How Powerful are You? gSPIN: Bringing Power Analysis to Your Finger Tips

Vinodkumar Prabhakaran
Stanford University
Stanford, CA, USA
vinod@cs.stanford.edu

Michael Saltzman
Yahoo!
San Jose, CA, USA
mik854@gmail.com

Owen Rambow
CCLS, Columbia University
New York, NY, USA
rambow@ccls.columbia.edu

ABSTRACT

We present the SPIN system, a computational tool to detect linguistic and dialog structure patterns in a social interaction that reveal the underlying power relations between its participants. The SPIN system labels sentences in an interaction with their dialog acts (i.e., communicative intents), detects instances of overt display of power, and predicts social power relations between its participants. We also describe a Google Chrome browser extension, namely gSPIN, to illustrate an exciting use-case of the SPIN system, which will be demonstrated at the demo session during the conference.

Categories and Subject Descriptors

I.2 [**ARTIFICIAL INTELLIGENCE**]: Natural Language Processing—*Discourse, dialogue and pragmatics*

Keywords

content analysis; power relations; dialog; email

1. INTRODUCTION

Social interactions often reflect the social context in which they occur through linguistic and dialogic patterns [12]. By social context, we mean the power relations between participants, their status, gender, and so on. Recent years have seen a growing interest in computationally analyzing correlates of social power in interactions on a variety of genres, ranging from Supreme Court transcripts [2] to email interactions [1, 4] to Wikipedia discussion forums [11, 2, 3]. However, there hasn't been much work on investigating the practical applications of a computer system that can automatically detect correlates of power in day-to-day communications. In this paper, we present an end-to-end power-prediction system called SPIN (Social Power in INteractions), that performs deep NLP-based power analytics on email interactions. We also demonstrate an use case of the SPIN system through a Google Chrome browser extension, named gSPIN, that seamlessly integrates the power of SPIN analysis with Gmail interactions. We will demonstrate the gSPIN system at the demo session in the conference.

WWW'16 Companion, April 11–15, 2016, Montréal, Québec, Canada.
ACM 978-1-4503-4144-8/16/04.
http://dx.doi.org/10.1145/2872518.2890557.

2. SPIN SYSTEM

The SPIN system predicts power relations between pairs of participants in written interactions. Unlike prior approaches (e.g., [4]) that require access to all messages exchanged between a pair for power prediction, the SPIN system predicts power relations based solely on single interactions, i.e., email threads. It uses clues from dialog structure, derived from the interaction meta-data (e.g., message count, replies per message) as well as from NLP-based computational models of dialog acts [9] and from overt displays of power (e.g., impoliteness), in addition to bag-of-words features from the message content.

2.1 Processing Pipeline

Figure 1 shows the 4-step processing pipeline of the SPIN system. The input to the system is an interaction thread in an XML format. The output provides an XML file with the power relations graph, as well as the original interaction annotated with dialog intentions and overt displays of power. Each processing step is described below.

2.1.1 Basic NLP stack

In this step, the xml formatting of the interaction thread is parsed and basic NLP steps (tokenization, sentence splitting, lemmatization, and part of speech tagging) are applied to the messages. We use the ClearTk [5] wrappers to perform these steps.

2.1.2 DA tagging

In this step, we apply dialog act tagging [10], which models the dialog structure of an interaction. A dialog act tagger automatically assigns dialog intentions to each sentence in the conversation. Specifically, we use the dialog act tagger from [6], which assigns each sentence in the messages to be of one of the four dialog acts:

- REQUEST-ACTION (requesting actions),
- REQUEST-INFORMATION (requesting information),
- INFORM (providing information)
- CONVENTIONAL (greetings, sign offs, etc.).

Our dialog act tagger is an SVM-based supervised learning system, which uses lexical ngram features and thread structure features. It obtains an accuracy of 92.2% on 5-fold cross validation (experiments are described in detail in [6]).

2.1.3 ODP tagging

In this step, we automatically identify instances of "Overt Display of Power" (ODP) in the message content, a notion that we

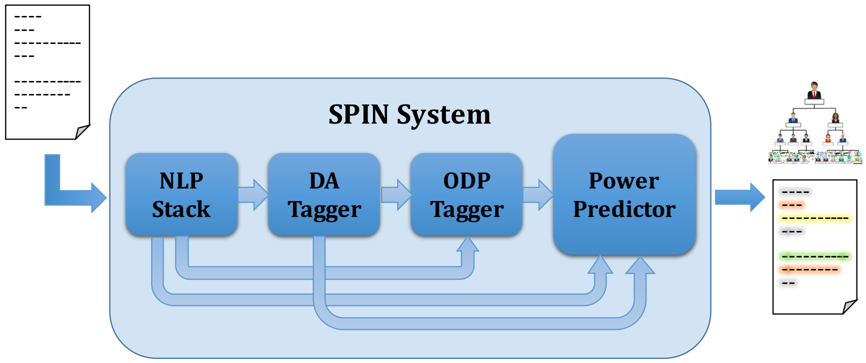

Figure 1: Social Power in INteractions (SPIN): Processing Pipeline
DA: Dialog Acts; ODP: Overt Displays of Power

introduced in [8]. We define an utterance to have ODP if it is interpreted as creating additional constraints on its response beyond those imposed by the general dialog act. For example, the dialog act of REQUEST-ACTION introduces the constraint that there be a response, but one could use linguistic forms to suggest that the response be one of a limited number of options. In effect, ODP captures the difference between the requests in the following two sentences:

- *I need your report by Friday*
- *Could you please try to send your report by Friday?*

Both sentences invoke the same request, but the former is an instance of overt display of power.

The ODP tagger (described in detail in [8]) is a binary SVM classifier which uses the lexical ngram features and dialog act features obtained from the prior step in order to make the prediction. The trained classifier model obtains an F-score of 54.2 (compared to a baseline of 10.4 if predictions were made at random) on 5-fold cross validation.

2.1.4 *Power prediction*

In the final step, we apply an automatic power prediction system that uses the lexical features, dialog act features, and ODP features in order to predict power relations between pairs of participants who interact within the given email thread. For each interacting participant pair, the system predicts the relationship between the first person and second person to be either *subordinate* or *superior*.

The power prediction system is trained on the Enron email corpus using the organizational hierarchy information. The corpus contains 36,196 threads in total, in which there are 15,058 interacting participant pairs that had a power relation. The system uses seven different feature subsets: positional (PST), verbosity (VRB), thread structure (THR), dialog acts (DA), dialog links (DL), overt displays of power (ODP), and word and part-of-speech ngrams (LEX). All the individual features and different experiments using different feature combinations are described in detail in [7]. Table 1 presents the results obtained by the system using different feature combinations. We use a majority class baseline assigning the first person to be always *superior*, which obtains 52.5% accuracy. We also use a stronger baseline using word unigrams and bigrams as features, which has an accuracy of 68.6%. The highest accuracy obtained without using any message content was 61.5%. On adding dialog act and overt display of power features, the accuracy improved to 62.5%. LEX features by itself obtain a very high accuracy of 70.7%, confirming the importance of lexical patterns in this task. The best performing system (BEST) uses a combination of lexical and structural features and obtains an accuracy of 73.0%.

Description	Accuracy
Baseline (Always Superior)	52.54
Baseline (Word Unigrams + Bigrams)	68.56
PST + VRB + THR	61.49
PST + VRB + THR + DA + DL + ODP	62.47
LEX	70.74
BEST combination	**73.03**

Table 1: Results on Power prediction.

3. gSPIN: A BROWSER PLUGIN

In order to demonstrate the utility of such a power analysis system for an end user, we created a Google Chrome browser extension, called gSPIN, that seamlessly integrates the SPIN analysis with Gmail. The gSPIN plugin enables users to apply SPIN analysis to their email threads and makes the power prediction results as well as the lower-level dialog act and ODP analyses available to them. It uses the Google Chrome Identity API in conjunction with the Gmail REST API to securely access and process the user's email threads.

3.1 Process Flow

The system architecture and the process flow of the gSPIN system are shown in Figure 2. Dotted arrows indicate communications that do not involve email content; solid arrows indicate email content being transferred. We describe below each step of the process, starting from the user initiating the gSPIN request.

1. User requests gSPIN analysis of the Gmail thread that is displayed in the Chrome browser.
2. gSPIN uses the Google Chrome Identity API to prompt the user to provide their mail credentials and give consent for gSPIN to read the email thread.
3. Google Chrome Identity API grants a read only oauth token after verifying the user's credentials.

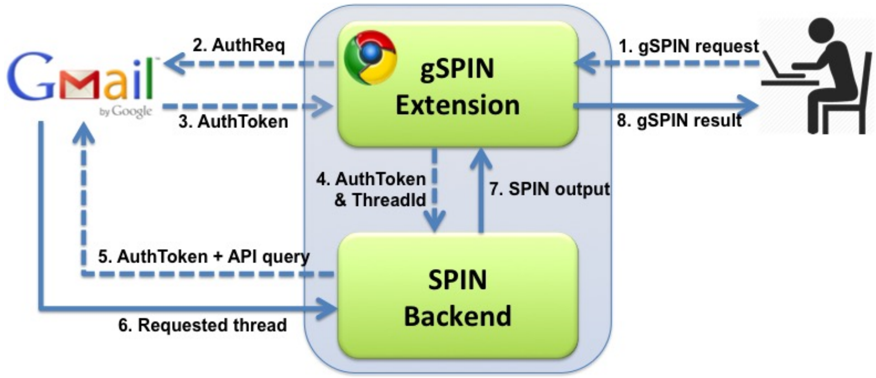

Figure 2: gSPIN plugin: process flow and system architecture.
Dotted arrows indicate communications without email content.
Solid arrows indicate email content being transferred

4. gSPIN sends the oauth token along with the thread identifier to the remote SPIN Backend server.

5. The SPIN Backend server requests the email thread contents using the oauth token, via the Gmail API.

6. The Gmail API returns the content of the requested email thread, which is then parsed into the proper format for SPIN analysis.

7. The SPIN system processes the email thread as per the processing pipeline shown in Figure 1 and returns the output to gSPIN.

8. gSPIN unpacks the SPIN output and displays it to the user in a pop-up window.

Steps 2 and 3 are performed only for the initial request or when an already obtained authentication token has expired. Tokens will typically last longer than the duration of a user's email interactions. In the current implementation, the SPIN Backend is running remotely

3.2 Sample Run

We demonstrate the functionality of gSPIN using a simulated email conversation using Gmail between two authors of this paper — Michael and Vinod. The email conversation is about the status of the SPIN system development, in which Vinod asks Michael about the status and instructs him to write a report by a certain date. The screen shot shown in Figure 3 shows the output produced by the gSPIN analysis. The first section of the output shows the power relations detected between pairs of interacting participants. The gSPIN system finds Vinod has power over Michael, based on the conversation in this thread. The second section of the output displays the original email thread marked with dialog acts and overt displays of power. The gSPIN system tags the sentence *The deadline for SPIN demo is March 1st* as an INFORM dialog act, where as *Is there any updates on putting together the system?* to be REQUEST-INFORMATION, and *Please write the report by Feb 28th!* as REQUEST-ACTION. In addition, the gSPIN also highlights the sentence *Please write the report by Feb 28th!* as an instance of overt display of power. The lower level dialog analysis results itself provides great benefits for the user, for example, it highlights parts of conversation that require immediate attention.

3.3 Discussion

The main objective of gSPIN system is to demonstrate the use case of a power analysis on user interactions, and hence is not designed to operate at scale. In order to perform power analysis at scale, the two main challenges are hardware limitations and security. The SPIN analysis is computationally intensive and it will require more dedicated resources to handle the potential load to the server/cluster. In terms of security, we will have to enforce stricter security precautions on the communication channels in order to gain the trust from the user to grant the SPIN server an authorization token to safely manage their emails.

Our aim was also to create an implementation that would be easily accessible to as many people as possible, using an already existing secure platform. Our choice of the Gmail – Google Chrome framework for this demonstration system is motivated by its wide reach and the secure platform it provides. Gmail is one of the popular email provider with an easy-to-use API combined with a Java library on top of that to provide easy manipulation of the email threads. In addition, Google Chrome offers the Identity API to facilitate user authentication through the browser. With the Identity API, all password authentication is handled safely and securely by the browser itself, and grants the user an oauth token to safely access email threads from the Gmail server. The oauth token can further be limited to a read-only token, with limited TTL, allowing the user greater peace of mind.

4. CONCLUSION

We presented the SPIN system which brings the power of deep NLP analytics to the end user to analyze his/her own social interactions. We also presented the gSPIN plugin, which illustrates how active research in NLP on the socio-pragmatics of dialog can be incorporated into an existing email environment. Using this plugin, a Gmail user can analyze his/her email conversations to infer the underlying power dynamics. A gSPIN version is available to download from the Google Chrome extensions store, and we will demonstrate it at the demo session in the conference.

In the future, we plan to expand the gSPIN functionality to analyze email drafts, so that users can reconfirm that their communication intentions are correctly represented, and that there are no inadvertent overt displays of power. We also plan to extend the SPIN analysis to other public email service providers (such as Yahoo! Mail), which provide a secure API to access their email databases.

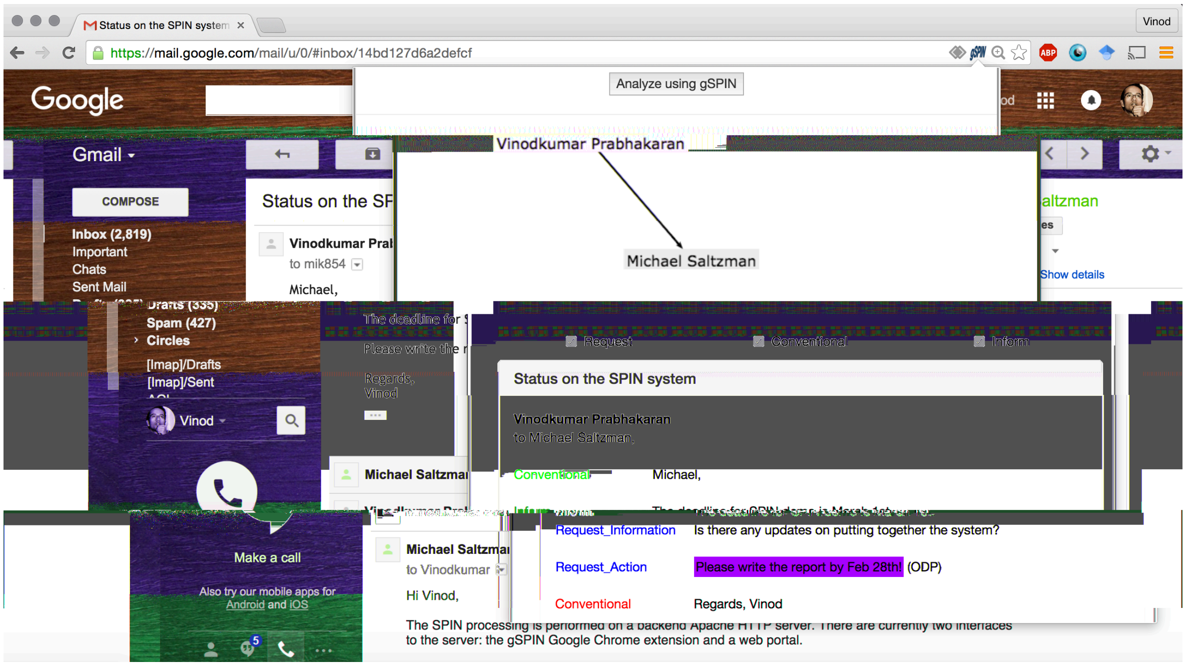

Figure 3: gSPIN at Work: output display of SPIN power analysis.
A simulated email conversation between two of the authors of this paper — Vinodkumar Prabhakaran and Michael Saltzman.
The SPIN system found Vinodkumar Prabhakaran to have power over Michael Saltzman in this fictional email conversation.

5. ACKNOWLEDGMENTS

This paper is based upon work supported by the DARPA DEFT Program. The views expressed are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government. Vinodkumar Prabhakaran and Michael Saltzman were affiliated with Columbia University during the time when this work was performed.

6. REFERENCES

[1] P. Bramsen, M. Escobar-Molano, A. Patel, and R. Alonso. Extracting social power relationships from natural language. In *ACL*, pages 773–782. The Association for Computational Linguistics, 2011.

[2] C. Danescu-Niculescu-Mizil, L. Lee, B. Pang, and J. Kleinberg. Echoes of power: language effects and power differences in social interaction. In *Proceedings of the 21st international conference on World Wide Web*, WWW '12, New York, NY, USA, 2012. ACM.

[3] C. Danescu-Niculescu-Mizil, M. Sudhof, D. Jurafsky, J. Leskovec, and C. Potts. A computational approach to politeness with application to social factors. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 250–259, Sofia, Bulgaria, August 2013. Association for Computational Linguistics.

[4] E. Gilbert. Phrases that signal workplace hierarchy. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, CSCW '12, pages 1037–1046, New York, NY, USA, 2012. ACM.

[5] P. V. Ogren, P. G. Wetzler, and S. Bethard. ClearTK: A UIMA toolkit for statistical natural language processing. In *Towards Enhanced Interoperability for Large HLT Systems: UIMA for NLP workshop at Language Resources and Evaluation Conference (LREC)*, 2008.

[6] A. Omuya, V. Prabhakaran, and O. Rambow. Improving the quality of minority class identification in dialog act tagging. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 802–807, Atlanta, Georgia, June 2013. Association for Computational Linguistics.

[7] V. Prabhakaran and O. Rambow. Predicting power relations between participants in written dialog from a single thread. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 339–344, Baltimore, Maryland, June 2014. Association for Computational Linguistics.

[8] V. Prabhakaran, O. Rambow, and M. Diab. Predicting overt display of power in written dialogs. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 518–522, Montréal, Canada, June 2012. Association for Computational Linguistics.

[9] J. R. Searle. *Speech acts: An essay in the philosophy of language*, volume 626. Cambridge university press, 1969.

[10] A. Stolcke, N. Coccaro, R. Bates, P. Taylor, C. Van Ess-Dykema, K. Ries, E. Shriberg, D. Jurafsky, R. Martin, and M. Meteer. Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech. *Computational linguistics*, 26(3):339–373, 2000.

[11] T. Strzalkowski, G. A. Broadwell, J. Stromer-Galley, S. Shaikh, S. Taylor, and N. Webb. Modeling socio-cultural phenomena in discourse. In *Proceedings of the 23rd International Conference on COLING 2010*, Beijing, China, August 2010. Coling 2010 Organizing Committee.

[12] T. A. Van Dijk. Structures of discourse and structures of power. In *Communication yearbook*, volume 12. 1989.