Construction of Evaluation Datasets for Trend Forecasting Studies

ICWSM 2023. Dataset track. Shogo Matsuno[2], Sakae Mizuki[1], and Takeshi Sakaki[1].
[1]Hotto Link, inc. {s.mizuki, t.sakaki}@hottolink.co.jp [2]Gunma University. s.matsuno@gunma-u.ac.jp

hottolink 群馬大学 GUNMA UNIVERSITY Socialgist 2nd author

We proposed construction method and created the trend forecasting task evaluation dataset. Scan QR code to access.

- Background: Trend forecasting studies lack a consistent, systematic evaluation protocol and dataset.
- Challenges: (non-)trending entity coverage, trend period annotation, domain coverage, and questionable public recognition proxy.
- Solution: Sample Wikipedia articles, conduct a public survey, and annotate **trending status**, **degree of trending**, **and trend period**.
- Advantages: Mitigates bias towards popular domains or non-trending entities, efficiently measures public recognition time-series.
- Contribution: Our *Trend Dataset* improves gold standard fidelity, expands entity coverage, and enables standardized evaluation.

Trend Dataset overview

Item	Value
# Entities	400
# trending	80
# non-trending	320
# Categories	21
Survey country	Japan
Survey period	2015~2019
# Respondents	510

Field	Example
Entity name	Splatoon 2/スプラトゥーン2
Domain	Art/Content-Game
Interest pattern	positive
Trending status	Trending
Degree of trending	0.546
Trend period	2017-07-23~2018-11-18

Evaluation tasks using the dataset

- Predict trending or not.
  E.g., Product popularity 🏬
  📏 F-value, accuracy

Pred/GT	trend	non-trend
trend	67	64
non-trend	13	256

- Learning to rank by degree.
  E.g., Best-selling TV games 🎮
  📏 MRR, nDCG

Rank	Prediction	Ground-truth
1st	Nier: Automata	Splatoon 2
2nd	Dragon Quest XI	Nier: Automata
3rd	Splatoon 2	Dragon Quest XI

- Predict trend period.
  E.g., Longevity of blockbusters 🎬
  📏 F-value

Entity	The Greatest Showman
Pred	2018-02-05~2018-06-01
GT	2017-12-24~2018-08-12

Sampling entities(=articles) from Wikipedia

🤔 How to avoid skew towards popular domains or non-trendings?
💡 Balanced sampling by incorporating <u>domain</u> and <u>interest patterns</u>.

1. <u>Domain</u> assignment by selecting closest category match.

2. Interest score calculation using article text and metadata.

$s = w_1x_1 + w_2x_2 + w_3x_3$

x_1: Number of pageview (log)
x_2: Number of editors (log)
x_3: Lexical patterns in text suggesting trends
- Quantify using Wikipedia2Vec[Yamada+, 16]

3. <u>Interest pattern</u> assignment using score and manual labeling.

Art/Content-Game

s rank	Entity
1st	Fortnite
2nd	Splatoon 2
3rd	Call of Duty
…	…
235th	Bravely Second
…	…
577th	VIPER –RSR-
578th	The Witch's House

High plausibility of trending
Top-k → positive: Fortnite, Splatoon 2, …

High plausibility of non-trending
Random → negative: VIPER –RSR-, Bravely Second, …

Long-standing popularity w/o recent trendiness
Mark → negative-popular: Pac-Man, Space Invaders, …

Measuring public recognition rates

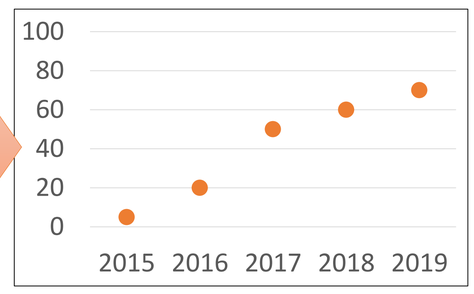
🤔 How to measure public recognition time-series efficiently?
💡 <u>Retrospective survey</u> and <u>interpolation</u> using internet search.

1. <u>Conduct survey</u>, measure the yearly public recognition rates.
   - Calculate each year-end values by accumulating choices.
   - Verification questions to filter out unreliable crowd workers.

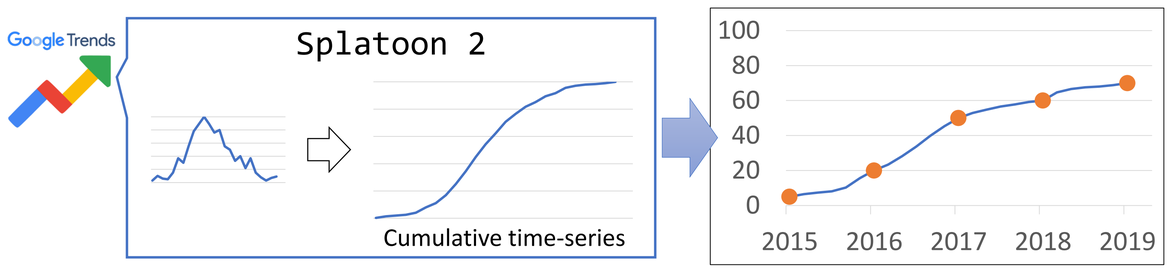
Q. *When* did you learn about Splatoon 2?

x 510

Choices	Ratio	Cum.
I knew since before 2015.	5%	5%
I knew since 2016.	15%	20%
…	…	…
I knew since 2019.	10%	70%
I don't know.	30%	100%

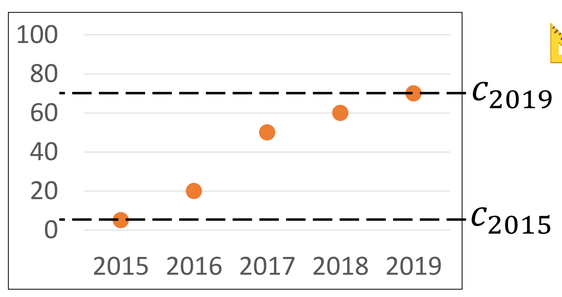


2. <u>Interpolate</u> weekly values using internet search volume.

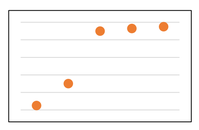


Annotating trend attributes

- **Trending status**: trending OR non-trending



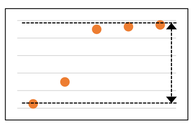
📏 $c_{2015} < 30 \wedge (c_{2019} - c_{2015}) \geq 25$

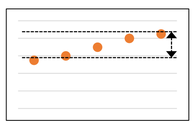

trending

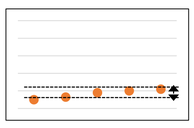

non-trending

- **Degree of trending**: 0~1

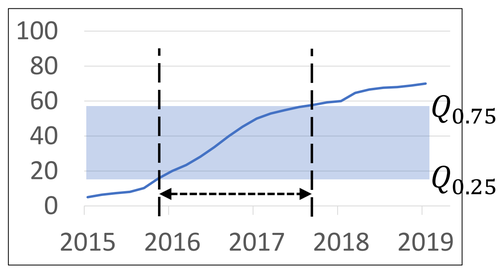
📏 $(c_{2019} - c_{2015}) \div 100$


0.90

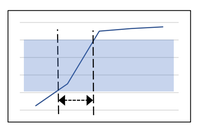

0.31

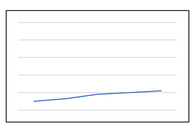

0.12

- **Trend period**: weekly time frame, limited to trending ones.



📏 Since 25%ile until 75%ile


2015-11-22 ~ 2016-08-21

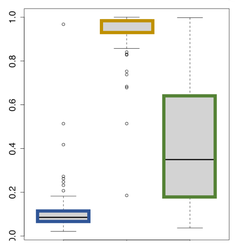

N/A

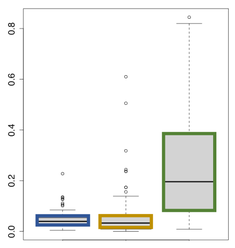
Analysis

➢ Interest patterns aid efficient trending entity collection.

Trending status	positive	negative	negative-popular
trending	80	0	0
non-trending	131	105	84

➢ Public survey aligned well with interest pattern assumptions.


Recognition rates as of 2019


Degree of trending $(c_{2019} - c_{2015})$

Interest pattern
- positive
- negative
- negative-popular

Limitations/Future works

- Longer survey periods: reliability of retrospective inquiries.
- More accurate trend periods: reliability of internet search data.