Analyzing Miscommunications.

Ravi Kiran Holur Vijay **Peter Thermos** **Henning Schulzrinne**

Computer Science Department, Columbia University, NY.

Abstract.

We explore some metrics that might possibly indicate occurrence of possible miscommunication during a task-oriented conversation. The metrics explored are motivated by real-world observations and cover lexical, structural and semantic attributes. We develop a test bed for evaluating the metrics and execute the metrics on a task-oriented corpus. Also, efforts were made to produce a task-oriented speech corpus in a multi-speaker and multi-cultural setting.

1. Introduction.

If we imagine ourselves as participants in a task-oriented conversation, we can see that miscommunications indeed do happen [1]. Further, if we think about the miscommunication detection and repair process adopted by humans, we can observe that it occurs at various logical levels - word, sentence and dialog. Based on some real world observations and previous works [1][2][3][4], we will explore some metrics corresponding to lexical, syntactic and semantic levels. At lexical level, we can calculate a metric that could be used to quantify the ambiguity of each word (both dependent and independent of context). At the structural level, we can calculate a metric based on syntactic priming [3][4]. At the semantic level, we can calculate a metric based on the sentimental polarity (positive or negative) of the sentence. In order to evaluate the metrics, we run the implementations of these on a task-oriented corpus [5].

Also, efforts were made to produce a task-oriented speech corpus in a multi-cultural and multi-speaker setting. For this, we used the game Counter Strike[6] to simulate a virtual world where players could communicate. The players involved in this exercise were from diverse cultural backgrounds. Also, English was not the native language in most of the cases.

2. Metrics.

2.1 Lexical level.

Word ambiguity induced by context of use can happen when a word can have multiple interpretations based on the context in which it is used. Here, the required context is provided by the sentence in which the word appears. For quantifying this ambiguity, we use two metrics – Score output of WSD algorithms and SD of sense frequencies from WordNet (SensesSD) [7]. For the first metric, we try out a couple of WSD algorithms [8] [9]. Both of these algorithms take a context and a target word as input and give the most probable sense for the target word as output, using WordNet as the dictionary. But, we need a score using which we could determine how probable the best probable sense was. For this purpose, we take the score as the number of overlaps of the sense with maximum overlaps and normalize it. For the second metric, we calculate the standard deviation of the frequencies of all the possible senses of a given word, using WordNet. After implementing and executing the above metrics on a task-oriented corpus, the following observations were made. If the score is less than 0 or more than 1 (like really, sorry, nice, left, stop, right, back, there etc.), we can see that the words might lead to potential misunderstanding more readily. Now, if the score is between 0 and 1, we need to look at SensesSD to decide if the word might lead to potential misunderstanding. Some words (here, there, work, fire, want, think etc.) are in this category, with low scores (less than .15) and high SensesSD (greater than 50). So, basically we need to identify words with low scores and high SensesSD, which might give us an approximate measure for the misunderstanding it can cause in the given context (sentence). A couple of graphs of Scores vs. No. of words is shown in *[Figure 1]* & *[Figure 2]*.

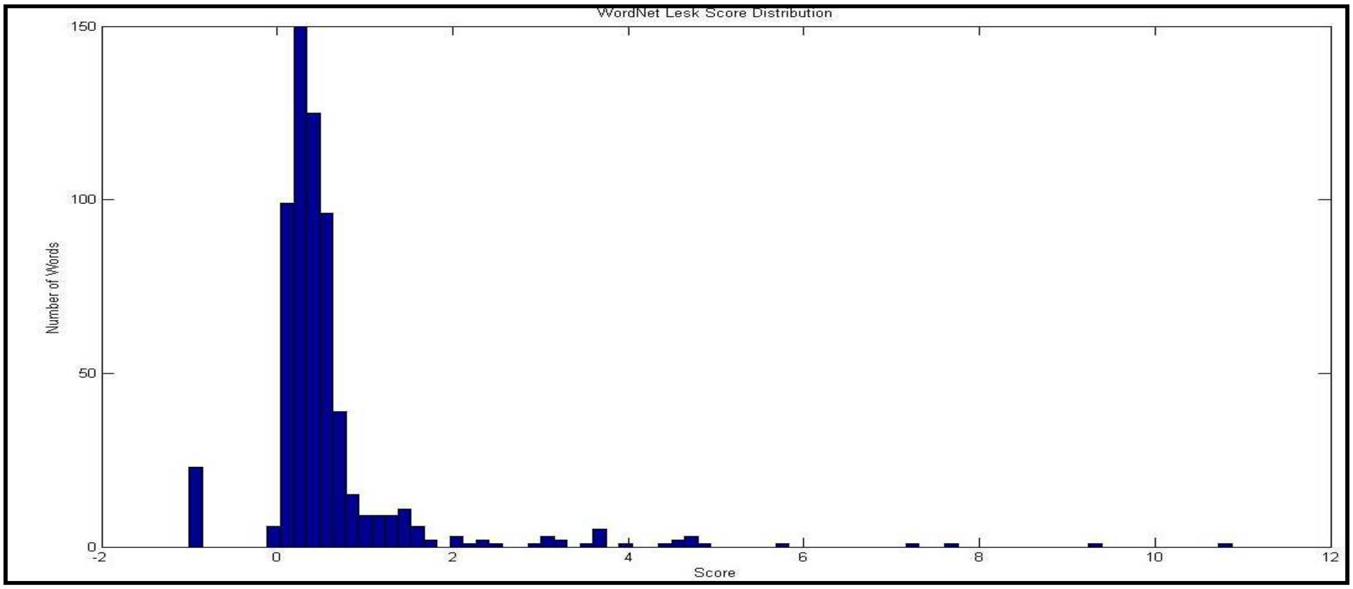

Figure 1 - Distribution of Scores for WordNet Lesk WSD algorithm - I

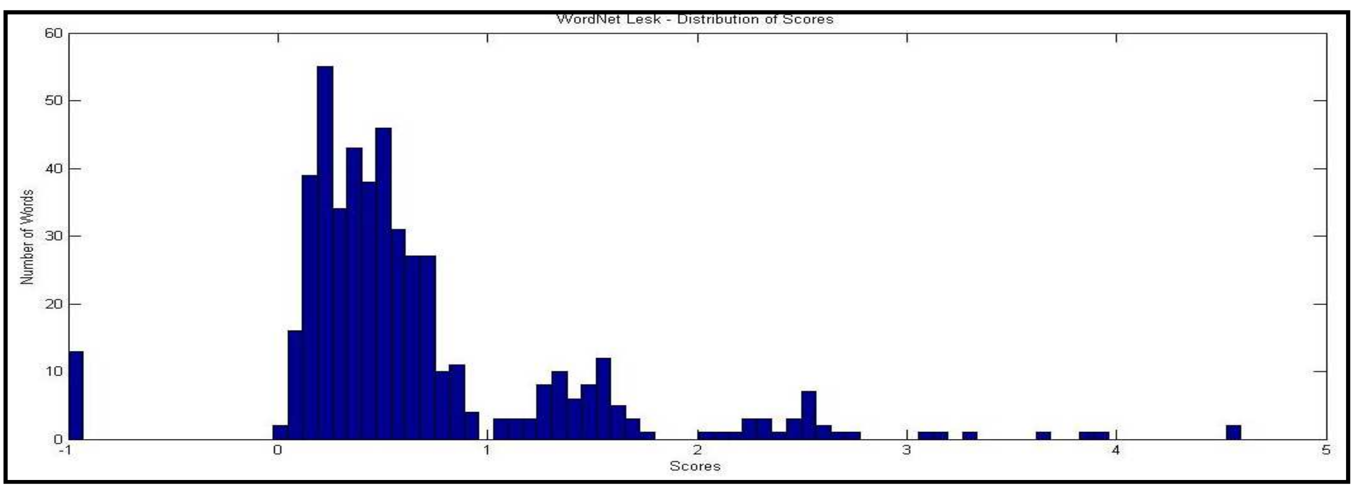

Figure 2 - Distribution of Scores for WordNet Lesk WSD algorithm - II

2.2 Structural level.

At the sentence level, we explore a metric based on Syntax priming [3][4]. We can use the metric to predict the task outcome using a regression model (SVM based regression) and thus evaluate how effective it is for predicting task success. The architecture is as shown in *[Figure 3]*.

The implementation was executed on the Map Task Corpus [5]. The task success metric used in Map Task corpus was the attribute to be predicted by the regression model. The performance of the evaluation is shown in *[Figure 4]*, the average error rate being -12.7622217.

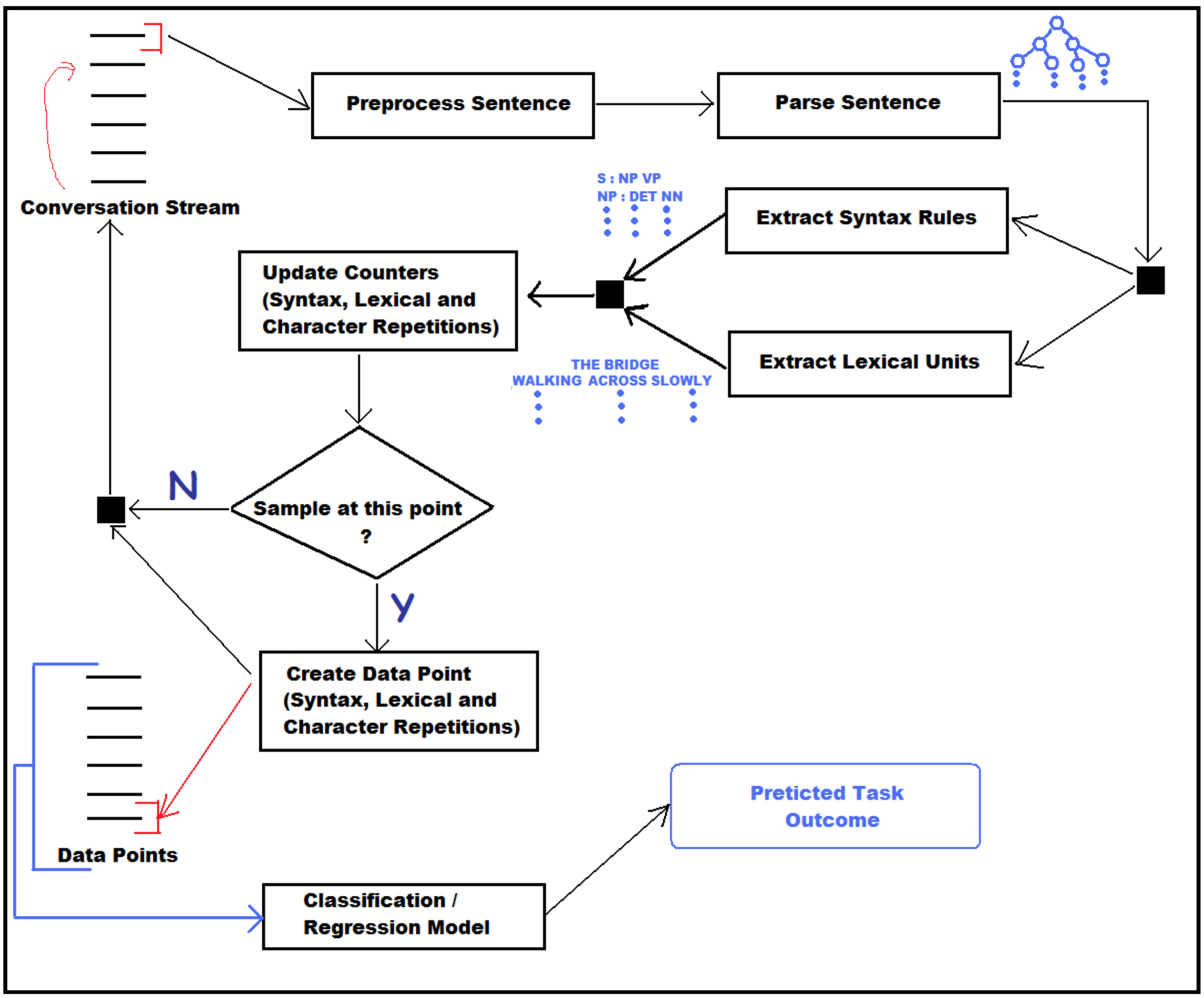

Figure 3 - Architecture for implementation of Structural level metrics.

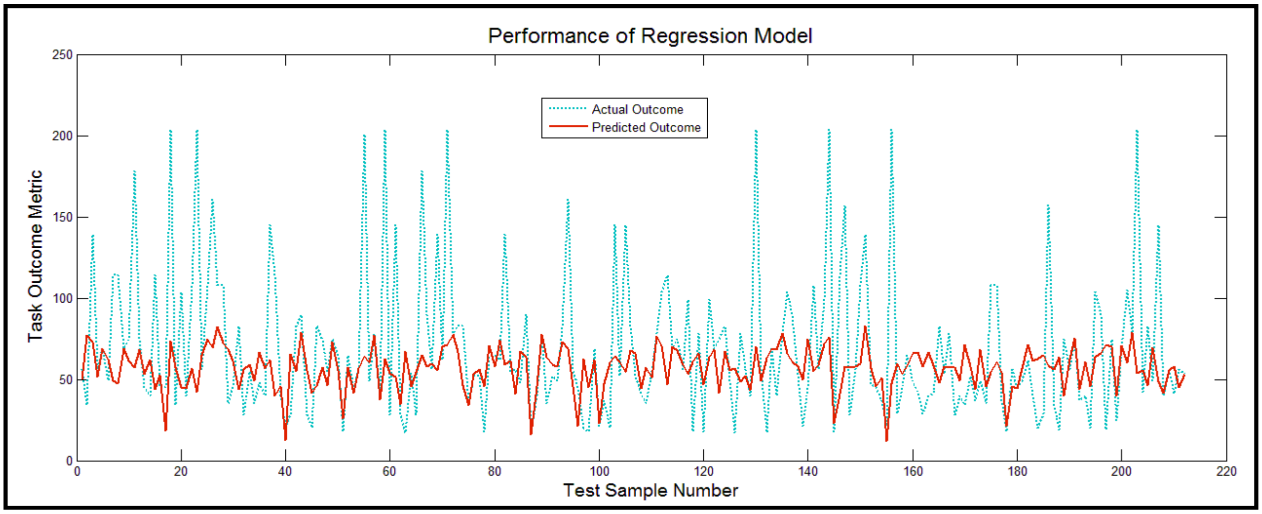

Figure 4 - Performance of Structural Priming based metrics.

2.3 Semantic level.

At the semantic level, we explore a metric related to sentiment (positive, negative and neutral) of the sentence. We define the sentiment of a sentence as the sum of the sentiment of the adjectives that the sentence contains. The sentiments of the adjectives were inferred using SentiWordNet[11]. When the implementation was executed on the Map Task corpus, the results obtained are as shown in *[Figure 5]*.

The correlation coefficients between Task metric and No. of Positive, Negative and Neutral sentences were -0.07, -0.007 and -0.1023 respectively.

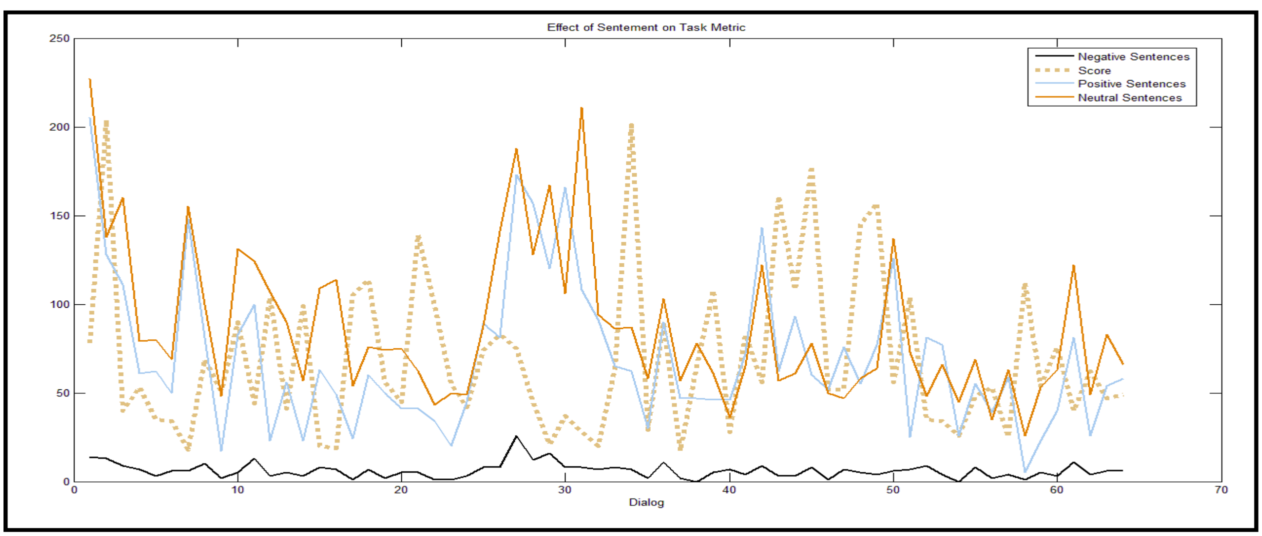

Figure 5 - Effect of Sentiment on Task Metric.

3. Task-oriented spoken dialog corpus.

In order to continue further explorations of the problem, efforts were made to produce a task-oriented spoken dialog corpus in a multi-cultural setting. For this, we used the popular FPS “Counter Strike” [6]. There were two teams, with approximately two to four players on each team. Each session consisted of games where teams of human players played against each other and where teams of human players played against the computer bots. The in-game conversations among players and the final scores were recorded for each of the games. The next step would be to transcribe the audios of each of the games into text, either manually or in a semi-supervised manner. A couple of screenshots from the game, along with the metrics used to measure the team’s success is shown in *[Figures 6, 7]*.

Figure 6 - In-Game world and Success Metrics I

Figure 7 - In-Game world and Success Metrics II

4. References.

[1] Words Are Mightier Than Swords … and Yet Miscommunication Costs Lives! - Stephen Poteet, Jitu Patel, Cheryl Giammanco.

[2] Fatal Words: Communication Clashes and Aircraft Crashes - Steven Cushing.

[3] TOWARD A MECHANISTIC PSYCHOLOGY OF DIALOGUE - Martin J. Pickering, Simon Garrod.

[4] Priming of Syntactic Rules in Task-Oriented Dialogue and Spontaneous Conversation - David Reitter, Johanna D. Moore, Frank Keller.

[5] Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G. M., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S. and Weinert, R. (1991). The HCRC Map Task Corpus. Language and Speech, 34, pp. 351-366.

[6] Counter Strike - http://store.steampowered.com/app/240/

[7] C. Fellbaum, editor. WordNet: An electronic lexical database. MITPress, 1998.

[8] Evaluating variants of the Lesk Approach for Disambiguating Words” by Florentina Vasilescu, Philippe Langlais and Guy Lapalme.

[9] Adapting the Lesk Algorithm for Word Sense Disambiguation to WordNet by Satanjeev Banerjee.

[10] Stanford Parser - http://nlp.stanford.edu/software/lex-parser.shtml

[11] Andrea Esuli and Fabrizio Sebastiani. SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. In Proceedings of LREC-06, 5th Conference on Language Resources and Evaluation, Genova, IT, 2006, pp. 417-422.

[12] Ian H. Witten and Eibe Frank (2005) "Data Mining: Practical machine learning tools and techniques", 2nd Edition, Morgan Kaufmann, San Francisco, 2005.