13th ICCRTS: C2 for Complex Endeavors

"Semantical Machine Understanding"

Topic 9: Collaborative Technologies

Ying Zhao, Chetan Kotak and Charles C. Zhou

Point of Contact: Charles C. Zhou

Quantum Intelligence, Inc.

3375 Scott Blvd, Suite 100 Santa Clara, CA 95054

408-203-8325

charles.zhou@quantumii.com

Semantical Machine Understanding

Abstract

Semantical Machine Understanding is the foundation for automatic sense and decision making of multinational, multicultural, and coalition applications. We show an innovative semantical machine understanding system that can be installed on each node of a network and used as a semantic search engine. Innovations of such a system include 1) text mining: extract concepts and meaning clusters based on contexts using pattern recognition and machine learning; 2) meaning learning: extract knowledge patterns that link human labeled meaning to raw data. The knowledge patterns can be applied to predict future data; and 3) collaborative meaning search: incorporate humans and machines to form a collaborative network to search and enhance the meaning iteratively.

In this paper, we also show the feasibility of using a semantic search architecture and discuss the two ways it is drastically different from current search engines: 1) indexes embedded in agents are distributed and customized to the learning and knowledge patterns of their own environment and culture. This allows data providers to maintain their own data in their own environment, but still share indexes across peers; 2) Semantic machine understanding enables discovery of new information rather than popular information.

Keywords: semantical machine understanding, text mining, decision making, sense making, semantic search, machine learning, sentiment classification, distributed search indexes

Background

Defense transformation has changed warfighting tactics from use of platform-based large-scale initiatives to quick reaction, team-based mobile force operations in discrete events. There are increased operations with joint, coalition, non-Government and volunteer organizations that require analysis of open-source (uncertain, conflicting, partial, non-official) data. Teams consist of culturally diverse partners with rapidly changing team members and various organizational structures. These characteristics put increasingly difficult demands on short turn-around, high stakes, crisis driven intelligence analysis. In order to respond to this challenge, more powerful information analysis tools are needed that can quickly extract meaning and intent from large volumes of data. There are a number of extant tools for data mining, including advanced search engines [1, 2] and key word analysis and tagging technologies [1], but better tools are needed to achieve advanced information discovery which provide more focused and directed content rather than line-item search results. The key to such a capability is the automated understanding of intent or meaning and the ability to represent it in a language/culture free format. Needed is a data/text analysis tool that can perform semantic search and provide language/culture-free information in a format that can be used for discovery of events, relationships and trends.

Semantical machine understanding is a very challenging task. Solutions have to be language/culture-free which mean it is better not to use linguistic based approaches. Commercial available tools for text analysis such as entity extractions are mostly based on linguistic based models to identify entities. This is also related to advanced search engines for information search and retrieval. Since meaning is often associated with human definition, what's really needed is an infrastructure incorporating human interactions in the loop to gradually enhance machine understanding (of human). An automatic mechanism is needed not only for data collection, but also for machine learning to reinforce its understanding when it encounters human interaction. However, because the involved parties can be distributed, culturally diverse partners with rapidly changing team members and various organizational structures, it is difficult to assume any meaning can be static and in a centralized location. Therefore, a peer-based infrastructure is needed. It is increasingly interesting to apply peer-to-peer (P2P) technologies to store, locate and understand information with agent-like applications distributed among a grid of computers. Each application is considered itself as a peer or node among a network of similar applications. This infrastructure allows the network to be "fault-tolerate", "distributed", and "self-scalable". With all the great advantages of a P2P concept, the current architectures lack the technology to learn from experience or human interactions.

Objectives

The objective of this paper is to demonstrate a semantical machine understanding system using three data sets (NEO transcriptions from NAVAIR, Katrina Blogs, and sentiment reviews from web) and two use case areas (decision making and sense making).

In our approach, we make an assumption that samples of historical intelligence analysis data are available which include the following:

- Observations: free-text, open vocabulary sentences
- Meaning: corresponding meaning of observations made by human analysts using keywords or free-text, open vocabulary sentences

Semantical machine understanding is achieved by combining innovations in text mining, meaning learning and collaborative search as shown in Figure 1. Text categorization extracts concepts and meaning clusters from free text input based on contexts. Meaning learning discovers knowledge patterns that link human labeled meaning to raw data. The knowledge patterns are applied to predict the meaning of new data. Collaborative meaning search incorporates human and machines in a loop to form a collaborative network in order to search and enhance the meaning iteratively.

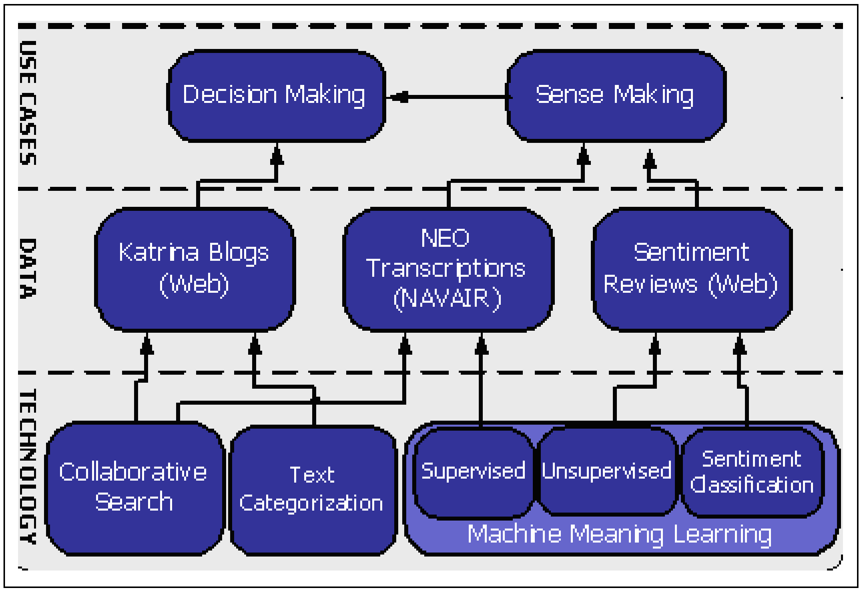


Figure 1. Semantical Machine Understanding

Technology

Text Mining

A text categorization and concept extraction mining technique – context-concept-cluster (ccc) model (US Patent Pending) – is shown here. The advantage of such a text mining technique over traditional information retrieval [3] is the ability to capture the cognitive level of understanding of text observations using only a few concepts.

Machine Meaning Learning

The process of learning the meaning from human labeled data (e.g. supervised learning) is illustrated in Figure 2. A train set of sentences with both observations and their labeled meaning are presented to a machine learning system. The system first generates a text categorization model that groups the sentences into categories by similarity. The system then generates a correlation model between the categories and the "real meaning" assigned by human analysts. The system also leaves out a data set for testing and evaluating. The test set is fed into the same model and a meaning is predicted for each sentence in the test set. The predicted meaning is then compared with the real meaning to evaluate the accuracy.

Collaborative Meaning Search

We show a collaborative meaning search to further improve meaning prediction. Each agent (either human or machine) generates its own meaning model of assigning (predicting) a meaning to the raw information observation. Each agent also holds a peer list showing how an agent is socially connected with other agents. The true meaning of a piece of information is the combination of predictions from an agent's own meaning model and the meanings from agents it is connected to in the social network.

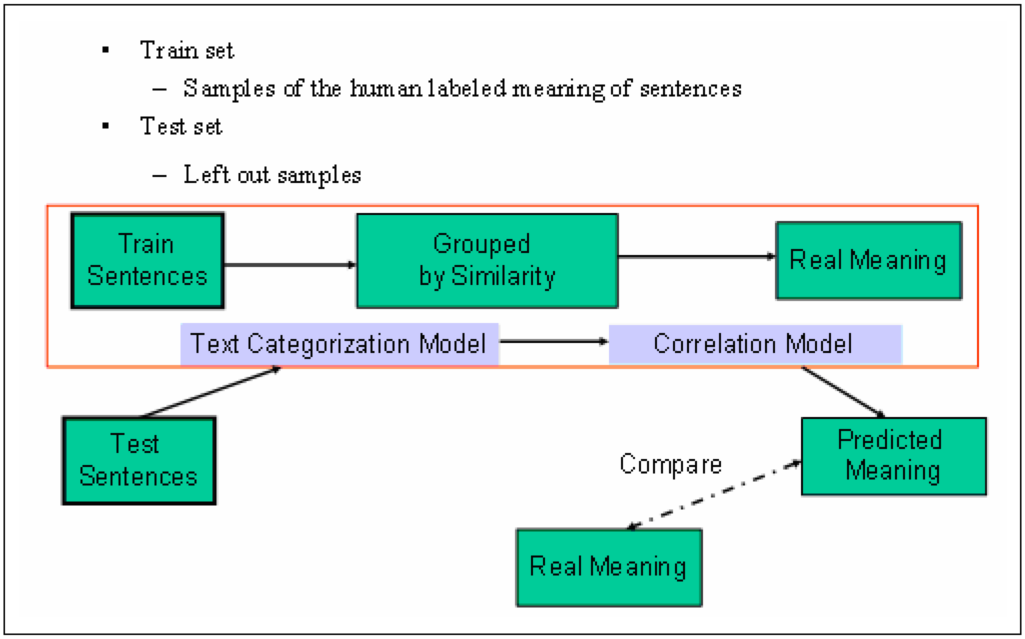


Figure 2. The process of meaning learning and prediction.

Use Cases

Sense Making Using NEO Transcription Data

One of the more studied scenarios for team collaborations is the Red Cross NEO (Noncombatant Evacuation Operation) scenario. Here, a team of experts in weapons, environment and intelligence work together to develop a course of action using assets available to rescue some Red Cross workers trapped in the middle of guerilla warfare on a remote island. Human analysts provide the optimal solutions and a scoring matrix is provided to evaluate alternative solutions. We were able to get three face-to-face NEO scenario team problem solving transcripts FS-2, FS-3, and FS-4 from NAVAIR. Figure 3 shows an example of a team problem solving transcript. The text observations (sentences) are the communications and conversations that were recorded during the team problem solving session. The meanings are defined as macrocognitive stages and states (processes)[22]. The stages are the human labels indicting if a piece of conversation can be categorized to a psychological stage such as "KC - Knowledge construction" or "TPS - Team problem solving". Examples of the states (processes) include:

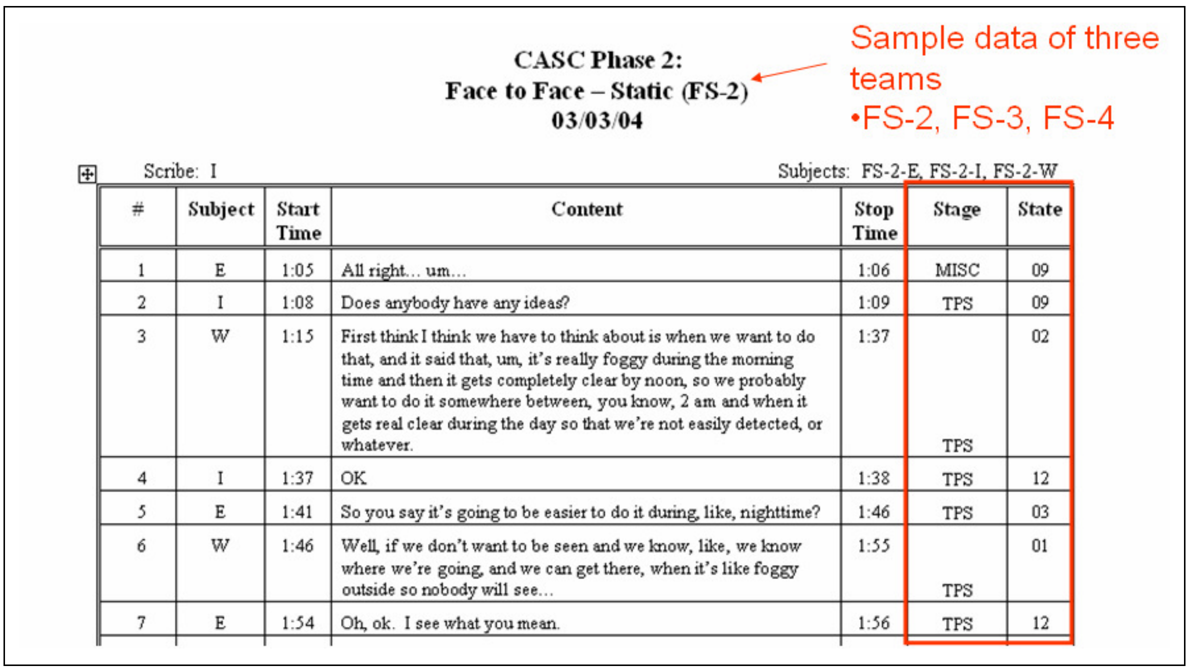
- 3: Macrocognition itk: individual task knowledge development
- 9: Macrocognition ica: iterative information collection and analysis
- 8: Macrocognition kio: knowledge interoperability development
- 12: Macrocognition cmm: convergence of individual mental models to team mental model
- 4: Macrocognition tk: team knowledge development

- 2: Macrocognition imm: individual mental model construction
- 10: Macrocognition tsu: team shared understanding development
- 6: Macrocognition vrm: individual visualization and representation of meaning

As one can see, the meaning "stages" or "states" may not be directly associated with what's been said in the transcripts, but rather the representation of the hidden intention a team member might try to convey during the team problem solving process. Important questions that psychologists try to answers are:

- Can these stages and states (processes) be predicted from what collectable measures such as transcripts and body languages?
- How can processes be tracked and identified automatically?

The primary information that is used for prediction is the "Content" field in the transcripts.



CASC Phase 2:
Face to Face – Static (FS-2)
03/03/04

Scribe: I

Subjects: FS-2-E, FS-2-I, FS-2-W

#	Subject	Start Time	Content	Stop Time	Stage	State
1	E	1:05	All right… um…	1:06	MISC	09
2	I	1:08	Does anybody have any ideas?	1:09	TPS	09
3	W	1:15	First think I think we have to think about is when we want to do that, and it said that, um, it's really foggy during the morning time and then it gets completely clear by noon, so we probably want to do it somewhere between, you know, 2 am and when it gets real clear during the day so that we're not easily detected, or whatever.	1:37	TPS	02
4	I	1:37	OK	1:38	TPS	12
5	E	1:41	So you say it's going to be easier to do it during, like, nighttime?	1:46	TPS	03
6	W	1:46	Well, if we don't want to be seen and we know, like, we know where we're going, and we can get there, when it's like foggy outside so nobody will see…	1:55	TPS	01
7	E	1:54	Oh, ok. I see what you mean.	1:56	TPS	12

Figure 3. NEO Transcription data from NAVAIR.

In the field "Content", human analysts label the conversations and some also label the characteristics of the communications including body languages. For example, if a person is pointing to the map or writing notes, that communication is noted as "points to the map" and "writes notes" in brackets. We were able to extract this body language communication and use the features as additional input dimensions for prediction. We also applied "Part of Speech (POS) Tagging" with a tool from "Stanford Log-linear Part-Of-Speech Tagger". This information was also used as additional input dimensions for prediction.

We explored three different settings for learning and predicting the meaning of sentences. Following settings were used to generate the three predictive models:

- Setting 1: Use content only
- Setting 2: Use content and features (body languages, questions, statements, etc.)
- Setting 3: Use content, features and previous states

The first setting result is shown in Table 1. The train set is half the data from FS-3 and the test set is remaining half. Table 1 shows only the test set accuracies for the states that are higher than 40% and the overall accuracy includes all the states. As one can see from Table 1, the content is highly correlated (80%) with State 12, i.e. convergence of individual mental models to team metal model. Also, the overall accuracy is improved as more input dimensions are added for prediction.

The second setting is shown in Table 2. The train set is all the data from FS-3 and the test set is data from FS-2. Table 2 shows the discrepancy of the two teams (FS-2 and FS-3) is reflected in the prediction accuracy such as State 12.

Table 1: State Prediction Test Result 1

Setting 1:Train FS-3; Test FS-3	Content	Content Features	Content Features Previous States
Overall	**30%**	**35%**	**49%**
State 3: individual task knowledge development		44%	
State 12: convergence of individual mental models to team mental model	80%	85%	58%
State 6: individual visualization and representation of meaning	58%	46%	
State 4: team knowledge development			71%
State 10: team shared understanding development			80%
State 8: knowledge interoperability development	48%	45%	48%
State 9: iterative information collection and analysis			40%

The third setting is shown in Table 3. Data from both FS-2 and FS-3 were used as train sets and FS-4 was used as the test set. Table 3 shows that adding diversified teams as train sets helps to compensate for the difference among the teams, therefore, improveing the overall accuracy as well as some specific states. Previous states are very helpful for predicting State 4 (team knowledge development) and State 10 (team shared understanding development), indicating a team might stay in team knowledge or understanding development state for a while to allow team members to take turn to express their views during a problem solving process.

Table 2: State Prediction Test Result 2

Setting 2: Train FS-3; Test FS-2	Content	Content Features	Content Features Previous States
Overall	**20%**	**31%**	**48%**
State 3: individual task knowledge development		44%	69%
State 12: convergence of individual mental models to team mental model	82%	40%	42%
State 6: individual visualization and representation of meaning	50%	57%	57%
State 4: team knowledge development			76%
State 10: team shared understanding development			66%
State 8: knowledge interoperability development		48%	
State 9: iterative information collection and analysis			

Table 3: State Prediction Test Result 3

Setting 3: Train FS-2 and FS-3; Test FS-4	Content	Content Features	Content Features Previous States
Overall	**30%**	**30%**	**50%**
State 3: individual task knowledge development	72%	40%	81%
State 12: convergence of individual mental models to team mental model	67%	63%	41%
State 6: individual visualization and representation of meaning	50%		50%
State 4: team knowledge development			76%
State 10: team shared understanding development			85%
State 8: knowledge interoperability development			
State 9: iterative information collection and analysis			44%

Table 4 shows the result for predicting the stages, where four models (content, content + features + previous stages, content + features + previous stage prediction, content POS tags) are generated using half of the data from FS-2 as training and the other half for testing. FS-2 is used because the labels are done with a higher consistency and therefore POS is more accurate. We found that POS and previous states are helpful for predicting stages. Also, since previous stages are helpful in the prediction for both states and stages, we used the iterative approach here where predicted stages are used as surrogates for the real previous stages in the prediction. As one can see from Table 4 the predicted stages do not perform as good as the real ones, nevertheless, it improves the accuracy from the one without it.

Table 4: Stage Prediction Result

Setting 1:Train FS-2; Test FS-2	Content	Content Features Previous Stages	Content Features Previous Stages Prediction	Content POS Tags
Overall	**61%**	**74%**	**61%**	**65%**
KC	6%	60%	38%	44%
TPS	91%	80%	72%	80%

From these experiments, the overall accuracies of predicting psychological states and stages of team problem solving may not be very practically yet. However, we found that outputting a confidence (a prediction accuracy estimated from the training data) for each future prediction is very helpful. High confidence of machine learning is associated with a higher accuracy of prediction for unseen data. For example, in our example, the predictions with confidence >=0.79 have an overall accuracy is 89%. Based on this, a machine model can be used to automate part of sense and decision making tasks of human analysts when there is a high confidence prediction. With this setup, human analysts can focus on the ones with low confidence. Such collaboration between machine and human analysis can partially reduce manpower and the current workload of human analysts. In the collaborative meaning search, we show a systematic way to improve accuracy by iterating between human and machine models.

Table 5 shows a collaborative meaning performed among the three models for predicting states using the setting in Table 1. The predictive accuracy is improved overall as well as in most of the individual states. The three collaborators are the three models shown in Table 5 - use content only, use content and features, and use content, features, and previous states. By looking at the three models collaboratively, the “correct” and “true” meaning or labels are picked up based on which collaborator has a higher confidence.

Collaborative meaning search is important for distributed team decisions and multi-culture environments where a domain/culture specific model and meaning prediction can be picked up because there is a higher confidence where the model is made, for example, either from a richer training data or from a better domain/culture specificity.

Table 5: State Prediction Using Collaborative Meaning Search

Setting 1:Train FS-3; Test FS-3	Content	Content + Features	Content +Features+Previous States	Collaborative Meaning Search
Overall	**30%**	**35%**	**49%**	**54%**
State 3: individual task knowledge development		44%		44%
State 12: convergence of individual mental models to team mental model	80%	85%	58%	72%
State 6: individual visualization and representation of meaning	58%	46%		62%
State 4: team knowledge development			71%	74%
State 10: team shared understanding development			80%	72%
State 8: knowledge interoperability development	48%	45%	48%	52%
State 9: iterative information collection and analysis			40%	

In summary, we found that the most effective way to improve the prediction of macrocognitive stages and states is a collaborative meaning search that selects the best prediction based on a confidence measure.

Decision Making Using Katrina Blogs

We adapted the NEO scenario to the context of Katrina disaster management in August 2005. We develop an Evacuation Operation: Katrina rescue scenario. Katrina disaster management involved many people, multiple organizations and government agencies. In this case, there was background and expert information that came from official sources such as government agencies or news organizations. More importantly, there were vast amounts of open source information from sources such as blogs during that time frame. Since it is difficult to get the archived data from official sources, we adapted the NEO background and expert information for the Katrina scenario for proof of concept. We collected approximately 300 blog entries over four days (Aug. 28th, 29th, 30th and 31st, 2005). Blog entries are dynamic, real-time data that are used to compensate for "official" data. We use this data to illustrate a decision making framework using semantical machine understanding. In the scenario, team collaborators have to make a series of recommendations on what assets (transportation, personnel and route etc) to choose and why (information that supports a recommendation). For example, to decide on transportation, they can search for "helicopter" and "boat". The search returns the numbers of matches from the two repositories (5,3) and (2,1) for "helicopter" and "boat" respectively. At this point, the transportation decision seems to go for a helicopter since it has more matched capability and knowledge. However, when adding blogs as the new repository, the search returns 17 blogs containing "helicopter" and 20 blogs for "boat".

When analyzing the 17 blogs containing "helicopter", they may discover that the 17 blogs containing "helicopter" can be grouped into a few distinct and meaningful categories that:

- Confirm and corroborate the current official information: helicopters are performing rescuing jobs.
- Discover new information: the number of helicopters was very limited (only four were used in rescue) and people were shooting at them.
- Discover new information: helicopters might have fuel concern since all the gas stations are not available.

At this point, the user might aggregate and estimate the impact of new information and feasibility of an action. The result indicates that he/she recommends NOT to use a helicopter as the transportation because of the risk factors from the newly discovered information. At this point, he/she may recommend using a boat for the rescue.

International Disaster Relief Effort (PACOM)

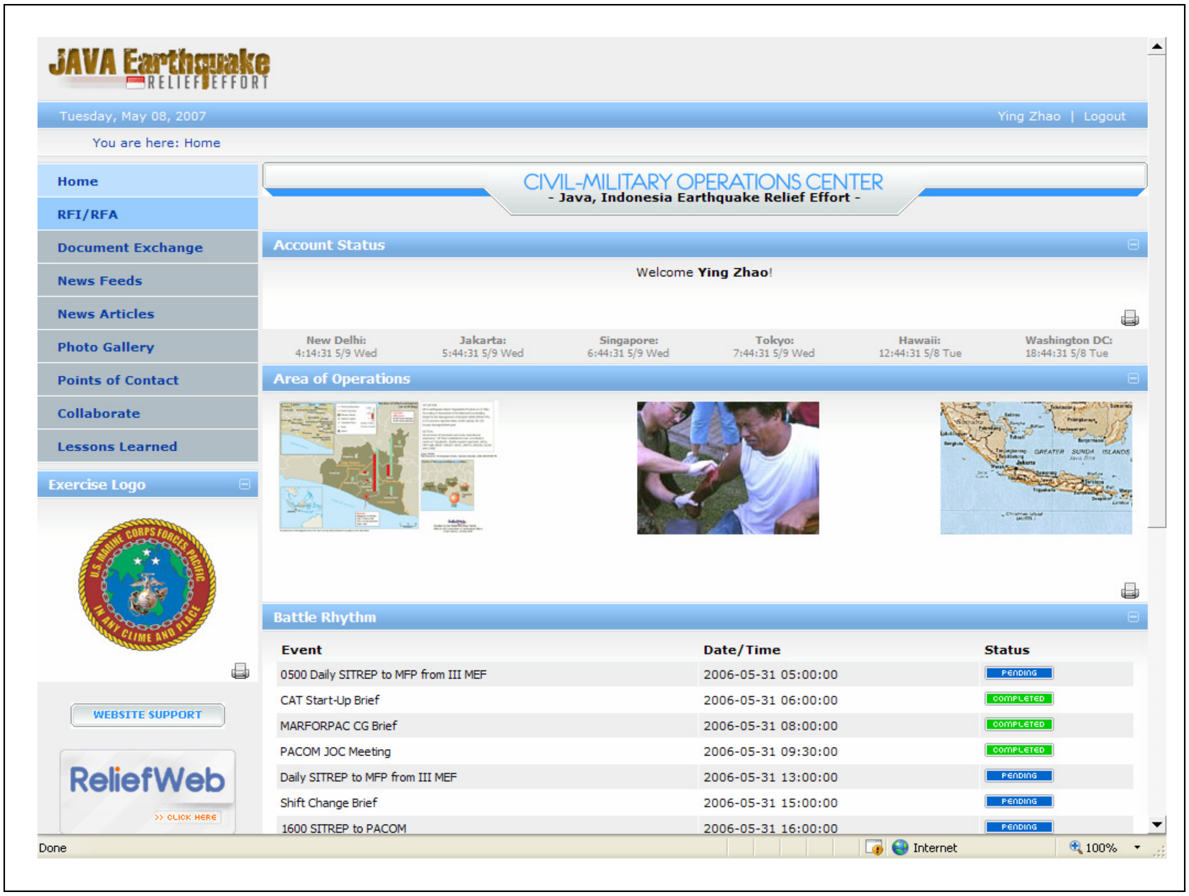


Figure 4. Java Earthquake Relief Effort Website

To see how this decision making framework might map to a real emergency operation business, we have looked at the Java Earthquake Relief Effort website (see Figure 4). The Java Earthquake Relief Effort website includes organizations, information types, available data and reference websites. Approximately 30 countries participated in unclassified information exchange through the website. Figure 5 is a summary of the review. A domain expert explained how the business process is related to the data as shown in Figure 5. Situation Reports (SITREPs) and Request for Action (RFA) are created by participating commands daily (includes a 24hr summary and forecast for 24-78hrs). Orders are decisions that are communicated to everyone and provide authority using the structured United States Message Text Format (USMTF). There are steps to act on in a real-life emergency operation:

Step 1: gather/store information (SITREPs, RFA, websites, news, etc...)

Step 2: visualize data

Step 3: present data to decision makers (SITREPs, briefings)

Step 4: communicate decision (orders)

Step 5: action (RFAs)

In summary, using this example, we found that information gathering (SITREPs, RFA, websites, news, etc), data presentation and decision making are the areas that Semantical Machine Understanding can help. Because of the diversified document types and collaborative partners, a semantic search engine that interprets the meaning and decides the value of a piece of information could be very helpful.

Sentiment Classification and Unsupervised Learning

As shown before, in order to perform a semantic search for decision making, the key factor is to decide what’s the meaning given a piece of. In real-life, human labeling meaning of information may be very expensive; one of simplifications is to get human labeled meaning as “positive” or “negative”, “good” or “bad”, “pros” or “cons” (to a decision, for example). Recent years have seen rapid growth in on-line discussion groups and review sites where a crucial characteristic of a piece of posted information is a special sentiment, or overall opinion towards a decision of subject matter. For example, whether a product is worth buying based on how many number of the positive or negative reviews it has.

Sentiment classification or annotation of phrases and texts is related to topical or text categorization. Sentiment analysis of blogs, review sites and online forums has attracted substantial interest for recent years in the field of natural language processing. Commercialization potential is huge for market intelligence. Traditionally, companies have captured such opinions through customer satisfaction surveys and focus groups in order to understand their users’ needs and improve products and services to meet these demands. The emergence of vast amounts of opinions online in the form of professional product reviews make it imperative to automate sentimental understanding of large amount information.

Organizations

- PACOM
  - United States Pacific Command
- MARFORPAC
  - Marine Forces Pacific
- III MEF
  - Marine Expeditionary Force

Information Types

- SITREPs
  - TXT, DOC
- Message/Orders
  - TXT, RTF
- Reports/Briefs
  - PPT
  - *Information: Template of information required (battle rhythm, overview, significant events, G-X, protocol, concerns/issues)*
- News
  - News from feeds (e.g. Yahoo feed)
  - HTML
- RFI/RFA
  - Request for Military Assistance
  - *How this ties to the process?*

Available Data

Organization/Information Type	SITREPs	Messages/Orders	Reports/Briefs
PACOM		x	
MARFORPAC			
III MEF	x		x

Information Type	Available
Articles	
News	Many
RFI/RFA	

Reference Websites

- USAID
  - Relief effort help; includes some data sheets
- US Department of State
  - Information on Indonesia
- World Bank Group
  - Information on Indonesia
- CIA World Fact Book
  - Information on Indonesia
- WHO (World Health Organization)
  - Situational Reports

Figure 5. Java Earthquake Website Review

Sentiment classification is a very challenging and difficult task as well. First, sentiment requires more understanding than the usual topic-based classification. Therefore, using machine learning and text categorization often results in low performance comparing to the fact that distinguishing positive from negative reviews is relatively easy for humans. Thirdly, little work has been done in trying to create a sentiment classifier that can operate across new domains, let alone across cultures. Conversely, domains may share language to convey sentiment. For example, people often use many of the same words to describe what they liked and disliked about movies and books. A domain-specific approach, however, requires training data in every domain with labeled meaning. Much of the previous work [4-7] has dealt with single domain classification where here are large amounts of labeled data available for training.

Automatic methods of sentiment annotation at the word level employ different techniques that can be grouped in two categories: (1) corpus-based approaches and (2) dictionary-based approaches. The first group includes methods that rely on syntactic or co-occurrence patterns of words in large texts to determine their sentiment [7-11]. However, most work in this area has relied on some level of human supervision – supervised learning -- in the form of hand-tagging or word-list construction. Popular corpora for labeled train data set including movie reviews split into positive and negative [12]; product data and opinions from online review boards which ranks range from +3 to -3 [13] and General Inquirer which places a list of words into various descriptive categories such as “positive”, “negative”, “pain”, “pleasure”, ”yes” and “no” [14] [15]. The majority of dictionary-based approaches use WordNet information to acquire sentiment marked words [13, 16] to create training sets for automatic sentiment classifiers [17] and to measure the similarity between candidate words and sentiment-bearing words [18]. Only recently has cross-domain training data been collected via a collaborative infrastructure. For example, RateItAll http://www.rateitall.com/ [19] is an online repository of consumer written reviews on a wide variety of topics including products and services. The reviews each have a rating, 1 to 5 stars, assigned by the author. Once submitted to RateItAll, the reviews do not go through an editorial process and are presented as is. Based on RateItAll, a sentiment classification called Reasoning Through Search (RTS) technique [20] uses existing labeled data and query formation strategies to estimate the sentiment of a text review. The classification system leverages domain relatedness where a total of 106,961 reviews from these 13 domains (actors, books, colleges, destinations, drinks, electronics, food, movies, music, restaurants, software, sports and video games). The reviews were rated between 1 and 5, the negative reviews were those with 1 or 2 stars, and positive reviews were those with 4 or 5 stars. Reviews with 3 stars were to be neutral.

Various supervised machine learning strategies (Naıve Bayes, SVM, Maximum Entropy, etc.) and the feature sets such as unigrams, n-grams, adjectives, etc. are used to train the classifiers [5]. Unsupervised learning applied in the area is rare. One unsupervised learning technique employs the mutual information measure between document phrases and the words “excellent” and “poor”, where the mutual information is computed using the statistics gathered by a search engine [6]. The number of hits (matching documents) returned from a search engine is then used decide a sentiment. A relaxation labeling method [21] is also an unsupervised method to extract and analyze opinion phrases corresponding to features as opposed to classifying the entire document.

Using the idea of sentiment classification, we have adopted a labeled sentiment word list from General Inquirer http://www.wjh.harvard.edu/~inquirer/ to illustrate a decision making mechanism which is able to decide if a piece of information “positive” or “negative”.

For example, in the Katrina Scenario a decision between “helicopter” and “boat” can be made after sending the context keywords “helicopter” and “boat” to a collaborative meaning search engine. The returning pieces of information are then grouped into “positive” and “negative” where a recommendation can be made based how many positive or negative information with respect to the search context. Sentiment classification is the key for the semantic search based decision making.

Apply iterations to improve sentiment classification and decision making

Sentiment classification and decision making traditionally require some forms of supervised learning. However in real applications, human labeled sentiments and decisions are expensive. It is thus more important to develop iteratively unsupervised learning to achieve the same goal. We want to illustrate a process that starts with a very small number of human labels and gradually and iteratively expand the label sets. We also compiled a second sentiment word list using WordNet (http://wordnet.princeton.edu/) and the method described in [13] which allows us to build a second sentiment search model. We use both to illustrate an iterative and collaborative approach that is capable of improving sentiment classification and decision making with little supervision as possible. In order to illustrate the process, we use a public data set (5331 positive and 5331 negative movie review sentences from http://www.cs.cornell.edu/people/pabo/movie-review-data) where human analysts label each sentiment. However, we only use the human labeled sentiments for validation, not for training. We apply an iterative approach as follows to decide the positive or negative sentiment of a sentence.

- Step 1: Start with a very short human labeled sentiment word list labeled positive or negative
- Step 2: Decide a sentence’s sentiment by counting how many positive or negative words from the short list appearing in a sentence. If more positive words than negative, the whole sentence is tagged positive
- Step 3: Extract a long list of words characterizing the meaning clusters in the whole data set. The list is discovered by the ccc algorithm.
- Step 4: For each word in the long list, estimate how likely it appears in the same sentence with the “positive” or “negative” words in the short list. Decide a word “positive” in the long list if it is more likely to associate with the positive words in the short list; “negative” otherwise.
- Step 5: Decide the sentences’ sentiment using the same decision rule in Step 2 and the long list generated from Step 4.
- Step 6: Iterate Step 4 and 5 a few times where the short and long list are merged as one, however the sentiment predictions are improved every iteration until they converge.

Table 6: Iterative and Unsupervised Sentiment Classification of Movie Reviews

Accuracy	Step 2	Step 5	Step 6 (iteration 1)	Step 6 (iteration 10)
	52%	54%	59%	80%

As shown in Table 6, the accuracy, which is the rate of predicted sentiments that are the same as with the labels from human analysts for a validation data set, is greatly improved using the iterative approach. The method employs our context-concept-cluster algorithm to discover keywords in an unsupervised way with little human labels to start with.

Conclusion

In this paper, we demonstrate the feasibility for an innovative Semantical Machine Understanding system that includes text mining, meaning learning and collaborative search on three open-source (internet) data sets, collaborative team problem solving transcripts and two use cases of sense making and decision making. The key contribution of our work is to apply combined innovations in text mining, meaning learning and collaborative meaning search to construct a semantic search architecture that greatly improve sense and decision making for multinational, multicultural, and coalition applications.

Acknowledgements

This work is based on the research in a Phase I SBIR ONR Contract No. N00014-07-M-0071. We want to thank Dr. Mike Letsky at ONR, Dr. Warner Norman at NAVAIR and Mr. Jens Jensen at USPACOM for valuable discussion.

References

1. Gerber, C., *Smart Searching.* In Military Information Technology, 2005. **9**(9).
2. Foltz, P.W., *Quantitative cognitive models of Text and Discourse Processing.* In The Handbook of Discourse Processes. Mahwah, NJ: Lawrence Erlbaum Publishing, 2002.
3. Bill Frakes and R. Baeza-Yates, eds. *Information Retrieval Data Structures & Algorithms*. 1992, Prentice-Hall: Englewood Cliffs, New Jersay.
4. Alm, C.O., D. Roth, and R. Sproat, Emotions from text: machine learning for text-based emotion prediction. Proceedings of HLT/EMNLP, 2005.
5. Pang, B., L. Lee, and S. Vaithyanathan, Thumbs up?sentiment classification using machine learning techniques. Proceedings of EMNLP, 2002: p. 79-86.
6. Turney, P.D., *Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews.* Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, (ACL), Philadelphia, 2002: p. 417-424.

7. Turney, P.D. and M.L. Littman, *Measuring praise and criticism: Inference of semantic orientation from association.* ACM Transactions on Information Systems (TOIS), 2003. **21**(4): p. 315-346.

8. Grefenstette, G., et al., eds. *Validating the Coverage of Lexical Resources for Affect Analysis and Automatically Classifying New Words along Semantic Axes*. Computing Attitude and Affect in Text: Theory and Application, ed. J.G. Shanahan, Y. Qu, and J. Wiebe. 2006, Springer Verlag.

9. Hatzivassiloglou, V. and K.B. McKeown. *Predicting the Semantic Orientation of Adjectives.* in *In Proceedings of ACL-97, the 35th Annual Meeting of the Association for Computational Linguistics*. 1997. Madrid,Spain.

10. Kim, S.-M. and E. Hovy. *Determining the Sentiment of Opinion*. in *In Proceedings of COLING-04, the Conference on Computational Linguistics*. 2004. Geneva, Switzerland.

11. Yu, H. and V. Hatzivassiloglou. *Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences*. in *In Proceedings of EMNLP-03, the 8th Conference on Empirical Methods in Natural Language Processin*. 2003.

12. Pang, B. and L. Lee, *A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts.* Proceedings of the ACL. Data Available at http://www.cs.cornell.edu/people/pabo/movie-review-data, 2004.

13. Hu, M. and B. Liu, *Mining and summarizing customer reviews.* Proceedings of the ACM SIGKDD International Conferences on Knowledge Discovery & Data Mining. Data Available at http://www.cs.uic.edu/~liub/FBS/FBS.html, 2004.

14. *General Inquirer.* Data Available at: http://www.wjh.harvard.edu/~inquirer/, 1997.

15. Agrin, N., *Developing a flexible analysis techniques for multiple domains.* http://courses.ischool.berkeley.edu/i256/f06/projects/agrin.pdf., 2006.

16. Strapparava, C. and A. Valitutti. *WordNet-Affect: and Affective Extension of WordNet*. in *In Proceedings of LREC-04, the 4th Conference on Language Resources and Evaluation*. 2004. Lisbon, Portugal.

17. Esuli, A. and F. Sebastiani. *Determining the Semantic Orientation of Terms through Gloss Analysis*. in *In Proceedings of CIKM-05, the 14th ACM International Conference on Information and Knowledge Management*. 2005. Bremen, Germany.

18. Kamps, J., et al. *Using WordNet to Measure Semantic Orientation of Adjectives*. in *In Proceedings of LREC 2004, the 4th International Conference on Language Resources and Evaluation, volume IV*. 2004. Lisbon, Portugal.

19. *Rate-It-All: The Opinion Network.* http://www.rateitall.com/, 2006.

20. Sood, S., et al., *Reasoning Through Search: A Novel Approach to Sentiment Classification.* http://www.cs.northwestern.edu/~pardo/courses/EECS395-22-MachineLearning-Winter07/papers/sentiment-classification.pdf, 2007.

21. Popescu, A.-M. and O. Etzioni, *Extracting product features and opinions from reviews.* In Proceedings of HLT-EMNLP 2005, 2005.

22. http://www.onr.navy.mil/sci_tech/34/341/cki/publications.asp