

3rd International Workshop on Rumours and Deception in Social Media (RDSM)

C&LING
2020 | The 28th International
Conference on Computational
Linguistics



Fake or Real? A Study of Arabic Satirical Fake News

Hadeel Saadany, Emad Mohamed, Constantin Orasan

{h.a.saadany, e.mohamed2} @wlv.ac.uk, c.Orasan@surrey.ac.uk

Outline:

Part One:

- Is satirical news harmful?
- Satirical fake news vs. propaganda and hoax.
- Hypothesis:

Linguistic features per se suffice to detect satirical fake news with high accuracy

Part Two:

- ML models used for detecting Arabic Satirical Fake news.
- ML satirical fake news detection with accuracy = **98.56%**
- Conclusion.

Part One

Is Satirical Fake News humorous or harmful?



عاجل -الرئيس المكسيكي: نتفهم دوافع مصر لقتل السياح وأتساءل عن دوافع السياح الحقيقية للذهاب هناك

نشرت بواسطة: admin في أخبار عاجلة 19 ساعة مضت 0 11,122 زيارة

المكسيك - ريو دي جانيرو - وكالة تاكو للأبناء

صرح مصدر رئاسي من العاصمة المكسيكية ريو دي جانيرو تفهم المكسيك رئاسة وحكومة وشعبا وجيشا وشرطة وهيومن ترافكرز لموقف السلطات المصرية وأن لمصر كافة الحقوق للدفاع عن أراضيها ضد أي إرهاب محتمل أو أي سائح بدون تصريح انتظار.

وقال الرئيس المكسيكي انريكي اخلاسيو: "نحن لدينا صحراوات شاسعة في المكسيك وذهب مواطنين مكسيكان لصحراء مصر ينير لدي شئ من الشك عن دوافعهم الحقيقية للذهاب لمصر خصوصا أنهم لم يكن لديهم تصريح انتظار وكانت سياراتهم في الممنوع"..وأضاف متسانلا:"ما الذي جعلهم يذهبون هناك؟"

جاءت تصريحات الرئيس المكسيكي على هامش مشاركته في قمة "تاكو سوبريم" الاقتصادية الكبرى والتي رفع فيها علم مصر معلنا تضامنه مع الجيش المصري في حربه ضد الإرهاب والسياحة بدون تصريح.

وستقوم الأهرام المكسيكية بتغطية كاملة للموضوع..انتظرونا وانضموا لصفحتنا على الفيسبوك لمزيد من التغطيات العاجلة

<https://www.facebook.com/Alahraam2020>

\#thankyou_Henrico_Iglassios

Taking Satire Seriously

- The Onion
- The Beaverton
- Etc.

*"Congress Passes
Americans With No
Abilities Act"*

*"FIFA Frantically
Announces 2015
Summer World Cup In
United States"*



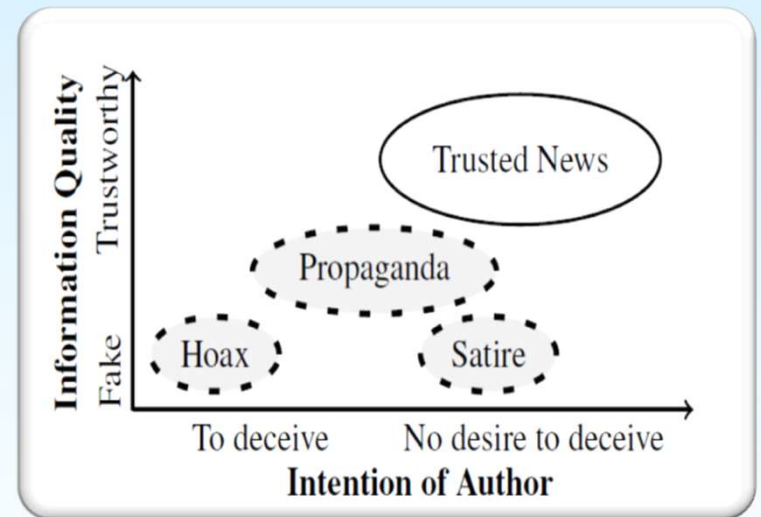
FIFA



Automatic fake news detection: Satire vs Hoax and Propaganda

Features:

- Propaganda and hoax: meta-data
 - The source
 - Reputation of the publisher
 - History of truthfulness
 - Linguistic
- Satirical: Linguistic
- Hoax and propaganda: low accuracy (20% - 70%)
- Satirical: high accuracy (85%-99%)



Part Two

Arabic Satirical Fake News Detection:

- **Data compiling:**

- Fake dataset: **(3185 articles)**
- Arabic satirical online news website ([Al Hudood](https://alhudood.net/), Mexican Ahram)
- Fact dataset: **(3710 articles)**
- BBC-Arabic, CNN-Arabic, Al-Jazeera news articles

<https://alhudood.net/>, <https://www.facebook.com/Alahraam2020/>

Linguistic-Statistical Analysis

- **Journalistic Register Measure**
- *Stated the spokesman (قال الناطق باسم), Reported by.. (تقرير أعده)*

$$J = \frac{1}{|\Omega|} \sum_{\omega \in \Omega} I_{\mathcal{C}}(\omega)$$

$$I_{\mathcal{C}}(\omega) = \begin{cases} 1 & \omega \in \mathcal{C} \\ 0 & \omega \notin \mathcal{C} \end{cases}$$

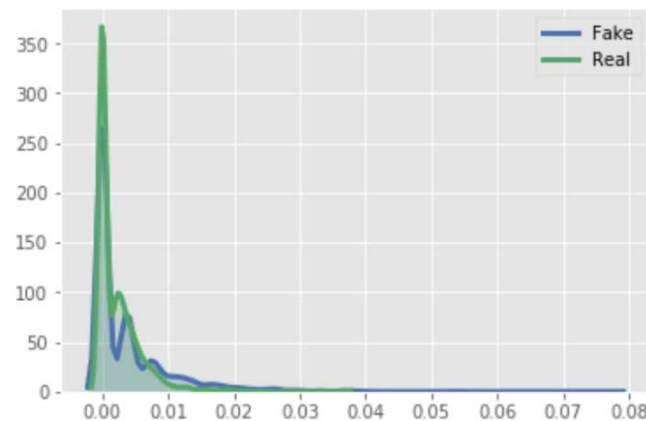


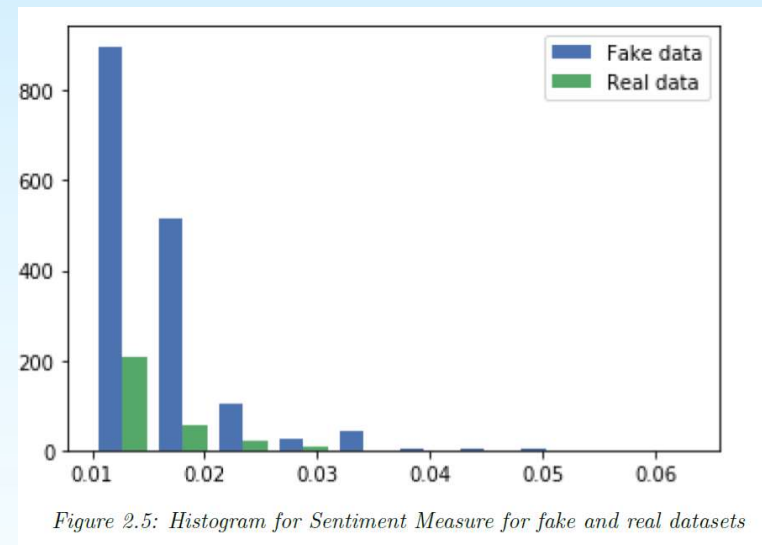
Figure 1: Probability Density Function of Journalistic Register Measure

Sentiment Intensity Measure

- Satire is created by emotional imbalance
- Statistically significant (statistic 3.27, p-value 0.001)

Subjectivity Measure

- The ratio of first-person plural verbs in a satirical fake article to a real news article was **2.2 to 1**.

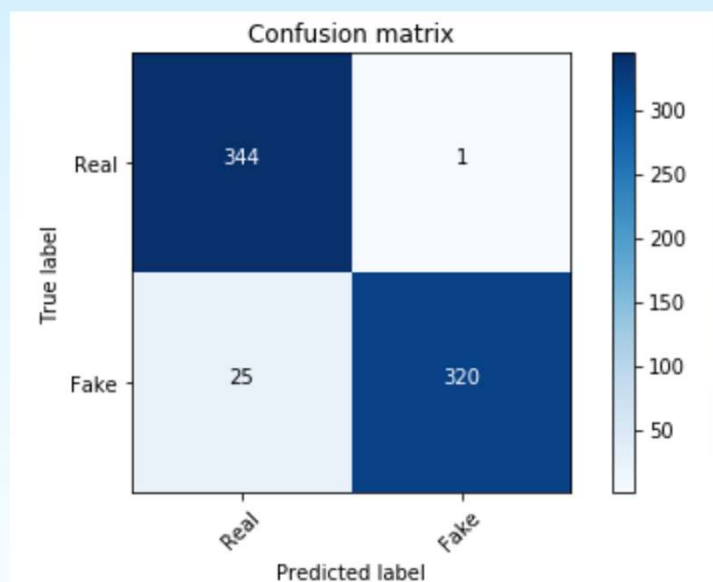


Classification Models and Results

| Model | Test Set | | | |
|--|--------------|--------------|--------------|--------------|
| | Acc | P | R | F |
| Multinomial NB (Count Vectors) | 96.23 | 96.47 | 96.23 | 96.24 |
| Multinomial NB (Word-Level TF-IDF) | 94.63 | 94.71 | 94.64 | 94.64 |
| Multinomial NB (N-Grams) | 86.08 | 88.03 | 86.09 | 86.22 |
| Multinomial NB (CharLevel) | 91.73 | 92.00 | 91.74 | 91.75 |
| Multinomial NB (Count Vectors segmented) | 95.79 | 95.97 | 95.80 | 95.81 |
| Multinomial NB (Word-Level TF-IDF segmented) | 94.49 | 94.65 | 94.49 | 94.51 |
| XGBoost with Count Vectors | 96.81 | 96.81 | 96.81 | 96.81 |
| CNN with pre-trained word embeddings | 98.59 | 98.49 | 98.61 | 98.49 |

Table 1: Summary of Accuracy (Acc), Precision (P), Recall (R) and F1 score (F) of classification models

Error Analysis:



Confusion matrix for baseline model

• Why a fake article is classified as real?

1. A perfect parody of the real news with a proliferation of journalistic clichés (15 phrases).
2. Written in MSA (Modern Standard Arabic), no slang.
3. Strong sentiment polarity but were not captured by the classifier (mostly coined hyperbole's specific of the author's style)

٢٠١٥/٠٨/١٧ | [مشاركة](#) | [✉](#) | [🐦](#) | [📞](#) | [f](#)

الحوثي يقترح مرور أوتوستراد "الضاحية الجنوبية - القدس" من صنعاء

جانب من الخريطة المقترحة للذهاب إلى الق

دعا زعيم أنصار الله عبد الملك الحوثي، رفيق دربه وسلاحه في المقاومة السيد "حسن نصر الله" لتمديد أوتوستراد يعرج في طريقه على صنعاء ليشرب كوباً من الشاي فيما هو متجه إلى القدس. وقال الحوثي أن بلاده تشكل بيئة أكثر جاذبية من دمشق لجميع الأطراف الراغبة في التوجه إلى ثاني القبلتين وثالث الحرمين الشريفين.

وكانت المفاوضات الجارية قد شهدت مؤخراً تقدماً ملحوظاً للحوثيين في مشارف "مثلث لعند"، وهو مثلث قائم الزاوية، كما يعلم الجميع. من جهتها تمسكت المقاومة الشعبية بمواقفها السابقة على ضرورة تحويل "المثلث" إلى "حلقة" أو دائرة مفرغة، أو تعديله ليصبح مثلثاً متساوي الأضلاع مثلاً.

أما فيما يخص الشؤون الإنسانية، فقد أكد مواطنون يمنيون قناعتهم أن المواد الغذائية متوفرة بكثرة في مكان ما، إلا أنهم يفضلون الزراعة البعلية والري بالتنقيط عوضاً عن الإسراف بالماء ولو كانوا على نهر جار.

وفي صعيد متصل، وصل صباح أمس وزير الخارجية السعودي إلى الجمهورية الإسلامية الإيرانية، في زيارة عائلية لتعميق الفجوة بين الشعبين الصديقين، حيث تباحث مع نظيره الإيراني أطرف جواد خفيف، آخر المستجندات في المنطقة، وسبل تطوير العلاقات الثنائية بين البلدين، كما أصدر بياناً مشتركاً يرفض أي تدخل في الشأن اليمني الذي يخص الإيرانيين والسعوديين وحدهم. ثم أضافا بعد برهة: واليمنيين، واليمنيين طبعاً.

Conclusion:

- A novel problem for Arabic Computational Linguistics
- Fake News detection accuracy is not the same for all types.
- Satirical fake news detection can be achieved by AI solutions.

Thank you