Cross-domain Gender Detection in Twitter

Mohsen Sayyadi and Giovanni L. Ciampaglia, Alessandro Flammini

School of Informatics and Computing

Indiana University, Bloomington, IN

Computational Approaches to Social Modeling Workshop of SOCINFO16

SCHOOL OF INFORMATICS AND COMPUTING

Introduction

- Computational approaches to social modeling increasingly rely on data from online social media
- Not all the required attributes are always available
- Some interesting attributes may not be available
  - Although vast amount of structured and unstructured data are available

Introduction

- Latent feature engineering on social network:
  - like age, gender, occupation
  - Supervised fashion
  - Unstructured textual data from online profiles along with other metadata are used
- Prior works are good but still not applicable sometimes
  - Usually domain specific
- Labeled training data are required
  - Often expensive to generate
  - Prior trained model are not general enough

Challenge

- In every project we should start from scratch generating gold standards:
  - Time consuming
  - Sometimes not easy in case there is no picture or descriptive name
- Cross domain classification usually fails
  - Textual features are not portable

Meeting the Challenge

- Employing more portable features along textual features:
  - screen names
  - profile avatar
- Using some advanced machine learning techniques
  - Train different models for different subsets

Cross Domain Classification

- It has not been address seriously in the literature
  - Although mentioned that trained features were not portable to new datasets
- Reuse models across different domains
  - Training on a labeled dataset in order to mine the same latent attributes in new unlabeled datasets

Contributions of This Project

- Propose a framework for gender detection on twitter
  - Using tweets, screen name and profile avatar
  - The trained model can be used for new datasets without need to build gold standard
- 1st time use of computer vision algorithm for gender detection of twitter users
- 1st model to be used for cross domain classification
- Best state of the art accuracy
  - 96% on the most famous benchmark (Ruths and Liu 2013)

Related Works

- Domain-specific tools for gender detection Gender detection:
  – For in speech transcriptions, blogs, movie reviews, e-mail and search queries
- On social media
  – link-based and group-based classification

Related Works

- Gender Detection on twitter:
  - Usually textual features are used
  - Sometimes self reported names are used for boosting the accuracy
  - Structural features analysis was not successful(although worked for Facebook)
- Profile avatar were used for building gold standard in several works but not as features

State of the Art

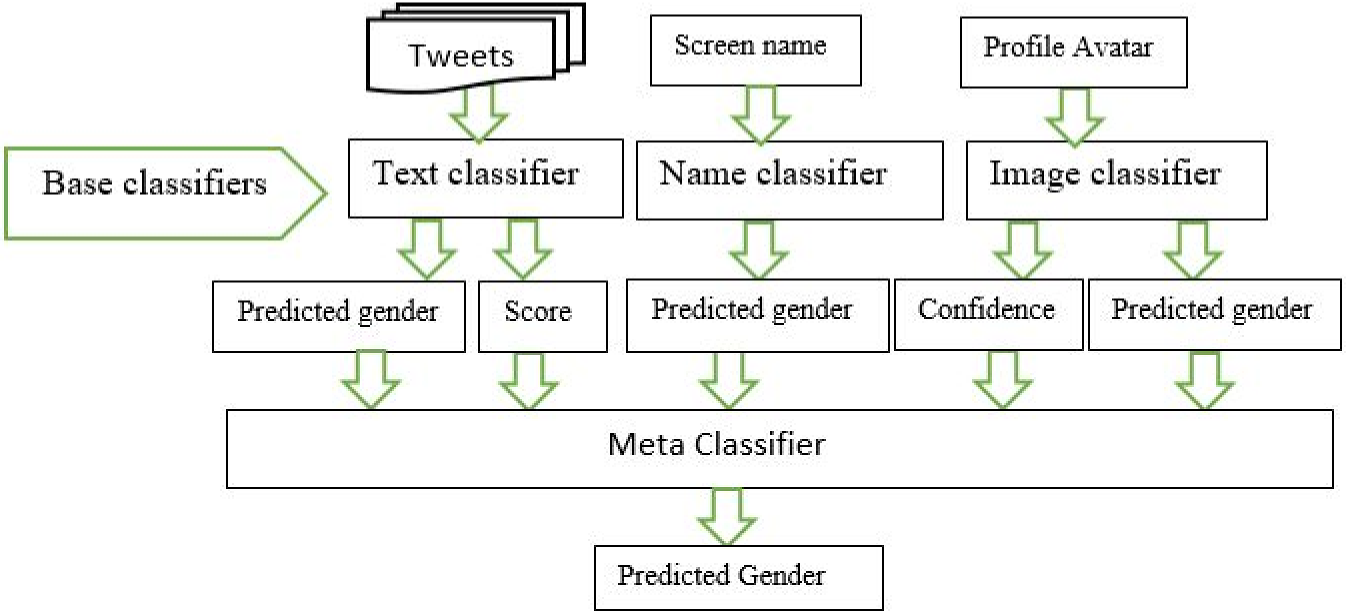
- Liu et al considered
  - textual features
  - First name
    - Most indicative signal of the gender of a person
- Best state of the art accuracy 86%
- Published their dataset

Proposed Framework

- stacked classier approach
  - Chaining multiple estimators
  - yields a more robust classier

- Rely on weak classifiers
  - Text classifier
  - Name classifier
  - Image classifier

Framework for Gender Inference



Data

- Dataset published by Ruths and Liu
- For each user, the numeric ID and a binary gender label are provided
- dataset is representative of the general Twitter population
- Selected a representative sample of users
  - who had posted at least 1,000 tweets over the lifespan of their accounts.
- Test Meta Classifier and analyze final results
- Twitter REST API were used to profile avatar and screen name and tweets
- Retweets and other simple form of near duplicates were removed

Final Datasets

- We derived two different datasets with different amount of tweets for each user

Dataset	Date range	Tweets (avg)	σ
D1	Jan 2014-Dec 2015	63	148
D2	Jan 2010-Dec 2014	530	871

Name Classifier

- Information in the self-reported screen name
- Although some people use non-descriptive nicknames
- Microsoft Discussion Graph Tool (DGT) can simplify the task
- DGT generates the label `unknown' when it is not able to classify a user with confidence.
- Coverage:
  - the fraction of cases for which DGT emits a label other than `unknown
- 88% accuracy with 51% coverage

Image Classifier

- Exploiting social media profile avatars has not been given much attention in the gender detection literature
  - at least compared to classification based on text and name
- Interestingly photos are used for building gold standards
- In prior work a sample of 15,000 random users and manually checking shows 57% of user profile pictures reflect the gender
- Publicly available Face++ library:
  - a naive deep learning face recognition tool
- Coverage:
  - the fraction of cases for which face++ could capture a face
- 87% accuracy with 32% coverage

Text Classifier

- For preprocessing step
  - removed stop word
  - transformed tweets into vectors of unigram
- Sparse vectors were fed to SVMLight
- 74% accuracy on D1 and 82% on D2
- More tweets per user usually leads to higher accuracy
- 74% accuracy on D1
- 82% accuracy on D2

Performance of Base Classifiers

- Performance of Individual classifiers

Classifier	Dataset	Acc.	Rec.	F-score	Coverage
Name	D1 + D2	88%	88%	88%	51%
Image	D1 + D2	87%	88%	88%	32%
Text	D1	74%	63%	68%	100%
Text	D2	82%	92%	86%	100%

- Correlation between base classifiers

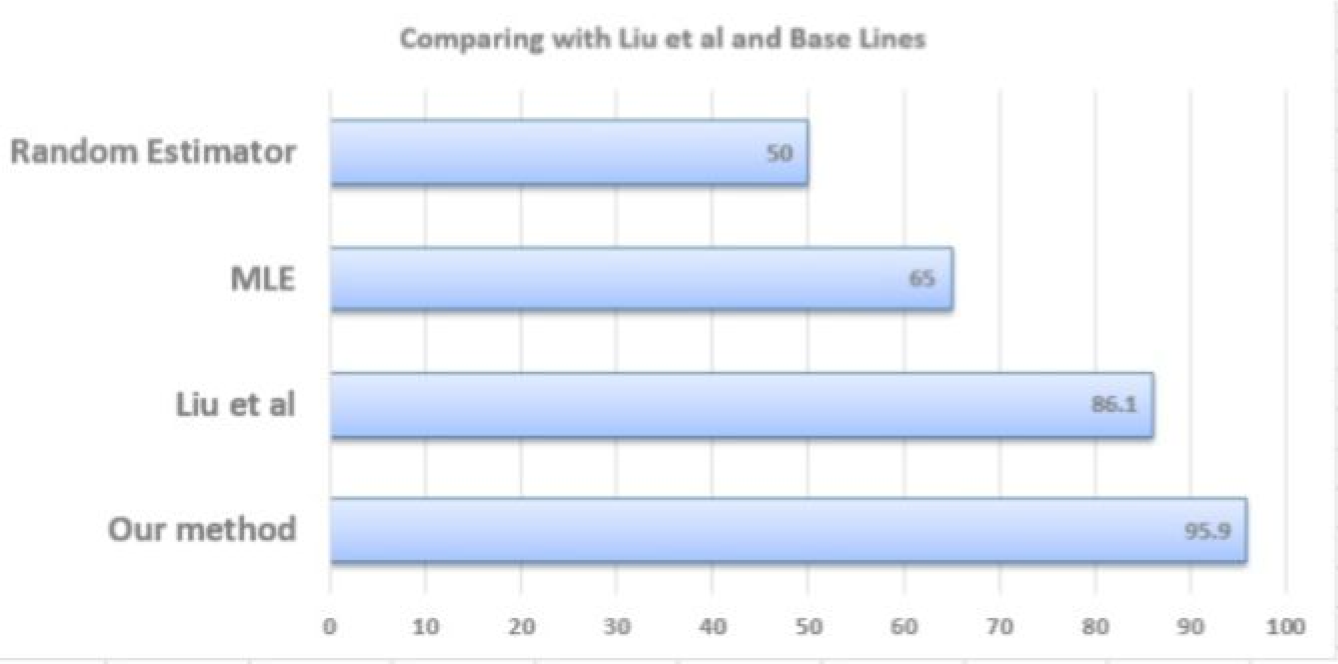
	Pearson	Spearman
Image Name	0.27	0.45
Image Text	0.34	0.56
Name Text	0.42	0.58

Performance of Meta Classifier

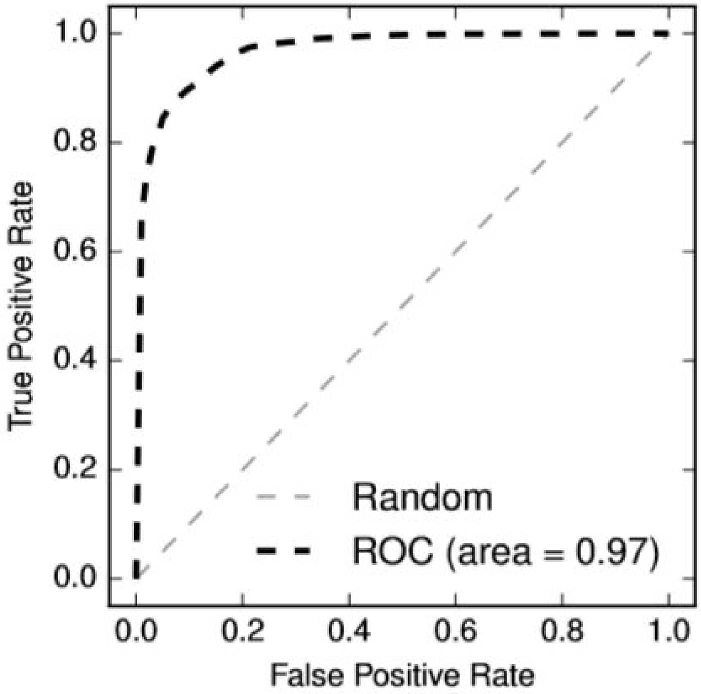
- Advanced ML techniques can find optimally weighted majority vote of weak classifiers
- Test with logistic regression also yields to better accuracy than previous state of the art

Dataset	Acc.	Rec.	F-score
D1	87.1%	88.4%	87.7%
D2	95.9%	97.1%	96.5%

Comparison with old method



ROC Diagram of stacked classifier



Cross domain classification results

- Performance of our stacked classifier for cross domain classification task on a different dataset
- BLM dataset
  - #BlackLivesMatter prominent US social movement
- results indicates
  - name and profile avatar are portable features
  - text cannot be relied for cross domain classification
- stacked classifier with inter domain classification
  - 93.4 percent in accuracy which shows applicability of our method

Results of gender detection on BLM

Dataset	feature set	type	Acc.
BLM1	text	Inter.	58.1%
BLM1	text	Cross.	58.9%
BLM1	text + face	Inter.	75.3%
BLM1	text + face	Cross.	63.3%
BLM1	text + name	Inter.	78%
BLM1	text + name	Cross.	67.8%
BLM1	face + name	Inter.	76%
BLM1	face + name	Cross.	76%
BLM1	text + face + name	Inter.	85%
BLM1	text + face + name	Cross.	72.8%

Dataset	feature set	type	Acc.
BLM2	text	Inter.	71.9%
BLM2	text	Cross.	59.4%
BLM2	text + face	Inter.	88.3%
BLM2	text + face	Cross.	62.6%
BLM2	text + name	Inter.	89.6%
BLM2	text + name	Cross.	63.5%
BLM2	face + name	Inter.	76%
BLM2	face + name	Cross.	76%
BLM2	text + face + name	Inter.	93.4%
BLM2	text + face + name	Cross.	71.1%

Conclusion

- Computer vision algorithm can be used to boost the gender classifier on twitter
- Employing a stacked classification framework can be suitable for mixing weak classifiers
- Using portable features cross domain classification is doable
  - No need to make new gold standards
- Gender detection in doable when amount of tweets is not high

Future Works

- Making more general model using boosted classifier
  - Using threshold classifier
  - If profile avatar or name are descriptive no need to consider text

- apply our framework to other platforms, like Google+
- Do more feature engineering for text classifier

References

- 1. Alvarez, A., Sierra, B., Arruti, A., Lopez-Gil, J.M., Garay-Vitoria, N.: Classifier subset selection for the stacked generalization method applied to emotion recognition in speech. Sensors 16(1), 21 (2016), http://www.mdpi.com/1424-8220/16/1/21
- 2. Bartlett, J., Norrie, R., Patel, S., Rumpel, R., Wibberley, S.: Misogyny on twitter (2014), http://www.demos.co.uk/files/MISOGYNY_ON_TWITTER.pdf, last accessed: 2016-05-16
- 3. Burger, J.D., Henderson, J., Kim, G., Zarrella, G.: Discriminating gender on twitter. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. pp. 1301{1309. EMNLP '11, Association for Computational Linguistics,
- Stroudsburg, PA, USA (2011), http://dl.acm.org/citation.cfm?id= 2145432.2145568
- 4. Burger, J.D., Henderson, J.C.: An exploration of observable features related to blogger age. In: AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs. pp. 15{20 (2006)
- 5. Ciot, M., Sonderegger, M., Ruths, D.: Gender Inference of Twitter Users in Non-English Contexts. In: Proceedings of EMNLP (2013)
- 6. Conover, M., Goncalves, B., Ratkiewicz, J., Flammini, A., Menczer, F.: Predicting the political alignment of Twitter users. In: Proc. 3rd IEEE Conference on Social Computing (SocialCom) (2011)
- 7. Davis, C.A., Ciampaglia, G.L., Aiello, L.M., Chung, K., Conover, M.D., Ferrara, E., Flammini, A., Fox, G.C., Gao, X., Goncalves, B., Grabowicz, P.A., Hong, K., Hui,
- P.M., McCaulay, S., McKelvey, K., Meiss, M.R., Patil, S., Peli Kankanamalage, C., Pentchev, V., Qiu, J., Ratkiewicz, J., Rudnick, A., Serrette, B., Shiralkar, P., Varol, O., Weng, L., Wu, T.L., Younge, A.J., Menczer, F.: OSoMe: The IUNI observatory on social media. PeerJ Preprints 4, e2008v1 (2016)

References

- 8. Deitrick, W., Miller, Z., Valyou, B., Dickinson, B., Munson, T., Hu, W.: Gender identification on twitter using the modified balanced winnow. Communications and Network Vol.04No.03, 7 (2012), http://www.scirp.org/journal/ PaperInformation.aspx?PaperID=22061
- 9. Fulper, R., Ciampaglia, G.L., Ferrara, E., Menczer, F., Ahn, Y., Flammini, A., Lewis, B., Rowe, K.: Misogynistic Language on Twitter and Sexual Violence. In: Proc. ACM Web Science Workshop on Computational Approaches to Social Modeling (ChASM) (2014), http://dx.doi.org/10.6084/m9.figshare.1291081
- 10. Garera, N., Yarowsky, D.: Modeling latent biographic attributes in conversational genres. In: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2 - Volume 2. pp. 710–718. ACL '09, Association for Computational Linguistics, Stroudsburg, PA, USA (2009), http://dl.acm.org/citation.cfm?id=1690219.1690245
- 11. Herring, S.C., Paolillo, J.C.: Gender and genre variation in weblogs. Journal of Sociolinguistics 10(4), 439–459 (2006), http://dx.doi.org/10.1111/j.1467-9841. 2006.00287.x
- 12. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Tech. rep., Technical Report 07-49, University of Massachusetts, Amherst (2007)
- 13. Joachims, T.: Making large-scale svm learning practical. LS8-Report 24, Universita¨t Dortmund, LS VIII-Report (1998)

References

- 14. Jones, R., Kumar, R., Pang, B., Tomkins, A.: ”i know what you did last summer”: Query logs and user privacy. In: Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management. pp. 909–914. CIKM ’07, ACM, New York, NY, USA (2007), http://doi.acm.org/10.1145/1321440. 1321573
- 15. Kıcıman, E., Counts, S., Gamon, M., De Choudhury, M., Thiesson, B.: Discussion graphs: Putting social media analysis in context. In: Intl. Conf. on Weblogs and Social Media (ICWSM-14). AAAI (June 2014), http://research.microsoft.com/ apps/pubs/default.aspx?id=210256
- 16. Lazer, D., Pentland, A.S., Adamic, L., Aral, S., Barabasi, A.L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., et al.: Life in the network: the coming age of computational social science. Science 323(5915), 721 (2009)
- 17. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature 521(7553), 436–444 (May 2015), http://dx.doi.org/10.1038/nature14539
- 18. Liu, W., Al Zamal, F., Ruths, D.: Using social media to infer gender composition of commuter populations. Sixth International AAAI Conference on Weblogs and Social Media abs/1405.6667 (2012)
- 19. Liu, W., Ruths, D.: what’s in a name? using first names as features for gender inference in twitter. AAAI Spring Symposium Series abs/1405.6667 (2013)
- 20. Ludu, P.S.: Inferring gender of a twitter user using celebrities it follows. CoRR abs/1405.6667 (2014), http://arxiv.org/abs/1405.6667
- 21. Mislove, A., Lehmann, S., Ahn, Y.Y., Onnela, J.P., Rosenquist, J.N.: Understanding the Demographics of Twitter Users. In: Proceedings of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM’11). Barcelona, Spain (July 2011)
- 22. Mohammad, S.: Portable features for classifying emotional text. In: Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. pp. 587–591. NAACL HLT ’12, Association for Computational Linguistics, Stroudsburg, PA, USA (2012), http://dl.acm.org/citation.cfm?id=2382029.2382123
- 23. Moreno, J.L.: Who shall survive?: A new approach to the problem of human interrelations. Nervous and Mental Disease Publishing Co, Washington, DC, US (1934)
- 24. Nguyen, D.P., Gravel, R., Trieschnigg, R.B., Meder, T.: ”how old do you think i am?” a study of language and age in twitter. In: Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media, ICWSM 2013, Cambridge, MA, USA. pp. 439–448. AAAI Press, Palo Alto, CA, USA (July 2013)
- 25. Nilizadeh, S., Groggel, A., Lista, P., Das, S., Ahn, Y.Y., Kapadia, A., Rojas, F.: Twitter’s glass ceiling: The effect of perceived gender on online visibility. In: Tenth International AAAI Conference on Web and Social Media (2016)
- 26. Olteanu, A., Weber, I., Gatica-Perez, D.: Characterizing the demographics behind the #BlackLivesMatter movement. In: Proceedings of the AAAI Spring Symposia on Observational Studies through Social Media and Other Human-Generated Content (SSS’16 OSSM). Stanford, US (2016)


Questions?