

# چرا من را تگ کنید؟ شناسایی انگیزه های برچسب گذاری کامنت در اینستاگرام

ارائه دهنده: مرضیه یاوری - دانشجوی ارشد ورودی ۹۹

مرداد ۱۴۰۱

Expert Systems With Applications 202 (2022) 117171

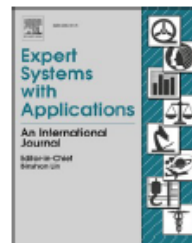


ELSEVIER

Contents lists available at [ScienceDirect](#)

## Expert Systems With Applications

journal homepage: [www.elsevier.com/locate/eswa](http://www.elsevier.com/locate/eswa)



Review

“Why tag me?”: Detecting motivations of comment tagging in Instagram<sup>☆</sup>



Jiwon Kang, Jeewoo Yoon, Eunil Park, Jinyoung Han<sup>\*</sup>

*Department of Applied Artificial Intelligence, Sungkyunkwan University, Seoul 03063, Republic of Korea*

<https://doi.org/10.1016/j.eswa.2022.117171>

Received 20 May 2021; Received in revised form 20 December 2021; Accepted 31 March 2022

Available online 7 April 2022

0957-4174/© 2022 Elsevier Ltd. All rights reserved.

## Expert Systems with Applications

### COUNTRY

United Kingdom



Universities and research  
institutions in United Kingdom

### SUBJECT AREA AND CATEGORY

Computer Science

- └ Artificial Intelligence
- └ Computer Science Applications

Engineering

- └ Engineering (miscellaneous)

### PUBLISHER

Elsevier Ltd.

### H-INDEX

225

### PUBLICATION TYPE

Journals

### ISSN

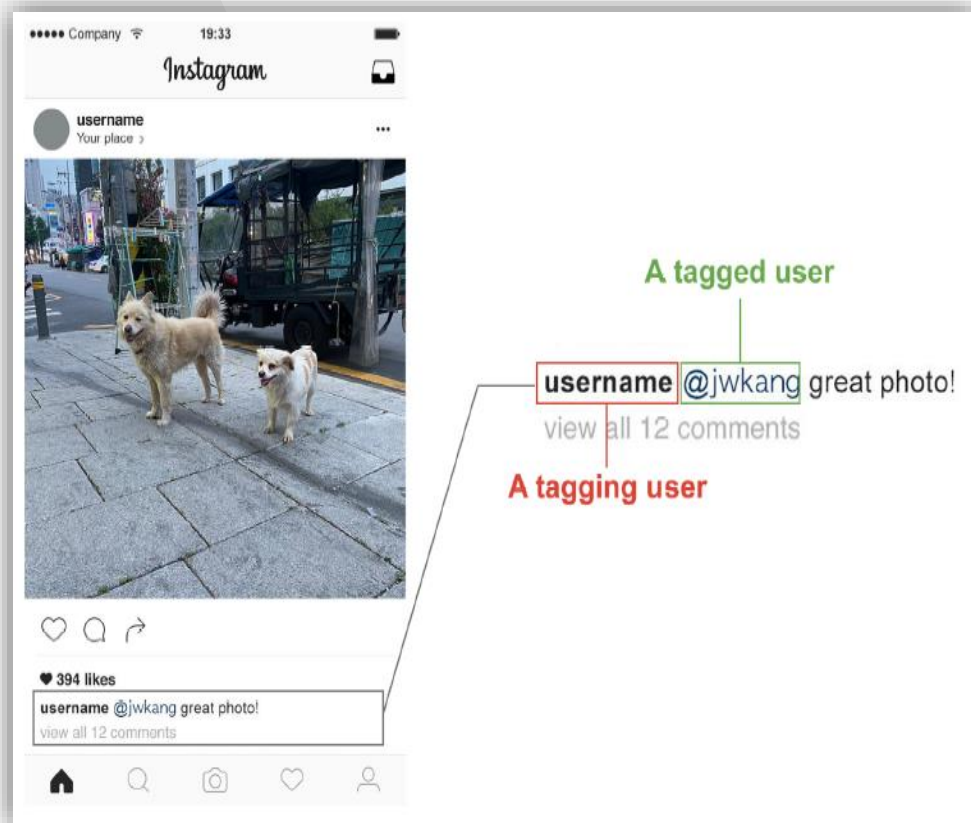
09574174

## فهرست مطالب

- ▶ مقدمه
- ▶ مجموعه داده
- ▶ کاربرد برچسب گذاری
- ▶ انگیزه برچسب گذاری
- ▶ مدل طبقه بندی انگیزه برچسب گذاری
- ▶ جمع بندی

۱.

مقدمه



- ▶ اینستاگرام شبکه اجتماعی محبوب در بارگذاری عکس و ویدیو
- ▶ قابلیت تگ کردن افراد در پست و نظرات یک مکانیسم اصلی برای برقراری ارتباط با دیگران

- ▶ کم بودن توجه به درک انگیزه های تگ کردن کاربر در کامنت
- ▶ مفید بودن درک انگیزه تگ کردن کاربر برای بازاریابی یا ارائه دهندگان محتوا





- ▶ تحلیل شیوه های تگ کردن با جمع آوری داده های تگ شده کاربر
- ▶ مطرح کردن سه سوال
  - I. تا چه حد تگ کردن کاربر در اینستاگرام رایج است؟
  - II. آیا ویژگی های زبانی متمایز در یک نظر با برچسب گذاری کاربر وجود دارد، در مقایسه با یک نظر بدون برچسب کاربر؟
  - III. چگونه کاربران دوست خود را در نظرات خود تگ می کنند؟
- ▶ ایجاد یک مدل مبتنی بر یادگیری برای طبقه بندی انگیزه تگ کردن کاربر



۲.

مجموعه داده

▶ شناسایی موضوعات محبوب در اینستاگرام

▶ شامل طنز، حیوان خانگی، سلبریتی، رسانه، بازی و طبیعت

▶ شناسایی حساب های محبوب

▶ براساس گزارشات رتبه بندی وب سایت های Statista و Socialblade

▶ انتخاب ۲۷ حساب

statista 

 SOCIAL BLADE

InstaLooter

Selenium

▶ جمع آوری پست های عمومی حساب های انتخاب شده

▶ تاریخ : ۱ ژانویه تا ۱۶ سپتامبر ۲۰۱۸

▶ ابزار: Selenium WebDriver و InstaLooter API

▶ از هر حساب

- آخرین ۹۹۴ پست

- آخرین ۵۰۴ کامنت

▶ مجموعه داده جمع آوری شده

▶ شامل ۹۵۰۱ پست، ۳۹۱۳۵۷۵ نظر و ۳۰۶۷۳۸۳ کاربر

- ▶ اطلاعات جمع آوری شده از هر پست
  - ▶ url پست، زمان ارسال، نام کاربری، شناسه کاربری ارسال، تعداد کل نظرات، تعداد لایک، متن پست
- ▶ اطلاعات جمع آوری شده از هر کامنت
  - ▶ نام کاربری کامنت، متن کامنت، نام کاربری(های) تگ شده
- ▶ اطلاعات جمع آوری شده از هر کاربر
  - ▶ تعداد کل پست‌های آپلود شده، تعداد دنبال شونده ها، تعداد دنبال کنندگان

۳.

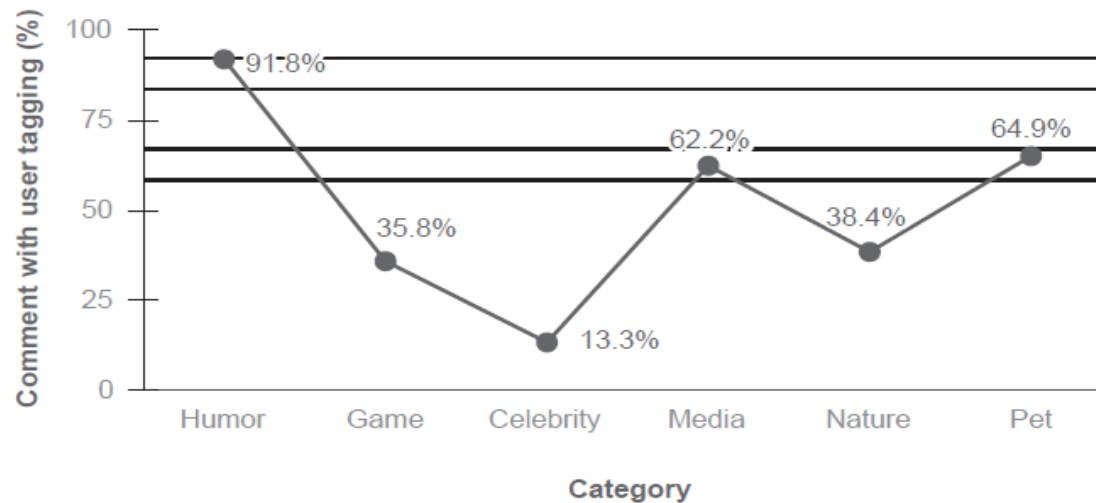
کاربرد برچسب گذاری کاربر

دنبال پاسخ به سؤالات

- ▶ (۱) تا چه اندازه برچسب گذاری کاربر در اینستاگرام رایج است؟
- ▶ (۲) آیا ویژگی های زبانی متمایز در یک نظر با برچسب گذاری کاربر در مقایسه با یک نظر بدون برچسب کاربر وجود دارد؟
- ▶ (۳) اینکه کاربران چگونه دوستان خود را در نظرات خود برچسب گذاری می کنند؟

## کاربرد برچسب گذاری کاربر (...)

- ▶ (۱) تا چه اندازه برچسب گذاری کاربر در بین دسته‌های محتوا در اینستاگرام رایج است؟
- ▶ بررسی میانگین نسبت نظرات با برچسب گذاری کاربر در شش دسته
  - ▶ گسترده بودن تگ کردن کاربر
  - ▶ میانگین نسبت نظرات با برچسب گذاری کاربران برای همه پست‌های جمع آوری شده ۵۴.۸٪ است.



## کاربرد برچسب گذاری کاربر (...)

(۲) آیا ویژگی های زبانی متمایز در یک نظر با برچسب گذاری کاربر در مقایسه با یک نظر بدون برچسب کاربر وجود دارد؟

- ▶ بررسی احساسات آشکار از نظرات با یا بدون برچسب گذاری کاربر با استفاده از LIWC
- ▶ حذف کامنت های برچسب گذاری شده بدون متن
- ▶ استفاده از نظرات نوشته شده به زبان انگلیسی
- ▶ فیلتر کردن نام کاربری تگ شده در متن
- ▶ استفاده از ۱,۰۱۱,۰۳۱ کامنت با نام کاربری تگ شده و ۹۸۳,۱۹۵ کامنت بدون برچسب گذاری کاربر
- ▶ انجام آزمون t برای هر دسته LIWC برای کامنت با و بدون برچسب گذاری کاربر



## کاربرد برچسب گذاری کاربر (...)

▶ به طور کلی کامنت با نام کاربری تگ شده

▶ تحلیلی کمتر

▶ کمتر بودن احساسات منفی

▶ شامل بودن کلمات اجتماعی و غیر رسمی بیشتر نسبت به کامنت ها بدون نام کاربری تگ شده

Features	w/ tagging		w/o tagging		F statistic	df	P-value
	M	SD	M	SD			
Analytic	44.96	40.65	57.21	40.07	-214.4	1,993,859.19	< 0.001
Negative emotion	3.16	10.7	3.39	11.99	-14.11	1,957,751.42	< 0.001
Social words	11.39	17.4	9.72	16.77	69	1,994,058.58	< 0.001
Informal words	11.94	24.66	7.39	20.25	142.43	1,940,019.47	< 0.001

## کاربرد برچسب گذاری کاربر (...)

(۳) کاربران چگونه دوستان خود را در نظرات خود برچسب گذاری می کنند؟

$$\text{TaggingEntropy} = - \sum_{i=1}^{N_u} p_i^u \log(p_i^u)$$

فرمول محاسبه آنتروپی برچسب گذاری

▶  $N_u$  تعداد کل کاربرانی است که توسط کاربر  $u$  تگ شده اند.

▶  $p_i^u$  بخشی از نظرات است که کاربر  $i$  توسط کاربر  $u$  تگ شده است.

▶ محاسبه آنتروپی برچسب گذاری

▶ (۱) یک کاربر دوستان خود را به طور مساوی تگ می کند یا روی دوستان خاصی در برچسب گذاری تمرکز می کند.

▶ (۲) یک کاربر توسط دوستان خاص یا چند دوست به طور مساوی تگ شده است.

▶ محاسبه فرکانس

▶ (۱) اندازه گیری فرکانس برچسب گذاری، که تعداد نظرات برچسب گذاری کاربر است که کاربر نوشته است.

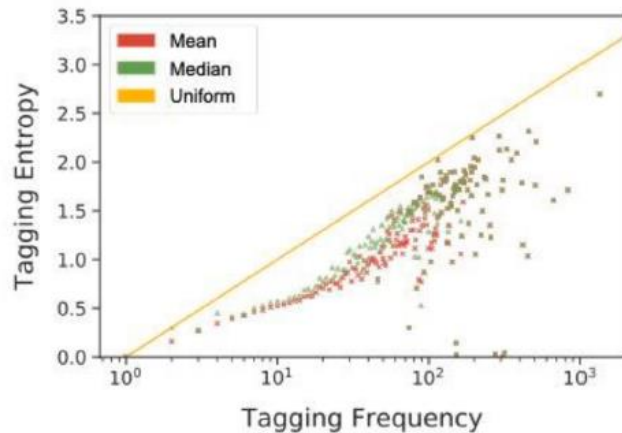
▶ (۲) اندازه گیری فرکانس برچسب گذاری شده، که تعداد نام کاربری وی است که توسط دیگران تگ شده است.

## کاربرد برچسب گذاری کاربر (...)

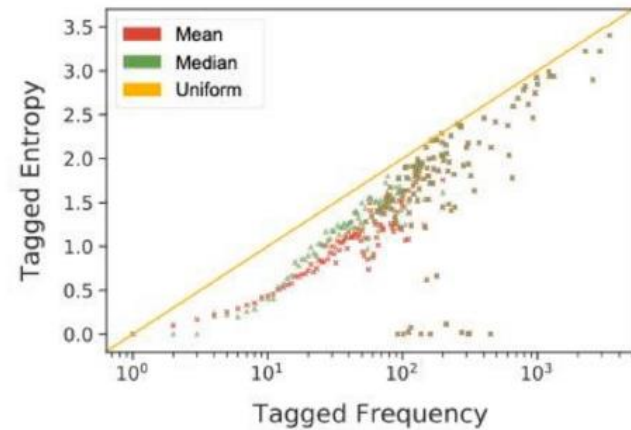
شکل (a) همبستگی بین فرکانس برچسب گذاری و آنتروپی برچسب گذاری متناظر آن

شکل (b) همبستگی بین فرکانس برچسب گذاری شده و آنتروپی برچسب شده

هر دو آنتروپی برچسب گذاری و برچسب گذاری شده کمتر از حالت یکنواختی هستند که در آن کاربر به طور مساوی سایر کاربران را برچسب گذاری می کند (یا توسط آنها برچسب گذاری می شود).  
 بیانگر رخ دادن برچسب گذاری کاربر در میان گروهی خاص از کاربران



(a) Tagging frequency vs. tagging entropy



(b) Tagged frequency vs. tagged entropy

## کاربرد برچسب گذاری کاربر (جمع بندی)

### پاسخ به سؤالات

(۱) تا چه اندازه برچسب گذاری کاربر در بین دسته های محتوا در اینستاگرام رایج است؟

▶ ۵۴.۸٪ از نظرات حاوی برچسب گذاری کاربر است، به این معنی که تگ کردن کاربر به طور گسترده در اینستاگرام استفاده می شود.

(۲) آیا ویژگی های زبانی متمایز در یک نظر با برچسب گذاری کاربر در مقایسه با یک نظر بدون برچسب

کاربر وجود دارد؟

▶ نظرات با برچسب گذاری کاربر تمایل بیشتری به اجتماعی و غیر رسمی بودن و کمتر منفی نسبت به نظرات بدون برچسب کاربر دارند

(۳) اینکه کاربران چگونه دوستان خود را در نظرات خود برچسب گذاری می کنند؟

▶ گروهی از کاربران که مرتباً با یکدیگر تگ می کنند و با یکدیگر تعامل دارند، وجود دارند و برچسب گذاری کاربر اغلب برای مکالمات بین فردی استفاده می شود.

۴.

انگیزه برچسب گذاری کاربر

## انگیزه برچسب گذاری کاربر

- ▶ براساس پاسخ سوالات مطرح شده
  - ▶ استفاده شدن تگ کردن کاربر برای رابطه اجتماعی
- ▶ استفاده از یک مدل مبتنی بر یادگیری برای شناسایی انگیزه‌های تگ کردن کاربران
- ▶ طبقه بندی انگیزه تگ کردن در سه دسته با بررسی عمل برچسب گذاری در داده ها
  - ▶ اطلاعات محور (information-oriented)
  - ▶ رابطه محور (relationship-oriented)
  - ▶ بحث محور (discussion-oriented)

## انگیزه برچسب گذاری کاربر (...)

### ▶ انگیزه اطلاعات محور

- ▶ اگر کاربر در مورد خود اطلاعاتی که از پست داده شده با کاربر(های) برچسب زده شده صحبت کند، انگیزه به عنوان انگیزه اطلاعات محور طبقه بندی می شود.
- ▶ به عنوان مثال، اگر در یک پست تصویری از سگی که لباس آووکادو پوشیده است نمایش داده شود و کاربر کامنتی مانند «lol!!» نوع مورد علاقه من از آووکادو!!» با برچسب گذاری کاربر، می توان نظر را به عنوان انگیزه اطلاعات گرا شناسایی کرد.

### ▶ انگیزه رابطه محور

- ▶ اگر کاربر با کاربر(های) برچسب گذاری شده مکالمه بین فردی داشته باشد و پست صرفاً مکالمات آن ها را واسط می شود یا تقویت کند یا حتی ذکر نشده باشد، انگیزه به عنوان انگیزه رابطه محور طبقه بندی می شود.
- ▶ به عنوان مثال، اگر یک پست در مورد یک عکاسی از منظره باشد، و کاربر با نگ کردن کاربر بنویسد «دن و من می گفتیم باید با شما بچه ها یک روز برویم»، آن کامنت را می توان به عنوان انگیزه رابطه محور شناسایی کرد.

### ▶ انگیزه بحث محور

- ▶ اگر یک کاربر و کاربر(های) برچسب گذاری شده اش در مورد پست بحث کنند، انگیزه به عنوان انگیزه بحث محور طبقه بندی می شود.
- ▶ به عنوان مثال، اگر پستی داستانی در مورد حقوق ادیان اقلیت نشان دهد، و کاربر بنویسد «به جز یک حقه برای ایجاد ادیان در یک نور منفی. Parody یک مذهب نیست» با برچسب گذاری کاربر، سپس نظر را می توان به عنوان انگیزه بحث محور طبقه بندی کرد.

## انگیزه برچسب گذاری کاربر (...)



- ▶ ایجاد داده‌های انگیزه برچسب گذاری کاربر
  - ▶ انتخاب تصادفی ۱۰۰ نظر با برچسب گذاری کاربر برای هر دسته محتوا مانند طنز یا بازی
  - ▶ ایجاد یک مجموعه داده با ۶۰۰ نظر برای شش دسته
  - ▶ تولید دو مجموعه داده مجزا
  - ▶ ۱۲۰۰ نظر مجموع دو مجموعه داده
- ▶ اختصاص هر مجموعه به دو حاشیه نویس
- ▶ طبقه بندی نظرات توسط حاشیه نویسان در یکی از سه مورد طبقه بندی
- ▶ استفاده از کامنت ها با برچسب گذاری کاربر که توسط دو حاشیه نویس، یکسان شناسایی شدن
- ▶ حذف نظراتی که در یکی از سه مورد طبقه بندی نمی شوند.
- ▶ تعداد داده های حاشه نویسی شده (مجموع: ۷۴۷ نظر)
  - ▶ انگیزه اطلاعات محور: ۳۱۳ نظر – ۴۲٪
  - ▶ انگیزه رابطه محور: ۳۶۹ نظر – ۴۹٪
  - ▶ انگیزه بحث محور: ۶۵ نظر – ۹٪



## انگیزه برچسب گذاری کاربر (...)

### شناسایی دقیق انگیزه ها

- ▶ انجام آزمون های ANOVA یک طرفه برای نشان داده ویژگی های زبان شناسی متمایز
- ▶ ویژگی های زبانی شامل امتیازات احساسات LIWC برای نظرات با برچسب گذاری کاربر با انگیزه های مختلف اطلاعات محور، رابطه محور و بحث محور

Features	Info. M	SD	Relat. M	SD	Discus. M	SD	F statistic	df	P-value
<b>LIWC sentiment scores</b>									
Social Words	7.3	17.26	17.9	21.54	12.21	7.21	26.59	(2, 744)	< 0.001
we-word	0.28	2.49	2.56	10.74	0.47	1.2	7.9	(2, 744)	< 0.001
Informal Speech	21.48	34.52	6.61	13.98	2.61	3.99	37.14	(2, 744)	< 0.001
Affect Words	22.97	32.6	8.55	13.87	7.08	5.54	36.01	(2, 744)	< 0.001
Positive emotion	17.77	29.32	6.46	12.86	3.31	4.01	29.08	(2, 744)	< 0.001
Negative emotion	4.72	17.17	2.07	6.01	3.31	4.02	4.22	(2, 744)	0.015
Cognitive activity	6.19	16.92	10.44	13.34	15.86	7.08	14.77	(2, 744)	< 0.001
<b>Other linguistic characteristics</b>									
Text length	4.95	5.23	10.57	11.33	53.54	38.96	312.47	(2, 744)	< 0.001
Portion of emoji (%)	8.15	16.86	4.12	11.1	0.33	1.17	12.85	(2, 744)	< 0.001

## انگیزه برچسب گذاری کاربر (...)

### نتایج آزمون

- ▶ (۱) نظرات اطلاعات محور: به احتمال زیاد حاوی کلمات غیررسمی مانند عامیانه اینترنتی هستند و کلمات عاطفی بیشتری را به صورت مثبت نسبت به دیگران نشان می دهند.
- ▶ (۲) نظرات رابطه محور: تمایل بیشتری به کلمات اجتماعی دارند.
- ▶ (۳) نظرات بحث محور: تمایل به داشتن کلمات بیشتری نسبت به سایرین دارند.
- ▶ کامنت‌های اطلاعات محور و رابطه محور بیشتر از نظرات بحث محور، کوتاه هستند و با شکلک‌ها صحبت می کنند.

۵.

مدل طبقه بندی انگیزه برچسب گذاری

## Features ▶

▶ استخراج ویژگی های ورودی از کامنت

▶ متن و طول آن

▶ امتیازات احساسی LIWC

▶ اطلاعات مربوط به ایموجی

▶ از -1 تا +1

▶ آموزش مدل های مبتنی بر یادگیری

▶ استفاده از متن کامنت به جای امتیازات LIWC

▶ حذف نامهای کاربری برچسب گذاری شده در متن ها

## مدل طبقه بندی انگیزه برچسب گذاری (...)

### طبقه بندی کننده ها

#### استفاده از الگوریتم های یادگیری ماشین

Naive Bayes (NB)

Logistic Regression (LR)

XGBoost (XGB)

Support Vector Machine (SVM)

#### مدل های مبتنی بر شبکه عصبی

مدل های مبتنی بر CNN

CNN(GloVe)

CNN(emoji2vec)

مدل های مبتنی بر Bert

tagBERT(BERT)

tagBERT(RoBERTa)

## مدل طبقه بندی انگیزه برچسب گذاری (...)



### ارزیابی

Model	Info.	Relat.	Discus.	Average
<b>NB</b>	73.84 (±3.21)	76.31 (±3.61)	59.83 (±13.49)	69.99 (±3.41)
<b>LR</b>	77.58 (±1.96)	80.02 (±2.77)	73.15 (±9.07)	76.92 (±3.78)
<b>XGB</b>	76.17 (±1.93)	77.61 (±2.2)	73.5 (±11.86)	75.76 (±4.74)
<b>SVM</b>	77.24 (±2.81)	80.73 (±2.33)	76.23 (±8.24)	78.07 (±4.09)
<b>CNN(GloVe)</b>	86.15 (±4.69)	82.73 (±5.23)	66.6 (±23.69)	78.49 (±7.16)
<b>CNN(emoji2vec)</b>	88.48 (±2.06)	86.01 (±0.89)	77.17 (±10.36)	83.89 (±3.72)
<b>tagBERT(BERT)</b>	88.36 (±1.0)	86.01 (±2.61)	75.65 (±10.05)	83.34 (±3.58)
<b>tagBERT (RoBERTa)</b>	87.19 (±5.61)	83.15 (±4.21)	80.83 (±2.9)	83.72 (±3.05)

تقسیم مجموعه داده

مجموعه آموزشی: ۷۰٪

مجموعه تست: ۳۰٪

محاسبه f1-score برای

هر مدل

با 95% confidence

## مدل طبقه بندی انگیزه برچسب گذاری (...)

عملکرد بهتر مدل های مبتنی بر یادگیری عمیق از مدل های یادگیری ماشین

به طور خاص، tagBERT (RoBERTa) برای شناسایی نظرات برچسب گذاری بحث محور، عملکرد قوی ای ارائه می دهد.

توسعه یک مدل طبقه بندی انگیزه برچسب گذاری بر اساس tagBERT (RoBERTa)

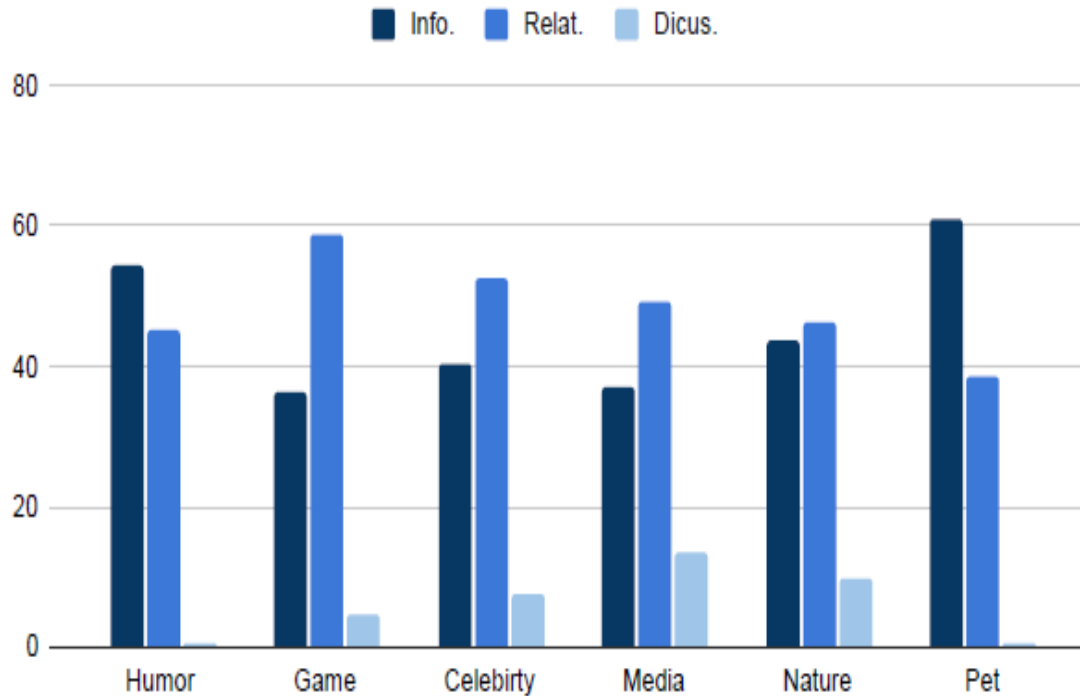
## مدل طبقه بندی انگیزه برچسب گذاری (...)

تجزیه و تحلیل انگیزه های برچسب گذاری در کل مجموعه داده

- ▶ اعمال مدل طبقه بندی پیشنهادی بر روی ۶۹۲۱۹۷ نظر با برچسب گذاری کاربر
- ▶ نظرات به زبان انگلیسی هستند
- ▶ نظرات حاوی بیش از دو کلمه



## مدل طبقه بندی انگیزه برچسب گذاری (...)



شناسایی کامنت ها با مدل پیشنهادی

▶ ۴۴.۰۸٪ کامنت ها : اطلاعات محور

▶ ۴۷.۷۴٪ کامنت ها: رابطه محور

▶ ۸.۱۸٪ کامنت ها: بحث محور

## مدل طبقه بندی انگیزه برچسب گذاری (...)

- ▶ پیدا کردن رابطه بین انگیزه های برچسب گذاری کاربر با پست های مربوطه
  - ▶ بررسی تعامل کاربر برای پست ( به عنوان نظرات)
  - ▶ ایجاد سه مجموعه پست
  - ▶ ۵ درصد برتر از کامنت های اطلاعات محور، رابطه محور و بحث محور
  - ▶ مقایسه با سایر پست ها

## مدل طبقه بندی انگیزه برچسب گذاری (...)

بیشتر بودن تعداد نرمال کامنت ها در دسته رابطه محور

▶ نشان دهنده پست با تعداد زیاد کامنت تگ شده برای معاشرت با کاربران تگ شده

▶ نبودن تفاوت آماری بین تعداد نرمال شده کامنت برای پست ها در دسته برچسب گذاری اطلاعات محور و بحث محور

Motivation	Top 5% posts		The other posts		T statistic	df	P-value
	M	SD	M	SD			
<b>Info.</b>	1.16	1.66	1.22	1.31	0.69	5431	0.49
<b>Relat.</b>	1.67	1.46	1.19	1.32	-5.29	294.77	< 0.001
<b>Discus.</b>	1.28	1.17	1.21	1.34	-0.84	5431	0.4

۶.

خلاصه

- ▶ بررسی عملکرد فعلی تگ کردن کاربران در اینستاگرام
- ▶ جمع آوری داده های برچسب گذاری کاربر در مقیاس بزرگ
- ▶ دسته بندی داده در شش دسته رایج (طنز، حیوان خانگی، سلبریتی، رسانه، بازی و طبیعت)
- ▶ ۵۴.۸٪ نظرات حاوی برچسب گذاری کاربر
  - ▶ استفاده گسترده از تگ کردن
- ▶ تجزیه و تحلیل متن کامنت ها
  - ▶ نظرات با برچسب گذاری دارای کلمات غیر رسمی نسبت به کامنت های بدون برچسب گذاری
  - ▶ یک کاربر توسط مجموعه ای از کاربران تگ می شود
- ▶ مدل پیشنهادی
  - ▶ ۴۴.۰۸ درصد در دسته اطلاعات محور، ۴۷.۷۴ درصد در دسته رابطه محور و ۸.۱۸ درصد در دسته بحث محور

# سپاس