Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)

Interactive Gender Inference with Integer Linear Programming

Shoushan Li[1,2], Jingjing Wang[1,2], Guodong Zhou[1,2*], Hanxiao Shi[3]
[1] School of Computer Science and Technology, Soochow University, Suzhou, China
[2] Collaborative Innovation Center of Novel Software Technology and Industrialization
[3] School of Computer Science and Information Engineering, Zhejiang Gongshang University, China
{lishoushan, 20134227025, gdzhou}@suda.edu.cn, hxshi@mial.zjgsu.edu.cn

Abstract

Interactive gender inference aims to infer the genders of the two involved users in a communication from the interactive text. In this paper, we address this task by proposing a joint inference approach which well incorporates label correlations among the instances. Specifically, an Integer Linear Programming (ILP) approach is proposed to achieve global optimization with various kinds of intra-task and extra-task constraints. Empirical studies demonstrate the effectiveness of the proposed ILP-based approach to interactive gender inference.

Introduction

Gender inference is concerned with determining whether a user is *male* or *female* by analyzing the user-generated content. This task has received considerable attention in several research communities, such as natural language processing and social network analysis. Whereas most of earlier studies on gender inference focus on formal texts with only a few candidate users, e.g., in Email (Corney et al., 2002) and Blog (Schler et al., 2006), more and more researchers have recently turned their attention to scenarios involving informal texts and tens of thousands of users, e.g., in social media (Miller et al., 2012).

In social media, it is worthwhile to highlight that the large number of users are not in isolation but correlated to each other. Therefore, in this scenario, a user-generated text is normally shared by several users instead of a single user. For instance, in Figure 1, user **A** posts a message and his friends **B** and **C** comment on his message. Here, the message "*So amazing a Louis Vuitton's…*" and the comment "*Wow, the necklace matches your earrings so perfect! I really envy you, sistah!!*" becomes an interactive text from **B** to **A**, denoted as **B** $\rightarrow$ **A**, from which, we are able to infer not only the gender of user **B** to be *female*, but also that of user **A** to be *female* due to the occurrence of word "*sistah*" in the comment. In the same spirit, we can infer **C** $\rightarrow$ **A** to be *male* $\rightarrow$ *female* from the interactive text due to the occurrence of nominal phrase "*my wife*" in the comment "*You look so sexy today! I wanna give my wife the same one. Where could I find it?*". To differentiate from previous studies, we refer to the task of inferring both genders of the involved two users from an interactive text as interactive gender inference, while the traditional one, the task of inferring the gender of individual user from the message text as individual gender inference.

Message:

User **A**: *So amazing a Louis Vuitton's necklace for Chrismas gift. Love you baby husband. Am I girly with it? Haha.*

Comments:

User **B**: *Wow, the necklace matches your earrings so perfect! I really envy you, sistah!!*

User **C**: *You look so sexy today! I wanna give my wife the same one. Where could I find it?*

- **Individual Gender Inference**
  - **Input**: message text;
  - **Output**: **A**: *female*
- **Interactive Gender Inference**
  - **Input**: interactive text (message + comment);
  - **Output**: **B** $\rightarrow$ **A**: *female* $\rightarrow$ *female*

Figure. 1. An example of message and corresponding comments in social media

Beyond conventional applications of individual gender inference, such as intelligent marketing and automatic advertising (Mukherjee and Liu, 2010; Burger et al, 2001; Volkova et al, 2013), interactive gender inference has its own unique applications. For example, it could be used to enable human-computer interaction more humanized and friendly. More specifically, when responding to a human with known gender, the computer could select a gender-aware response from

[*] Corresponding author

many possible candidates to make the user more comfortable. Such scenario exactly needs the technology of interactive gender inference.

Following the footstep of individual gender inference, we could simply adopt a straightforward approach to interactive gender inference, which conceptualizes it as a multi-category classification problem with 4 categories, i.e. *male_to_male* (*mm*) *male_to_female* (*mf*), *female_to_male* (*fm*), and *female_to_female* (*ff*). However, it is worthwhile to note that interactive gender inference exhibits a number of specific characteristics that might be promising in facilitating the improvement on the inference performance. One distinct one, quite different from individual gender inference, is that the label of one instance may be not in isolation but highly correlated with those of other instances from either the same task or another task (e.g., individual gender inference).

On one hand, the label of one instance from interactive gender inference might be correlated with those of other instances from the same task. This is because different instances might share the same user. For instance, in Figure 1, $\mathbf{B} \rightarrow \mathbf{A}$ and $\mathbf{C} \rightarrow \mathbf{A}$ share the same user **A** and their categories should be the same in terms of user **A**. (e.g., *mm* and *fm* are possible but *mm* and *ff* are impossible). This could be utilized to keep gender consistency for a certain user. In this way, the data sparseness and information scarcity problems can be much alleviated. For instance, it might be hard to tell whether $\mathbf{C} \rightarrow \mathbf{A}$ is *mf* or *mm* due to the lack of deterministic information for gender inference from the interactive text. However, once the instance of $\mathbf{B} \rightarrow \mathbf{A}$ is confidently predicted as *ff* , we can certainly infer $\mathbf{C} \rightarrow \mathbf{A}$ to *mf* because user **A** has been predicted as *female* in $\mathbf{B} \rightarrow \mathbf{A}$.

On the other hand, the label of one instance from interactive gender inference might be correlated with those of other instances from the task of individual gender inference. This is because one instance in interactive gender inference might share the same user as the instances from the task of individual gender inference. This could help obtain better performance in interactive gender inference by introducing more knowledge from the task of individual gender inference. For instance, in Figure 1, it might be hard to tell whether $\mathbf{B} \rightarrow \mathbf{A}$ is *mf* or *mm* due to the lack of deterministic gender information in the interactive text. However, once the instance of user **A** is confidently predicted as *female* in the task of individual gender inference, we can certainly infer $\mathbf{B} \rightarrow \mathbf{A}$ to *mf*. Moreover, this could benefit individual gender inference as well. For instance, once $\mathbf{B} \rightarrow \mathbf{A}$ is confidently inferred as *ff* in interactive gender inference, we can know that the genders of users **B** and **A** are *female* and *female* respectively in individual gender inference.

In this paper, we address interactive gender inference by incorporating label correlations among the instances with a joint inference approach. First, we model interactive gender inference as a 4-category classification problem, similar to individual gender inference, and adopt a standard machine learning-based approach with a variety of features including basic bag-of-words features, F-measure features (Heylighen and Dewaele, 2002; Nowson et al., 2005), and POS sequence patterns (Mukherjee and Liu, 2010). Then, we achieve global optimization via a joint inference approach, named Integer Linear Programming (ILP), where the label correlations among the instances in the same task and across different tasks are modeled as two kinds of constraints, namely intra-task and extra-task constraints respectively. Specifically, a global objective function is minimized with the obtained posterior probabilities of the test instances. Empirical studies demonstrate that both intra-task and extra-task constraints are very effective in interactive gender inference.

The remainder of this paper is organized as follows. Section 2 overviews related work on gender inference. Section 3 introduces data collection. Section 4 describes our joint approach to interactive gender inference. Section 5 presents the experimental results. Finally, Section 6 gives the conclusion and future work.

2 Related Work

In the last decade, many studies have been devoted to gender inference (also called gender classification) from several research communities, such as natural language processing and social network analysis, differing primarily in different styles of text and different types of features.

As an important style of text, blog has been a focus of previous studies on gender inference. Schler et al. (2006) exploit the differences in writing style and content between *male* and *female* bloggers to determine an unknown author's gender on the basis of a blog vocabulary. Yan and Yan (2006) present a Naive Bayes (NB) classification approach to identify the genders of weblog authors. Nowson et al. (2006) utilize the weblog data to construct a feature set for automatic gender detection. Mukherjee et al. (2010), Peersman et al. (2010), and Gianfortoni et al. (2011) focus on exploring more effective features to improve the performance.

As another important style of text, email has attracted considerable attention on gender inference. Corney et al. (2002) extract various kinds of content-free features from email text, such as style markers and structural patterns, to classify the user gender. Mohammad et al. (2011) show that there are marked differences across genders in how they use emotion words in work-place email.

Recently, micro-blog has been becoming more prevalent in gender inference due to the rapid growth of social networks. Burger et al. (2011) describe the construction of a large multilingual dataset labeled with the genders of Twitter users. Miller et al. (2012) identify the genders of Twitter users using Perceptron and Naïve Bayes with selected *n*-gram features. Ciot et al. (2013) conduct the first assessment of latent attribute inference in various languages beyond English, focusing on gender inference of Twitter users. Alowibdi et al. (2013) explore language-independent gender inference on the basis of Twitter text.

Unlike all above studies which infer the gender of a single user, our study aims to identify the genders of two interactive users. To the best of our knowledge, this is the first attempt to address this task.

3 Data Collection

Our data is collected from Sina Micro-blog (http://weibo.com/), a famous Micro-blogging platform in China. In this platform, local users publish short messages and other users are allowed to respond to the messages. Here, the response text is called comment. From the website, we crawl each user's homepage which contains user information (e.g. *name*, *gender*, *verified type*), messages and corresponding comments. The data collection process starts from some randomly selected users, and then iteratively gets the data of their followers and followings. To guarantee the reliability of the data, we remove those unsuitable users who are verified as organization or have less than 50 followers or 50 followings.

Table 1 shows the statistics about the *male* and *female* users with the number of average messages each user posts and the number of average comments each user gets from other users.

Table 1: Statistics of the *male* and *female* users

	#Users	#Average messages	#Average comments
male	21539	141	77
female	32029	150	102

In this study, we assume that an interaction from user **A** to user **B** happens when user A writes a comment as a response to a message posted by **B**. Thus, the instance of the interactive text from **A** to **B** contains all interactive texts from **A** to **B**.

4 Interactive Gender Inference

In social media, the users and their interactions can be represented as a graph, i.e., $G=(V,E)$ where V is the set of $|V|=N$ users and $E \subset V \times V$ is a set of $|E|=M$ user interactions, i.e. interactive edges among users. Here, interactive edge $e_{i,j} \in E$ is directed and represented by the interactive text from $v_i \in V$ to $v_j \in V$.

Specifically, **Interactive Gender** can be represented as a triple $(e_{i,j}, r_{ij}, P_{ij})$, where $e_{i,j} \in E$ is an edge; $r_{ij} \in R$ $R=\{mm, mf, fm, ff\}$ is the label associated with the edge $e_{i,j}$. P_{ij} is the detailed probability distribution over the labels in *R*.

4.1 Four-category Classification

In standard supervised classification, a predictor *f* is trained to map an input vector *x* into the corresponding class label *y*. In interactive gender inference, the input vector *x* is a feature vector extracted from on the interactive text, containing both the message and the comment associated with $e_{i,j} \in E$.

Formally, the objective of four-category classification is illustrated as follows:

$$f(x) \rightarrow y \\ \text{Where } y \in R \text{ and } R=\{mm, mf, fm, ff\} \tag{1}$$

In this first step, this predictor is used to predict the interactive genders of all the edges in the test data. That is to say, we obtain all the triples $(e_{i,j}, r_{ij}, P_{ij})$ where $e_{i,j}$ is one sample in the test data. Specifically, P_{ij} contains the probabilities of the sample belonging to each category, i.e.,

$$P_{ij} = < p(r_{ij}=mm), p(r_{ij}=mf), p(r_{ij}=fm), p(r_{ij}=ff) > \tag{2}$$

Note that a variety of effective features has been proposed in previous studies for individual gender inference. We borrow some of them into our task of interactive gender inference. Specifically, the following kinds of features are investigated in this study:

(1) **Bag-of-words features (BOW)**, basic word features popularly utilized in gender inference.

(2) **F-measure feature**, a unitary measure of text's relative contextuality using the POS frequencies, i.e.,

$$F = 0.5 * [(freq.noun + freq.adj + fre.prep + freq.art) \\ -(fre.pron + freq.verb + freq.adv + freq.int) + 100] \tag{3}$$

Where *freq.x* denotes the frequency of POS tag *x*. As proposed in Heylighen and Dewaele (2002) and Nowson et al. (2005), this feature explores the notion of implicitness of text.

(3) **POS sequence patterns**, POS sequence features extracted using the algorithm, as proposed in Mukherjee and Liu (2010), to capture the writing styles of users from different genders.

Although other features, e.g. language features, have been recently proposed to boost the performance of gender inference (Alowibdi et al., 2013), they are not applicable in our task due to their dependence on specific languages or tasks.

4.2 Global Optimization

Alternatively, the label in interactive gender inference could be represented as a 4-dimention vector and the value of each element in the vector is either 1 or 0 denoting whether the corresponding label (i.e., *mm*, *mf*, *fm*, or *ff*) is assigned or not. For instance, r_{ij}=[1,0,0,0] denotes the assigned label is *mm* and $r_{ij}[k]$ denotes the *k*-th label in the vector and $P_{ij}[k]$ denotes the probability belonging to the *k*-th label.

4.2.1 ILP with Intra-task Constraints

In this subsection, we leverage the label correlations among the instances in the same task to boost the inference performance. Specifically, the objective function can be defined as follows:

$$\min \sum_{e_{i,j} \in E} \sum_{k=0}^{3} r_{ij}[k] \cdot (-\log P_{ij}[k]) \tag{4}$$

Subject to:

$$r_{ij}[k] \in \{0,1\} \tag{5}$$

$$\sum_{k=0}^{3} r_{ij}[k] = 1 \tag{6}$$

Where formula (6) implies that the task of interactive gender inference is a single-label classification problem. That is, the label of an instance could be only one option in *R*, i.e., $\{mm, mf, fm, ff\}$. Note that the above objective function aims to minimize the similarity between the label vector and the probability vector. When no other constraints are available, the best solution to the above function is to assign the sample with 1 to the label which takes the maximum posterior probability.

For interactive gender inference, the constraints implied in the correlations among the instances in the same task could be categorized into the following three categories:

(C1) Intra-Left-Left constraint: When the left node of an edge is the same as the left one of another edge, the genders of the two left nodes should be the same. For instance, instances **A** $\rightarrow$ **B** and **A** $\rightarrow$ **C** should have the same gender on the left node, i.e., **A**.

$$(r_{ij}[0] + r_{ij}[1]) - (r_{il}[0] + r_{il}[1]) = 0 \quad \forall e_{i,j} \in E \wedge e_{i,l} \in E \tag{7}$$

(C2) Intra-Right-Right constraint: When the right node of an edge is the same as the right one of another edge, the genders of the two right nodes should be the same. For instance, instances **B** $\rightarrow$ **A** and **C** $\rightarrow$ **A** should have the same gender on the right node, i.e., **A**.

$$(r_{ij}[0] + r_{ij}[2]) - (r_{lj}[0] + r_{lj}[2]) = 0 \quad \forall e_{i,j} \in E \wedge e_{l,j} \in E \tag{8}$$

(C3) Intra-Left-Right constraint: When the left node of an edge is the same as the right one of another edge, the genders of the left and right nodes should be the same. For instance, instances **A** $\rightarrow$ **B** and **C** $\rightarrow$ **A** should have the same gender on the shared node, i.e., **A**.

$$(r_{ij}[0] + r_{ij}[1]) - (r_{li}[2] + r_{li}[3]) = 0 \quad \forall e_{i,j} \in E \wedge e_{l,i} \in E \tag{9}$$

It is worthwhile to note that due to the same nature of another constraint (i.e., the Intra-Right-Left constraint) as the Intra-Left-Right constraint, we omit it in our description.

4.2.2 ILP with Extra-task Constraints

In this section, we address ILP with extra-task constraints from individual gender inference, which determines the gender of a single user from the message of the interactive text.

Formally, **Individual Gender** can be represented as a triple (v_i, r_i, P_i), where $v_i \in V$ is a node; $r_i \in R'$ $R' = \{male, female\}$ is a label associated with the node. For the convenience of computation, we assume $r_i = 1$ if the label is *male*; Otherwise, $r_i = 0$. P_i is the probability distribution over the labels in R'.

As a result, the objective function (4) becomes:

$$\begin{aligned} \min & \sum_{e_{i,j} \in E} \sum_{k=0}^{3} r_{ij}[k] \cdot (-\log P_{ij}[k]) \\ & + r_i \cdot (-\log p_i) + (1 - r_i) \cdot (-\log(1 - p_i)) \\ & + r_j \cdot (-\log p_j) + (1 - r_j) \cdot (-\log(1 - p_j)) \end{aligned} \tag{10}$$

Where p_i denotes the posterior probability of the *i*-th user belonging to the *male* category, obtained from individual gender inference.

Accordingly, the constraints implied in the correlations among the instances from the extra task of individual gender inference could be categorized into the following two categories:

(C4) Extra-Left constraint: When the left node of an edge in the task of interactive gender inference is the same as the one in the task of individual gender inference, the genders of the two nodes should be the same. For instance, instance **A** $\rightarrow$ **B** in interactive gender inference and instance **A** in individual gender inference should have the same gender of user **A**.

$$r_i - (r_{ij}[0] + r_{ij}[1]) = 0 \quad \forall e_{i,j} \in E \wedge v_i \in V \tag{11}$$

(C5) Extra-Right constraint: When the right node of an edge in the task of interactive gender inference is the same as the one in the task of individual gender inference, the genders of the two nodes should be the same. For instance, instance **B** $\rightarrow$ **A** in interactive gender inference and instance **A** in individual gender inference should have the same gender of user **A**.

$$r_j - (r_{ij}[0] + r_{ij}[2]) = 0 \quad \forall e_{i,j} \in E \wedge v_j \in V \tag{12}$$

5 Experimentation

5.1 Experimental Settings

Data: The data set contains 53,675 users, as described in Section 3. From this data set, we select the largest two separate user groups (i.e. without interactions between the two groups), one with 20191 users as the training group and the other with 9339 users as the test group. Since it is difficult to infer the interactive gender if the two involved users have too few interactive comments, we omit those instances with less than 10 comments. Table 2 shows the statistics about the instances in each interactive gender category. Table 3 shows the statistics about the instances in each individual gender category in both training data and test data.

Table 2: #Instances in each category of interactive gender inference

#	#Training Data	#Test Data
mm	2883	1109
mf	4462	1599
ff	10954	3395
fm	4596	1591

Table 3: #Instances in each category of individual gender inference

#	#Training Data	#Test Data
male	3006	1761
female	4623	1253

Features: Three types of boolean features, including bag-of-words, f-measure, and POS pattern features, are adopted in our experiments. To get word and POS features, we use the tookit ICTCLAS (http://www.ictclas.org/ictclas_download.aspx) to perform word segmentation and POS tagging on the Chinese text.

Classification Algorithm: Three different classification algorithms are investigated, including naïve Bayes (NB), support vector machines (SVM), and maximum entropy (ME). Specifically, the NB and ME algorithms are implemented with the Mallet Toolkit[2] and the SVM algorithm implemented with the SVM-light Tookit[3]. Besides, all the outputs are provided with probability/confidence from these tools. The ILP is solved by the lp_solve 5.5.2.0 Toolkit[4].

Evaluation Measurement: The performance is evaluated using the standard accuracy measurement.

Significance test: *T*-test is used to evaluate the significance of the performance difference between two approaches (Yang and Liu, 1999).

5.2 Experimental Results without Joint Inference

Without joint inference, a standard 4-category classification approach is adopted to classify the interactive text instances in the test data. Table 4 shows the performances of this approach when different classification algorithms and different kinds of features are leveraged. From this table, we can see that, among the three classification algorithms, SVM and ME yields similar results, which are much better than those achieved by NB. Besides, with regard to the features, we find that leveraging more sophisticated features, either f-measure or POS pattern features, could yield better performance than using BOW features only. Especially, when both these two kinds of features are used, the performance reaches the best of 0.622 by ME. In the following experiments involving interactive gender inference, we only use the ME classification algorithm with all features due to their best performance.

Table 4: Performance of interactive gender inference without joint inference when different classification algorithms and different kinds of features are leveraged

Features	NB	SVM	ME
BOW	0.539	0.595	0.605
BOW+F-measure	0.545	0.610	0.613
BOW+POS-pattern	0.556	0.615	0.615
All features	0.560	0.619	0.622

5.3 Experimental Results with Joint Inference

In joint inference, the ILP approach is proposed to incorporate various kinds of label constraints to achieve global optimization.

5.3.1 Intra-task Constraints

Table 5: Performances of interactive gender inference with joint inference when different kinds of intra-task constraints are utilized

	Accuracy
Baseline	0.622
ILP(C1)	0.679
ILP(C2)	0.655
ILP(C3)	0.760
ILP(C1+C2+C3)	0.771

In this experiment, both the message and the comment text is used to predict the interactive gender of an edge, while the message text is used to predict the individual gender of a user. Table 5 shows the performance of the ILP approach when different kinds of constraints are used. From this table, we can see that all kinds of constraints are capable of improving the performance. Among the three kinds of constraints, C3

[2] http://mallet.cs.umass.edu/

[3] http://svmlight.joachims.org/

[4] http://web.mit.edu/lpsolve_v5520

(the Intra-Left-Right constraint) performs best. When all these constraints are utilized, the inference performance reaches the best of 0.771 in accuracy, 0.149 better than the baseline. This indicates the beneficiary of label correlations among the instances in the same task to interactive gender inference. Significance test show that our ILP approach with either single type or all types of intra-task constraints all significantly outperforms the baseline approach (p-value<0.01)

5.3.2 Extra-task Constraints

In this experiment, we use the message text only to infer the gender of a single user, i.e., individual gender inference.

As a preliminary experiment, Table 6 shows the performance of individual gender inference. From this table, we can see that, similar to interactive gender inference, it is beneficial to employ f-measure and POS pattern features in individual gender inference. As far as the classification algorithms are concerned, SVM and ME perform much better than NB.

Table 6: Performance of individual gender inference when different classification algorithms and different kinds of features are leveraged

	NB	SVM	ME
BOW	0.785	0.855	0.861
BOW+F-measure	0.793	0.860	0.863
BOW+POS-patterns	0.796	0.863	0.864
All features	0.805	0.870	0.873

Table 7: Performance of interactive gender inference with joint inference when additional extra-task constraints are utilized

	Accuracy
Baseline	0.622
ILP(C1+C2+C3)	0.771
ILP(C1+C2+C3+C4)	0.797
ILP(C1+C2+C3+C5)	0.780
ILP(C1+C2+C3+C4+C5)	0.809

Table 7 shows the performance of our ILP approach (i.e. with joint inference) to interactive gender inference when additional extra-task constraints are utilized. From this table, we can see that when extra-task constraints, either C4 or C5, are leveraged, our ILP approach yields much better performance than the baseline. When both C4 and C5 constraints are used, an impressive improvement of 18.7% over baseline is obtained. Significance test show that our ILP approach with different types of constraints all significantly outperforms the baseline approach (p-value<0.01).

It is worthwhile to note that our ILP approach could also benefit the task of individual gender inference when extra-task constraints are employed. Table 8 shows the performance of individual gender inference where baseline means standard 2-category classification using the message text only. From this table, we can see that our ILP approach outperforms the baseline (0.873 vs. 0.890), even when the performance of the baseline is at a very high level. Furthermore, significance test shows that the improvement is significant (p-value<0.05). This indicates the usefulness of label correlations in both interactive and individual gender inference.

Table 8: Performance of our ILP approach to individual gender inference using extra-task constraints

Individual Gender Inference	Accuracy
Baseline	0.873
ILP Approach	0.890

6. Conclusion

In this paper, we investigate a novel task of gender inference, namely interactive gender inference, and address it with an ILP approach by fully incorporating the label correlations among instances. Specifically, label correlations are modeled as constraints in ILP in two scenarios: intra-task and extra-task. Empirical studies demonstrate that the ILP approach with intra-task and extra-task constrains performs significantly better than the baseline on interactive gender inference. Moreover, our ILP approach benefits the task of individual gender inference as well.

In our future work, we would like to explore more features, e.g. syntactic features, to improve the performance of interactive gender inference. Moreover, given the flexible framework of ILP approach, we would like to incorporate crucial gender information from other tasks, e.g. gender inference with social connections, to further enhance the gender inference performance.

Acknowledgments

This research work has been partially supported by three NSFC grants, No. 61273320, No.61375073, No.61331011, one NSF grant of Zhejiang Province No.LY13F020007, one National Social Science Foundation of China No. 14BTQ047, and Collaborative Innovation Center of Novel Software Technology and Industrialization.

References

[Alowibdi *et al*., 2013] Alowibdi J., U. Buy and P. Yu.2013. Language Independent Gender Classification on Twitter. *In Proceedings of 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 739-743.

[Burger *et al*., 2011] Burger J. and J. Henderson and G. Kim and G. Zarrella. 2011. Discriminating Gender on Twitter. *In Proceedings of EMNLP-11*, pp. 1301–1309.

[Corney *et al*., 2002] Corney M., O. Vel, A. Anderson and G. Mohay. 2002. Gender-Preferential Text Mining of E-mail Discourse. *In Proceedings of ACSAC-02*, pp. 282-289.

[Ciot *et al*., 2013] Ciot M., M. Sonderegger and D. Ruths. 2013. Gender Inference of Twitter Users in Non-English Contexts. *In Proceedings of EMNLP-13*, pp. 1136–1145.

[Gianfortoni *et al*., 2011] Gianfortoni P., D. Adamson and C. Rosé. 2011. Modeling of Stylistic Variation in Social Media with Stretchy Patterns. *In Proceedings of EMNLP-11*, pp. 49–59.

[Heylighen *et al*., 2002] Heylighen F., and Dewaele, J. 2002. Variation in the contextuality of language: an empirical measure. *In Proceedings of Foundations of Science*, 7, pp. 293–340.

[Mukherjee *et al*., 2010] Mukherjee A. and B. Liu. 2010. Improving Gender Classification of Blog Authors. *In Proceedings of EMNLP-10*, pp. 207-217.

[Miller *et al*., 2012] Miller Z., B. Dickinson and W. Hu. 2012. Gender Prediction on Twitter Using Stream Algorithms with N-Gram Character Features. *International Journal of Intelligence Science*, Vol. 2, No. 4, pp.143-148.

[Mohammad *et al*., 2011] Mohammad S. and T. Yang. 2011. Tracking Sentiment in Mail: How Genders Differ on Emotional Axes. *In Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis(2011)*, pp.70-79.

[Nowson *et al*., 2005] Nowson S., Oberlander J., Gill, A. J., 2005. Gender, Genres, and Individual Differences. *In Proceedings of the 27th annual meeting of the Cognitive Science Society*, pp. 1666–1671.

[Nowson *et al*., 2006] Nowson S. and J. Oberlander. 2006. The Identity of Bloggers: Openness and Gender in Personal Weblogs. *In Proceeding of AAAI-06*, pp. 163-167.

[Peersman *et al*., 2011] Peersman C., W. Daelemans, L. Van Vaerenbergh. 2011. Predicting Age and Gender in Online Social Networks. *In Proceedings of SMUC-11*, pp. 37-44.

[Schler *et al*., 2006] Schler J., M. Koppel, S. Argamon and J. Pennebaker. 2006. Effects of Age and Gender on Blogging. *In Proceedings of AAAI-06*, pp. 199-205.

[Volkova *et al*., 2013] Volkova S., T. Wilson and D. Yarowsky. 2013. Exploring Demographic Language Variations to Improve Multilingual Sentiment Analysis in Social Media. *In Proceedings of EMNLP-13*, pp. 1815–1827.

[Yan *et al*., 2006] Yan X. and L. Yan. 2006. Gender inference of Weblog Authors. *In Proceedings of AAAI-06*, pp. 228-230.

[Yang and Liu, 1999] Yang Y. and X. Liu. 1999. A Re-Examination of Text Categorization Methods. In *Proceedings of SIGIR-99*, pp. 42-49.