Auth-Integrate
Toward Combating False Data on the Internet

Romila Pradhan, Sunil Prabhakar

PURDUE UNIVERSITY
Basis of approaches to combat false data

- Incorporating user interaction
- Assessing claims collectively
- Assessing claims individually
Claims are assessed individually or in a network setting

• Different forms of fabricated data
  • deception, fake reviews, vandalisms, controversies, hoaxes, rumors

• Leverage linguistic cues to detect false data
  • aspects of language (e.g., tone, stance, objectivity, hedges, negation) to infer correctness of claims

• Utilize structure of specific community networks to identify misinformation
  • vandalism/controversies/hoaxes in Wikipedia
  • rumors on microblogging websites and social media
  • fake reviews in the services business
Multiple data conflicts resolved using truth discovery techniques

- Characterize data sources through quality measures (e.g., accuracy, precision, recall, FPR)
- Use techniques (e.g., Bayesian analysis, probabilistic graphical models, optimization and probabilistic soft logic) to jointly infer correctness of claims and credibility of sources
- Solutions strictly limited to structured data conflicts
- Strong assumption that sources are honest
Interacting with users is important

• Fact-checking websites (e.g., Snopes, PolitiFact, FactCheck) act as vanguards of truth
• Data management problems often seek human input to improve their effectiveness
  • User does not always have to be an expert
  • Advances made in crowdsourcing research and data management tasks can help in expediting the task of verifying facts
Basis of approaches to combat false data

Assessing claims individually

- different forms of false data
- linguistic cues
- community structure

Assessing claims collectively

- structured truth discovery
- infer source credibility and claim correctness

Incorporating user interaction

- prioritize questions
- manage misinformation
Distinguish correct from incorrect data, and provide explanations

Fusion resources
Knowledge graph
Master data

Articles from data sources

Data Items
Entity resolution

Sources
Source dependencies

Claims

<table>
<thead>
<tr>
<th>Entities</th>
<th>Sources</th>
<th>Claims</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>E₁</td>
<td>S₁</td>
<td>C₁₁</td>
<td>t₁</td>
</tr>
<tr>
<td>E₁</td>
<td>S₂</td>
<td>C₁₂</td>
<td>t₂</td>
</tr>
<tr>
<td>E₁</td>
<td>S₃</td>
<td>C₁₃</td>
<td>t₃</td>
</tr>
<tr>
<td>E₂</td>
<td>S₁</td>
<td>C₂₁</td>
<td>t₄</td>
</tr>
</tbody>
</table>

C₁₁ is a “fact”
C₁₂ is “rumor”

Knowledge management module

Misinformation manager
Expert
Crowd

Get feedback from users
Identify “misinfluencers” and influential sources
Correct claims, place limiting campaigns
Implement corrective measures

Truth Discovery Module
Correct
Incorrect ✗
Explanations
Output

Articles from data sources

Misinformation manager
Expert
Crowd

Get feedback from users
Identify “misinfluencers” and influential sources
Correct claims, place limiting campaigns
Implement corrective measures

Knowledge management module

Truth Discovery Module
Correct
Incorrect ✗
Explanations
Output

Misinformation manager
Expert
Crowd

Get feedback from users
Identify “misinfluencers” and influential sources
Correct claims, place limiting campaigns
Implement corrective measures

Knowledge management module

Truth Discovery Module
Correct
Incorrect ✗
Explanations
Output
AuthIntegrate, an end-to-end system aimed at combating false data on the Internet

Foundations in DB and data mining. Research advances in the areas of IE, data fusion, adversarial ML and influence propagation. Key components:

- leverages authoritative resources of information to maintain knowledge and provenance related to data items, claims and sources
- presents false data detection as truth discovery of structured data
- engages user feedback and corrective measures to recognize influential sources, (limit) maximize dissemination of (mis)information