AI supporting content verification and analysis

Session: Automated Fact-Checking: The way to go?

Stuart E. Middleton
University of Southampton, Electronics and Computer Science
www.ecs.soton.ac.uk/people/sem

SciCar 2018
25th Sept 2018
Overview

• Verification tools and research
• AI supported content verification
  – Social media and web scale analytics
  – Digital image forensics
  – Digital text forensics
• Trends around AI and content verification
Speaker

- Dr Stuart E. Middleton
  - Senior research engineer
  - University of Southampton, Electronics and Computer Science (ECS), IT Innovation Centre

- Research
  - Computational linguistics and information extraction

- Interdisciplinary
  - Journalists (Deutsche Welle, REVEAL)
  - Law enforcement agencies (UK Border Force, FloraGuard; UK National Crime Agency, VIVACE)
  - Intelligence analysts from UK DSTL and UK Ministry of Defence
Verification tools and research

Verification tools and research

Web & Social Media
- Tineye
- Google Reverse Image Search
- Truly Media

Text
- FullFact.org
- PolitiFact.com
- BBC Reality Check
- Washington Post Fact Checker
- FirstDraftNews CrossCheck

Multimedia
- MAVEN Project FP7
- MediaEval Workshop
- Verifying Multimedia Use Task
- Web Multimedia Verification Workshop
- The Wild Web tampered image dataset

Workshops and Projects
- MAVEN Project FP7
- PHEME Project FP7
- SNOW Workshop
- RDSM Workshop
- TweetCred
- Factmata Google DNI
- Content Check ANR

Datasets and Fact Checking Sites
- FreeBase
- DBPedia
- YAGO2
- DBpedia
- FreeBase
- Exif.regex.info
- Google Reverse Image Search
- Truly Media Google DNI
Verification tools and research

**Multimedia**
- MediaEval Workshop
- Verifying Multimedia Use Task
- MAVEN Project FP7
- Web Multimedia Verification Workshop
- The Wild Web tampered image dataset
- Tineye.com
- Fotoforensics.com
- Exif.regex.info
- Truly Media Google DNI
- Izitru.com
- Eyewitnessproject.org
- Google Reverse Image Search

**Text**
- PHEME Project FP7
- ACL SemEval Workshop
- RumourEval Task
- SMART Project FP7
- SNOW Workshop
- RDSM Workshop
- IDIR Claim Buster
- Reconcile C3
- Webpage Credibility Dataset
- Exif.regex.info
- Meedan Check
- Factmata Google DNI
- Google DNI
-华盛顿 Post
- Fact Checker
- FirstDraftNews
- CrossCheck
- Emergent.info
- PolitiFact.com
- FullFact.org
- Truly Media Google DNI
- Web of Trust
- Freebase
- Hoax
- Twitter Trails
- Web Multimedia
- News & Journalism
- Text
- Multimedia

**Web & Social Media**

**Key**
- Datasets and Fact Checking Sites
- Workshops and Projects

© University of Southampton, 2018
Verification tools and research

How many cyclists are killed each year?

In brief

Claim
In 2016, 18,477 cyclists were injured in road accidents, including 3,499 who were killed or seriously injured.

Conclusion
Correct for Britain. The 18,477 figure includes deaths, of which there were 102.

PolitiFact
Storyful
Full Fact
Factmata
Google DNI
Fact Checker
Web & Social Media
News & Journalism
Text
Multimedia
DBPedia
ACL SemEval Workshop
RumourEval Task
MediaEval ... Image Search 
fotoforensics.com
eyewitnessproject.org
Reconcile C3
Webpage Credibility Dataset
exif.regex.info

Datasets and Fact Checking Sites
Workshops and Projects

Key
InVID Project H2020
Video Forensics

A plugin to debunk fake news and to verify videos and images

AI support for content verification

- Social media and web scale analytics
  - Information extraction
    - Tending/emerging events & factual claims
    - Context (profile, comments, network) > Verify source & claim
  - Classify image & videos from posted comments
    - Eyewitness, Fake/Genuine, First person report
AI support for content verification

- Social media and web scale analytics
  - Information extraction
    - Topic & event models, Entity extraction, OpenIE, Geoparsing
    - Social network analysis, Temporal traffic analysis
  - Classify image & videos from posted comments
    - Supervised & semi-supervised learning, Language models

https://reveal-jdss.it-innovation.soton.ac.uk/reveal_journalists_dss
AI support for content verification

- Social media and web scale analytics
  - Information extraction
  - Tending/emerging events & factual claims
  - Context (profile, comments, network) > Verify source & claim
- Image & video classification
  - What people say about URI > Eyewitness, Fake/Genuine

https://reveal-jdss.it-innovation.soton.ac.uk/reveal_journalists_dss

thampton, 2018
AI support for content verification

• Digital image forensics
  – Manual content inspection
    – Background landmarks, weather, signposts, insignia
    – Image metadata analysis
  – Manipulation detection algorithms

http://reveal-mklab.iti.gr/reveal/about.html
AI support for content verification

- Digital image forensics
  - Manual content inspection
  - EXIF metadata e.g. camera data for checking with author
  - Manipulation detection algorithms
    - Splice/copy-move detection using JPEG compression traces

http://reveal-mklab.iti.gr/reveal/about.html
AI support for content verification

 Metadata summary

<table>
<thead>
<tr>
<th>JppegComment</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>Creator: gd-jpeg v1.0 (using Livingston JPEG v2), quality = 75</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>File</th>
<th>description</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>File Name</td>
<td>Raw</td>
<td></td>
</tr>
<tr>
<td>File Size</td>
<td>3539 bytes</td>
<td></td>
</tr>
<tr>
<td>File Modified</td>
<td>Tue Jul 12 14:59:03</td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td>+03:00 2016</td>
<td></td>
</tr>
</tbody>
</table>

send this image to Google reverse image search

Double JPEG quantization inconsistencies (IQ)

JPEG Ghosts (GHOST)

JPEG blocking artifact inconsistencies (BLOCK)

Error Level Analysis (ELA)

Median Filtering noise residue [H-RIDEAN]

High frequency noise [WAVELET]
AI support for content verification

• Digital text forensics
  – Manual linguistic expert analysis
  – Author attribution, profiling & clustering
    – Academic > PAN conference series
    – Interest from law enforcement agencies (e.g. FloraGuard)

https://pan.webis.de/clef18/pan18-web/index.html
http://floraguard.org/
AI support for content verification

- Digital text forensics
  - Manual linguistic expert analysis
  - Author attribution, profiling & clustering
    - Supervised learning (word and character n-gram features), zip/checksum, repetitions, statistical profiling of vocabulary
AI support for content verification

- Digital text forensics
  - Manual linguistic expert analysis
  - Author attribution, profiling & clustering

PAN @ CLEF 2018

This is the 18th evaluation lab on digital text forensics. PAN will be held as part of the CLEF conference in Avignon, France, on September 10-14, 2018.

Evaluations will commence from January till June. We invite you to take part in any of the three tasks shown below.

- **Author Identification**
  - Given a document, who wrote it?
  - One subtask focuses on cross-domain authorship attribution applied in fanfiction and another subtask focuses on style change detection.

- **Author Profiling**
  - Given a document, what are its author's traits?
  - This task focuses on gender, whereas text and image may be used as information sources of tweets in English, Spanish and Arabic.

- **Author Obfuscation**
  - Given a document, hide its author.
  - This task works against identification and profiling by automatically paraphrasing a text to obfuscate its author's style. The tasks offered are author masking and obfuscation evaluation.

https://pan.webis.de/clef18/pan18-web/index.html
http://floraguard.org/
Trends around AI and content verification

• AI helping to handle larger content volumes
  – Machine learning increasingly automating pre-filtering of content prior to human verification work
  – Crowdsourcing being regularly used for fact checking
    – Eyewitness, Local experts, Volunteer fact checkers
  – Better integration of open data from trusted sources
  – Use of distributed AI and big data to scale up algorithms
    – Big data solutions from companies owning big datasets (e.g. Facebook algorithms to take down terrorist or child sex abuse content)

• Platforms as gateways to original content
  – Social media platform API's increasingly the only access to UGC
  – Metadata stripping by platforms could prevent many forensic techniques that needs access to original content
Trends around AI and content verification

• AI producing better fake content
  – Generative Adversarial Networks (GAN) for face swap / deepfake
    – AI can be trained to spot fakes, but AI can also be trained to produce ever more realistic fake images and videos
    – Future content might never be trusted without provenance or corroborating evidence

• Systems being designed to enhance trust in AI
  – Increased focus on AI algorithm transparency, bias, ethics
  – Better interfaces to verification AI to allow human-machine collaboration
  – Explainability of AI results so they can be understood
    – Journalists (and courts of law) need verification methods to be explainable and causal links from findings back to original evidence
Thanks you for your attention!

Any questions?

Dr Stuart E. Middleton
University of Southampton, Electronics and Computer Science, IT Innovation Centre

email: sem03@soton.ac.uk
web: www.ecs.soton.ac.uk/people/sem
www.it-innovation.soton.ac.uk
twitter:@stuart_e_middle