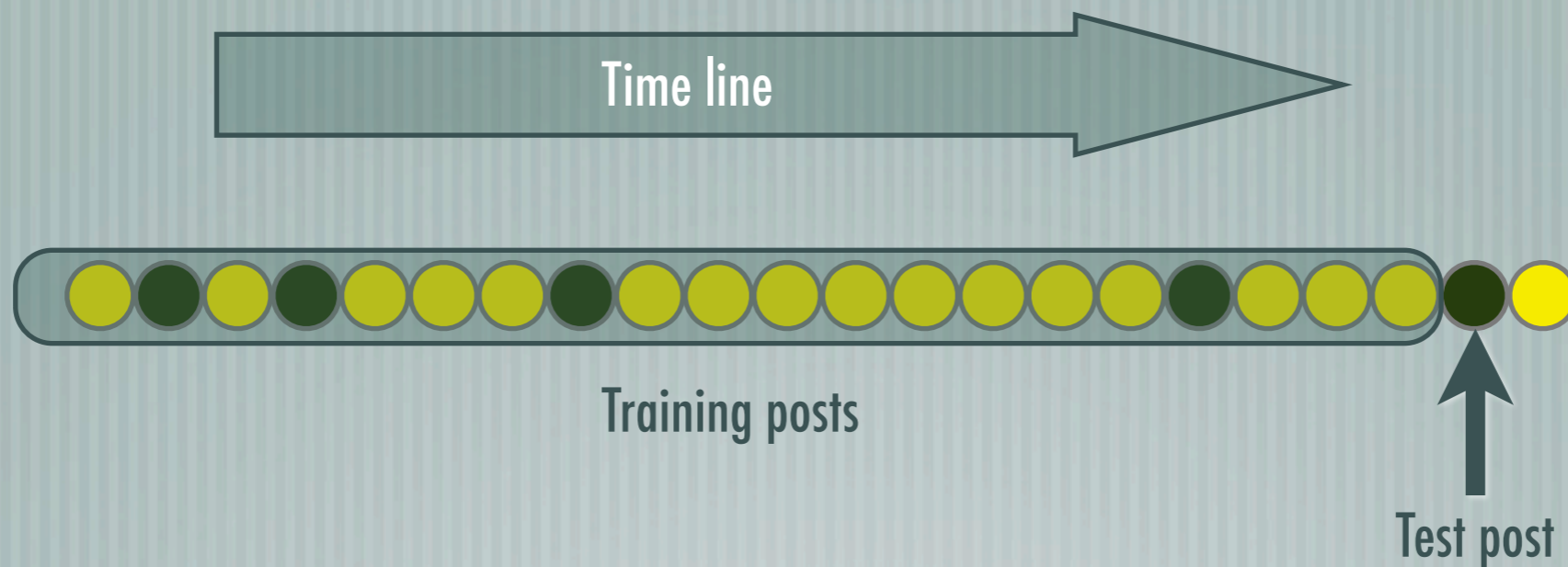
Online Evaluation



Online Evaluation



Online Evaluation



Time Sensitive Sampling

From 262,336 posts, randomly choose 668 posts, total 2307 records

Statistics under online evaluation mode

	Total	New	Old
Users	668	41	627
Items	668	602	66
Tags	2,307	321	1,986

Time Sensitive Sampling

From 262,336 posts, randomly choose 668 posts, total 2307 records

Statistics under online evaluation mode

In the real world, the problem is dominated by the need to predict tags for existing users when they tag new items.

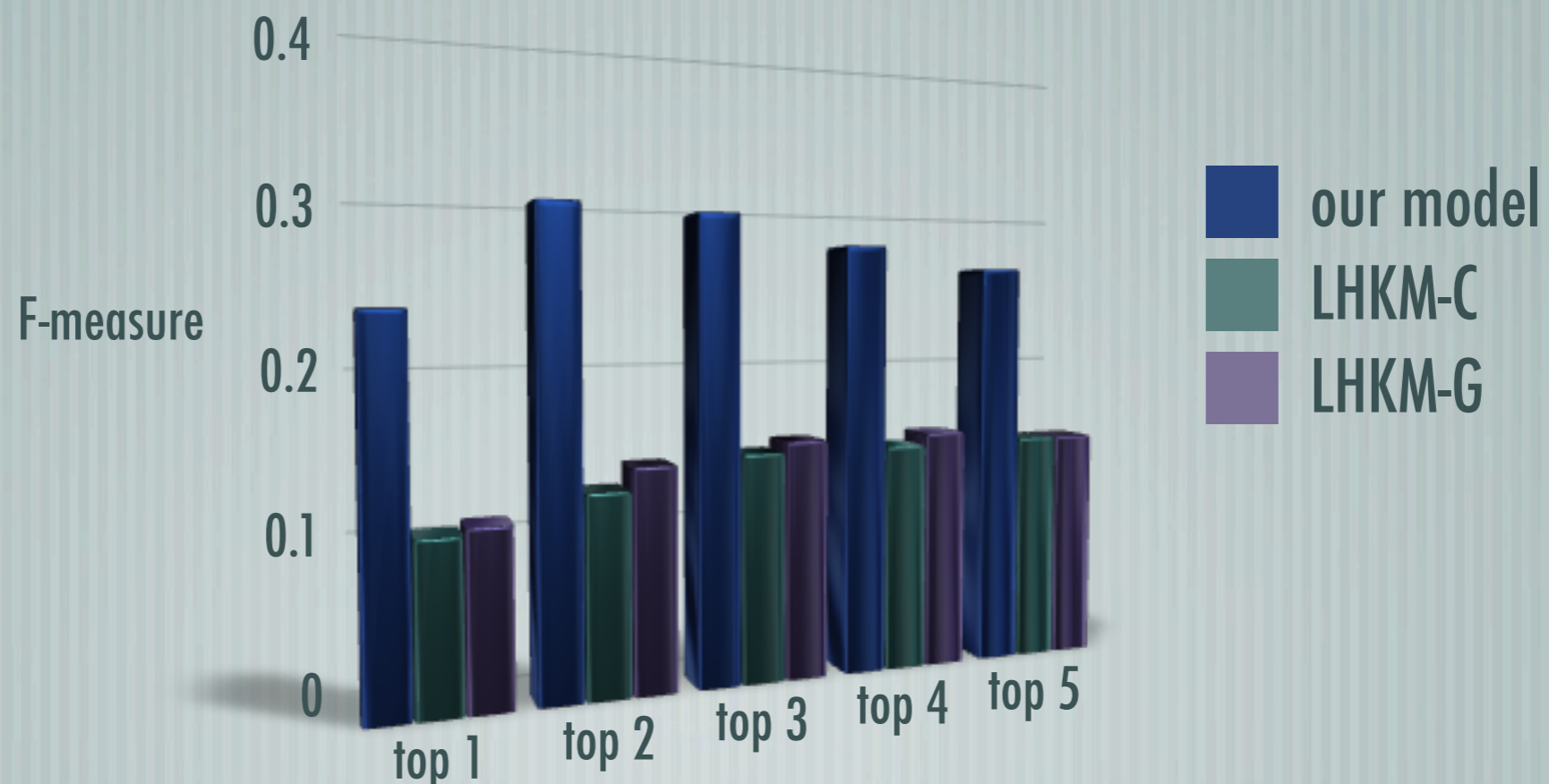
	Total	New	Old	
Users	668	41	627	6.1% new users
Items	668	602	66	90.1% new items
Tags	2,307	321	1,986	13.9% new tags

Experimental Results

— [Compare with Lipczak et al.'s methods

— Winner of content-based method ECML PKDD Challenge Workshop

— LHKM-C Content-based method and LHKM-G Graph-based method



Conclusions

— [Time-sensitive sampling shows that

— predict tags for existing users when they tag new items.

— [Online evaluation mode,

— more realistic

— [Our method,

— can improve F-measure by over 30%

— find that ego-centric effect is often high

Thank you!

Our poster will be in

Poster Session II & Demo Session

Date: Tuesday, July 27, 2010

Time: 5:45pm - 8:00pm

Location: Independence Center B, floor 1