Interaction Mining

Making business sense of customers conversations through semantic and pragmatic analysis

[1]Vincenzo Pallotta, [1]Lammert Vrieling
School of Business and Technology
Webster University Geneva
Switzerland

[1]Rodolfo Delmonte
Department of Language Sciences
Università "Ca Foscari" Venezia, Italy

Abstract

In this chapter we present the major challenges of a new trend in business analytics, namely Interaction Mining. With the proliferation of unstructured data as the result of people interacting with each other using digital networked devices, classical methods in text business analytics are no longer effective. We identified the causes of their failure as being related to the inadequacy of dealing with conversational data. We propose then to move from Text Mining towards Interaction Mining, and we make several cases for this transition in areas such as marketing research, social media analytics, and customer relationship management. We also propose a roadmap for the future development of Interaction Mining by challenging the current practices in business intelligence and information visualization.

1 Introduction

Via the Web a wealth of information for business research is ready at our fingertips. Analyzing this – unstructured – information, however, can be very difficult. *Analytics* has become the business buzzword distinguishing traditional competitors from 'analytics competitors' who have dramatically boosted their revenues. The latter competitors distinguish themselves through "expert use of statistics and modeling to improve a wide variety of functions" (Davenport, 2006, p. 105). However, not all information lends itself to statistics and models. Actually, most information on the Web is made for, and by, people communicating through 'rich' language. This richness of our language is typically missed or not adequately accounted for in (statistical) analytics (e.g. Text-mining) – and so is its real meaning – because it is hidden in *semantics* rather than form (e.g. syntax). In our efforts of turning unstructured data into structured data, important information – and our ability to distinguish ourselves from competitors – gets lost.

[1] The authors are co-founders of InterAnalytics, Geneva, Switzerland, www.interanalytics.ch.

Search engines (Büttcher, Clarke, & Cormack, 2010) have exploited statistical (frequency-based, TF-IDF[2]) methods to its extreme, but building indexes of Web content with keywords is not enough for understanding beyond keyword-based search. The use of semantics in search would be a great improvement and new generation search engines (Grimes, 2010) are starting to address this. Semantic search can be approached from several perspectives. The most common one is to go beyond word forms and consider concepts with their semantic relationships. Concepts can be extracted implicitly or explicitly. In the first case, the contexts of words in a document base determine the concept (Landauer, Foltz, & Laham, 1998). In the second case, concepts are assigned to word forms through a semantic lexicon or ontology such as WordNet[3].

However, semantics is not only necessary for search, it is necessary for any processing of information from the Web. After all, semantics simply means "making sense" and we would like to argue that sophisticated semantic analysis of content is a necessary tool for quality business research of Web data.

For instance, any business analyst in *fast moving consumer goods* (FMCG) is looking for so much more than just text when analyzing on-line focus group interview data. A FMCG analyst would analyze interaction, which would reveal shared language, beliefs and myths, argumentative reasoning, justifications, and changes of opinion or (re)interpretation of experiences (Catterall & Maclaran, 1997).

Focus group interview data is both qualitative and interactive. It is not just text; it is conversational data as people are responding to one another. As such traditional (manual) analysis of focus group data is labor intensive, complex, analyst dependent, inconsistent and subjective[4].

<< The key problem is that good analysis of unstructured data is costly, complex and time-consuming>>

However, the power of current state-of-the-art NLU[5] systems makes automated analysis of these - and other - types of unstructured data feasible and possible (Delmonte, Bristot, & Pallotta, 2010).

[2] TF-IDF stands for Term Frequency-Inverse Document Frequency. It is a common weighting schema in information retrieval and a statistical measure used to evaluate how important a word is to a document in a collection or corpus.
[3] http://wordnet.princeton.edu/
[4] Methods can be used to assess the level of subjectivity in analysis by comparing analyses of the same data performed by different analysts. This can be achieved by computing the Kappa agreement between the analysts (Cohen, 1960). Usually, subjectivity focus group analysis lies on the choice of the coding scheme and not just on the assignments of codes to text. This makes Kappa agreement test unusable.
[5] Natural Language Understanding

This chapter will put to rest the myth that computers cannot extract rich information from unstructured data[6] even from conversational data. We will put forward a new generation of “Interaction Mining” technology that is analyst independent, consistent regardless of the quantity of data, ‘Machine-like’ precision in its analysis in multiple languages and – compared to manual analysis – a quantum leap faster.

First, we will conduct first a survey of current technology for Interaction Mining by assessing the strengths, weaknesses and limits of current approaches such as Text Mining. Additionally, we will present a study in eliciting business requirements for Interaction Mining in different domains. We will present a new approach, which exploits information extracted from automatic analysis of conversational data, which solves some of the challenges highlighted in the requirements section. We will conclude the chapter by outlining a roadmap for research in Interaction Mining.

2 Interaction Mining

In this section we review some current approaches to Business Analytics such as Data Mining and Text Mining by assessing their benefit, strengths, weaknesses and limits to deal with conversational data. We then propose a novel paradigm that we advocate being more suitable for the analysis of interactions between customers. We call the new paradigm Interaction Mining and it stems from and extends standard approaches to Text Mining. We define Interaction Mining as analyzing *interaction* (or *conversational*) *business data* generating actionable insights.

2.1 Standard approaches in Business Analytics

Gartner defines *[business] analytics* (BA) as “leveraging data in a particular functional process (or application) to enable context-specific insight that is actionable” (Kirk, 2006). In BA, data collected from data collected through structured customers’ data (e.g. transactions, forms) are typically analyzed through *Data Mining* tools (Shmueli, Patel, & Bruce, 2010).

When dealing with unstructured data, data mining tools alone are insufficient. Text Business Analytics (TBA) aims at understanding business data using quantitative (statistical) methods resulting in actionable insights by means of *Text Mining* technologies. Text Mining looks at more unstructured data, which typically come in form of textual documents. Text Mining tools typically extract features from textual content in order to discover interesting patterns or classify these according to similarity. In a sense, Text Mining includes data mining as one of its component.

---

[6] See for example (Intertek, 2002): “The technology for searching and analyzing textual data is based on the ability of computers to handle the meaning (i.e., semantics) of content. While humans can read and understand texts, computers can not.”

Text Mining is about extracting statistically relevant (possibly unknown) patterns (or themes) from textual documents (Feldman & Sanger, 2006). Text Mining can be decomposed into three distinct components:

1. a feature extraction component, which transforms unstructured input into structured data;
2. a data mining component, which discovers statistically significant patterns from data;
3. a visualization component, which allows the user to visualize the relationships between the discovered patterns and map them into pre-defined categories.

For instance, a Text Mining system could help in building clusters of documents in a document base by looking at the similarity of their extracted features (e.g., words, concepts, named entities). The clusters are then visualized with their associated themes.

A Text Mining approach is more about coping with large-sized unstructured textual data repositories than about looking at finer-grained information contained in conversations. While we believe that Text Mining is extremely useful pattern discovery tool, it does not answer alone the questions of *why* these patterns are linked to the actual customer's behaviors and opinions.

2.2 From Text Mining to Interaction Mining

The presence of customers on the Web – through several communication channels – provides companies with a wealth of unstructured data, which is hardly transformed into actionable insights. Most of this unstructured information is user-generated and provided by customers in natural language, either in textual or in speech form.

With the establishment of Web 2.0 (Tapscot & Williams, 2006) we moved from user-generated content to user-generated conversations. Conversational data is clearly unstructured as the customers interact through free-text or speech and they are not constraint by any pre-defined structures when to use natural language.

By looking at conversations (i.e. interlinked contributions produced by several authors) as a process, we need to take a substantial different approach than Text Mining. It is the *process* itself that carries the semantics (true meaning) of the conversation and not just the aggregation of features extracted from individual contributions. In other words, we need to understand not only what is said, but also why and how it has been said.

Interaction Mining extracts rich information (semantic and pragmatic) from unstructured customers' interactions, namely conversations held between people and organizations through various communication channels (e.g. phone calls, email, social media, recorded meetings). Moreover, Interaction Mining makes sense of extracted information by means of appropriate data mining and visualization tools thus enabling what we call Interaction Business Analytics.

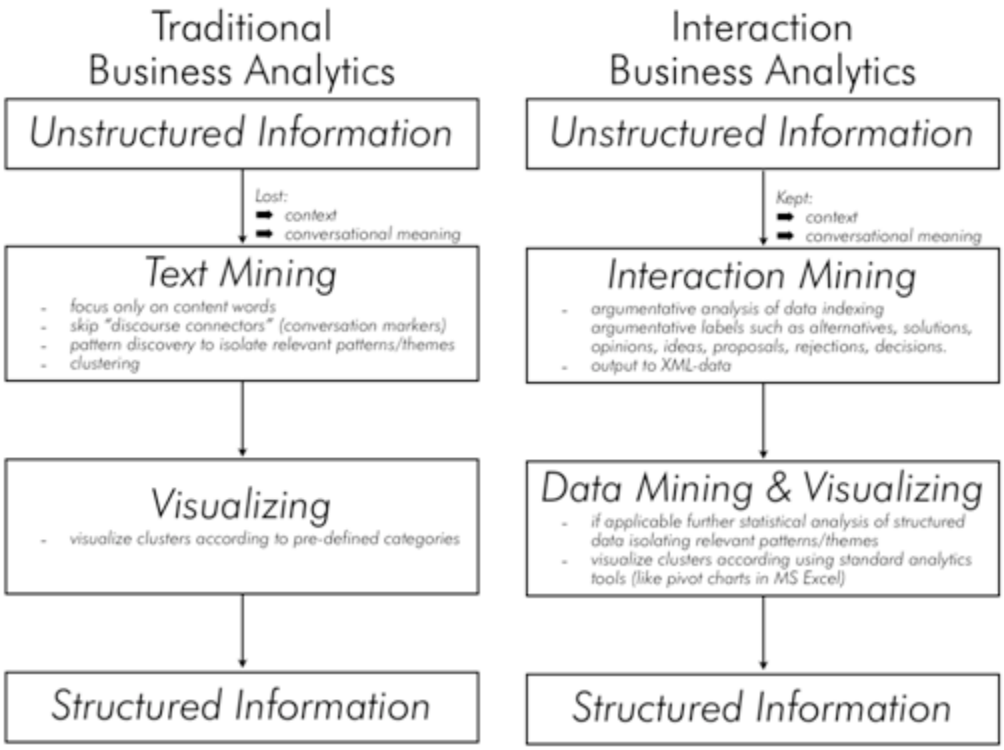

Figure 1 Comparison of Text Business Analytics (based on Text Mining) and Interaction Business Analytics (based on Interaction Mining)

The main similarities and differences between the two approaches are portrayed in Figure 1. The main difference lies in what type of information is extracted from conversation. We basically advocate that an Interaction Mining system should be able to make sense to the actions performed by language. In other words, the system should be able to analyze the *process* of conversation rather than just the *content* of conversation.

Conversations are indeed difficult to process with standard BA tools not only because it is unstructured language data but also because each individual contribution needs to be understood within its context (e.g. with respect to its position in the conversation).

The role of a contribution in conversations is no longer only informative as in a text documents. A contribution carries a *pragmatic force* that can steer the conversation along different directions and outcomes. Moreover, when dealing with unstructured data, Text Mining tools only look at content-bearing words and typically skip discourse connectors (e.g. conjunctions, prepositions) (Blakemore, 2002). These are extremely important in conversations since they carry a lot of the conversational meaning of each contribution. In other words, it is no longer possible to rely to the distributional semantics of documents based on the frequency of content-bearing words.

A simple but illustrative example is when negation is used in language. The sentence: “The judge declared that the CEO did not pay bribes” would always match the query “CEO pays bribes” in a standard content analysis system, but never will the query “CEO was innocent”. There are two main reasons for this simple failure:

1. Negative markers such as “not” are normally removed from indexes;
2. Even if negation would be accounted for, there is no way to infer that “not paying” entails “innocence”.

The above example is only one and maybe the simplest of the challenges posed by interaction mining. Other challenges are those related to various discourse-level phenomena such as anaphora and temporal resolution, detection of rhetorical relations, speech and dialogue acts, ellipsis, presupposition and conversational implicatures. Discussion of these issues is outside the scope of this chapter. The interested reader can refer to (Mitkov, 2003) for an overview.

Another case that clarifies the objectives of Interaction Mining and highlights the limitations of standard Text Mining techniques when applied to conversational data is Opinion Mining and Sentiment Analysis (Pang & Lee, 2008) and (Liu, 2010). Basically, nearly all approaches look at individual contribution and classify them into pre-defined categories (e.g. positive, negative, neutral attitudes). These approaches trivially fail in making sense of complex customer interactions where participant argue about a topic. Whatever algorithm is applied (e.g. based on machine learning or on a lexicon), the *context of the text* (i.e. where and when it appears) is never taken into account. In some cases it might be enough as long as the contribution directly refers to the topic under discussion (e.g. a review of a product). Very often, the text appears within a discussion or a conversation (e.g. comments, replies in blogs or micro-blogs) and negative contribution can refer to other contributions and not to the topic under discussion. A full example of this phenomenon is discussed in more details in section 3.3.

Of course, we also take into account other sources of context beyond the mere conversation, which are also neglected by Text Mining, such as knowledge of the user models and profiles, user’s interaction history, cultural and language settings. Any account of these contexts would improve the quality of understanding of the conversation. Any system that relies on features extracted from the surface form of text (e.g. words) will fail in taking into account such contexts that are implicit and not manifested in the text. In contrast, systems that encode extensive knowledge bases and ontologies will have better chances to contextualize their analysis on multiple dimensions as long as the context of the conversation can be detected and classified by the system. For lack of space and because it would result into a too specific discussion, we do not address purposely these aspects in this chapter.

2.3 Making sense of customers interactions

Until now, using automation in making sense of unstructured customer information was particularly challenging as it must first be turned into structured information and then analyzed with the appropriate tools. Often, the richness of the information source makes it impossible to perform a trivial transformation into quantitative data.

For instance, just counting the occurrence of certain terms is not enough to make sense of opinions expressed in users' product reviews such as those in e-pinions[7]. Looking at ratings and doing standard Opinion Mining is not enough if the company wants to discover the *root cause* of customers' disappointment. In such a case, one has to look closer at the reviews and "understand" exactly why the customers say what they say. Typically, this process is very time-consuming and complex especially in case where information must be harvested from multiple channels (e.g. blogs, social media, contact-centers, forums). The challenge here is to understand exactly what type of information needs to and can be extracted from unstructured conversational data and which methods can efficiently perform that task.

The term multichannel (or cross-channel) Interaction Mining has been introduced by several companies (Autonomy), (NICE) and (Verint) to describe the process of gathering the customers' voice from several communication channels and build actionable knowledge from it. All these products show an interesting trend: companies need to monitor customers' behavior from multiple sources in order to spot any arising issue in due time. Having this knowledge at hand, allow companies to react fast and fix the problem before it becomes unmanageable. An example of this is an organization that has a large amount of information spread across more than 100 different sources. Employees would spend about 50% of their time searching for the right type of information among these sources to answer customers' questions. What was needed was a system that would provide access to aggregate information over all different sources. Information needed to be automatically categorized, linked and delivered to the employees so that they can respond fast and accurate.

The minimum common denominator of these approaches is that unstructured data is extracted from several sources and turned into structured data for statistical quantitative analysis. Very few details are provided on what type of information is extracted but essentially most of the products are based on Search and Text Mining techniques and as such do not take real conversational data into account.

In summary, we believe that the benefit of applying Text Mining to conversational data is already major. However, we also believe that Text Mining provides only a slight contribution to what can be really extracted through a comprehensive account of natural language interaction between customers. Improvement of Text Mining techniques can be achieved by shifting the focus from content to process and have ways to extract features that allow us to understand the key points of interaction. One

[7] www.epinions.com

possible approach we present in this chapter is based on argumentative analysis and it will be detailed in section 4.1.

2.4 Related Work

As we already mentioned, Interaction Mining is a novel and emerging approach to business analytics and so far there are very few work that we can consider as strictly related. Loosely, we consider relevant Text Mining and Opinion Mining because these techniques can be used as a starting point for more elaborate analysis required for Interaction Mining's interaction mining. Data Mining and Information Visualization are clearly useful tools to further making sense to information extracted from conversational data.

One area of investigation we believe very relevant to Interaction Mining is that of Social Analytics[8]. This new trend includes the analysis of user-generated content published in social networks. Many big and small companies such as, for instance, SAS[9], ViralHeat[10], and Alterian[11] are trying to impose their signature on this area. However, we notice that the tools deployed for analyzing social media conversations are still based on standard Text Mining technology, with a social network analysis twist (Watt, 2003).

Relevant to Interaction Mining are also software architectures for deploying applications. Interaction Mining tools can be standalone or integrated with standard analytics architecture such as IBM's UIMA[12] or GATE[13]. While outside the scope of this chapter, we believe that industry standards for application development are fundamental and we advocate for an extensive use of them in the framework of Interaction Mining in order to ensure interoperability of applications.

Another area that is somehow related to Interaction Mining is Customer Relationship Management[14] (CRM). In CRM, the goal is to analyze the interactions with customers in order to predict future trends and improve the relationship over time. Of course, Interaction Mining would be beneficial in the analysis phase since it would enable better understanding of customers' behavior. So far, analysis of customers' behavior in CRM system is limited to Data Mining of transactional data (e.g. purchases, returns, churn rate in e-commerce websites). Very little work has been done in the area of analyzing unstructured interactions with customers (e.g. contact centers), let alone interaction between customers themselves. We see here a substantial impact of Interaction Mining for leveraging interaction information for boosting CRM performance.

---

[8] Gartner group mentioned Social Analytics as one of the top 10 strategic technologies that will have "significant impact" on the enterprise over the next three years (2011-2013).
[9] http://www.sas.com/software/customer-intelligence/social-media-analytics/
[10] http://www.viralheat.com/
[11] http://www.alterian.com/
[12] www.research.ibm.com/UIMA/
[13] http://gate.ac.uk/
[14] http://en.wikipedia.org/wiki/Customer_relationship_management

In the following section we outline some specific requirements for Interaction Mining beyond those that are already met by standard Text Mining techniques, namely the context-unaware analysis of semantic content of text for classification and clustering purposes.

3 Eliciting user requirements for Interaction Mining

The goal of this section is making a case for Interaction Mining by looking at limitations of current BA approaches, in particular of Text Mining and Opinion Mining. For that purpose, we look closer at the intrinsic nature of customers' interactions. Based on our observations, we distinguish between three broad classes of customer's interactions:

1. **Direct interaction between the company and the customer**. This type of interaction is either initiated by the customer by, for instance, calling the contact-center, or solicited by the company through feedback forms or surveys. These interactions are much more focused and issue-oriented and typically synchronous (with the exception of email exchanges).
2. **Indirect interaction between the company and customer**. This typically happens through the public broadcast of corporate messages or consumer-generated content in public forums. This interaction is often asynchronous and with a larger purpose than solving a particular issue. Both the customer and the company can initiate interactions of this kind. Channels used are typically social media, discussion forums, blogs and corporate websites.
3. **Customer-to-customer interaction**. This type of conversations are publicly recorded in discussion forums, chats, and other Web 2.0 collaborative tools, with the purpose of discussing about products or services provided by companies by sharing experiences and best practices.

Standard Text Mining tends to blur these distinctions that we want to make more explicit. We believe that the types of interaction are substantially different and different information extraction techniques need to be used.

For example, in direct interaction of customers with contact centers, one analysis goal can be to monitor agents' performance (e.g. the conversion rate, the problems solved rate). It is apparent that techniques for Opinion Mining are probably inappropriate here. Issue-oriented conversations show little sentiment towards the product but rather concerning the faced issue. A successful method should be able to provide information about whether or not the interaction leads to the customer's satisfaction (i.e. issue solved). This is usually signaled by a conventionalized exchange between the agent and the customer. While it might be possible to spot successful calls through Text Mining technique, the opposite is not, especially if one wants to figure out why the issue was not solved.

For instance, the occurrence of many questions and less assertions might signal a situation where the problem is not yet solved. Subsequently, a *conversational analysis* *system* must be able to at least detect the occurrence of questions and understand if the question is left unanswered or not.

In summary, applying inadequate analysis methods a great deal of information can be ost. The more interactive the conversation is, the higher is the need of focusing on micro-linguistic phenomena such as those that signal the real attitudes of participants and the roles of their contributions in the conversation.

We now examine three possible (among many others) domains of application for Interaction Mining: online focus groups research, quality monitoring of contact centers and mining public opinions in social networks. For each domain we first provide a summarized view on requirements for Interaction Mining. Then, we will provide some solutions in Section 4 and guidelines for future development in Section 5.

3.1 (Online) Focus-group research

We conducted a number of interviews with experts in the domain of qualitative analysis of focus group data and we summarize below our findings about the types of information that is typically considered relevant for the analysis:

- Group transcript into questions (or in our terminology “raised issues”) made by the facilitator.
- For each question/issue/problem, report those that appear to be direct answers to the question (e.g. proposals, ideas, solutions, and opinions).
- Identify requests for clarification/explanations to the initial question and/or to participants’ answers.
- Identify those conversations between participants that focus on a specific aspect of the initial question (e.g. elaboration of a theme).
- Identify those conversations between participants that seem to be off topic or that start a new theme.
- Classify participant’s contribution as factual and opinions/wishes.
- Highlight consensus patterns.
- Highlight agreement/disagreement patterns.

It is apparent that these elements are strongly related to the conversational process rather than to the textual content. One has to look at how themes are discussed and not just identify those themes.

Having a clear framework to understand what types of information is relevant for conducting qualitative analysis of focus group data is helpful also because it creates a minimum common denominator among analyses of comparable data.

Anybody who has analyzed even a few interviews knows how difficult this is. Skilled business analysts using qualitative analysis of conversations typically analyze unstructured data from customer interactions manually (e.g. online focus groups). This approach is expensive and complex but it has the advantage of providing rich explanations of the phenomenon. Proper analysis begins with a careful reading of all data. Next, one will need to assign open codes to all text by identifying the key word or words. Next, the codes will be grouped into broader categories or themes relevant to the research question, which will provide the basis for building the theory and interpretation of the findings. These themes are only verified when two or more groups include them in their discussion. Aforementioned process is often simplified in order to save time and cost through merely listening and/or reading and then summarizing the information.

Ideally, good analysis of focus group data would be consistent, systematic and objective[15]. Moreover, good analysis highlights shared language, what was taken for granted, what needs to be clarified, what was proposed, and what were the positives, negatives, and neutrals. Finally, good analysis distinguishes between facts, opinions, wishes, and statements, and it tracks opinions, agreements and disagreements. Unfortunately, this is typically not the case even if the researchers use advanced tools like Nvivo[16] or Atlas[17]. The latter tools facilitate consistent and systematic analysis but are only as smart as the user and consistently, let alone objectively, tracking the elements mentioned above requires more than a “mere human”. According to our research most analysts do not use these tools but ‘trust’ on their memory and summarizing skills.

Above-mentioned requirements of focus group data indicate that Interaction Mining with fast, objective and ‘Machine-like’ precision would be a differentiator.

3.2 Contact centers

Contact center analytics or “call analytics” is a very active field with many competitors. However, even “whole call analytics” is limited to the logistics of how a call is handled (call volume, call duration, time-to-answer, unnecessary transfers, managing partner transfers, and maximizing Interactive Voice Response (IVR)). Notable exceptions are companies specialized in Speech Analytics[18]. They use statistical analysis of (speech) calls and as such provide improvement to talk times and, for example, reduce the volume of audio for review. However, even if we consider the latter, there is still a lot that conversation analysis can improve for a contact center. For example, careful analysis of transcribed calls can discover correlation between call operators’ utterances types and conversion rate or the number of unmatched requests. The use of ‘association rules’ is often used in Market Basket Analysis[19] (customers who bought this book, also bought these…) but with the use of conversation analysis this can also be

[15] Objectivity is often not even included because it is typically considered to be impossible to be objective in analyzing interviews. There are, however, ways to increase objectivity such as using more analysts for the analysis and having a threshold of cross-analyst agreement. In reality this is most of the time too expensive and time consuming.
[16] http://www.qsrinternational.com
[17] http://www.atlasti.com
[18] http://en.wikipedia.org/wiki/Speech_analytics.
[19] Also referred to as Affinity Analysis: http://en.wikipedia.org/wiki/Affinity_analysis

used in contact centers. For example, agents who simply ask, “Shall I make the booking for you?” make more bookings (Subramaniam, 2008). In the case of contact centers, conversation analysis combined with data mining would make a “killer app”.

3.3 Opinion mining

As mentioned in Section 2.3, Opinion Mining approaches typically fail in making sense of complex customer interactions where participants argue about a topic. In Figure 2

we show an excerpt from a public conversation on Google Wave[20]. The topic of this conversation is Facebook’s privacy policy changes. If we just look at the last contribution, there are several positive terms (e.g. agree, understand, care, good), but the comment is clearly negative as it supports the previous negative statement stating that Facebook’s privacy changes are difficult to understand. For this specific case, most of Opinion Mining system would erroneously recognize the last contribution as positive, while it only confirms the previous negative comment on Facebook’s privacy policy and it should actually count as a negative one. In other words, each contribution must be interpreted within its conversational context and not in absolute terms as it is mostly done in current Opinion Mining technology.

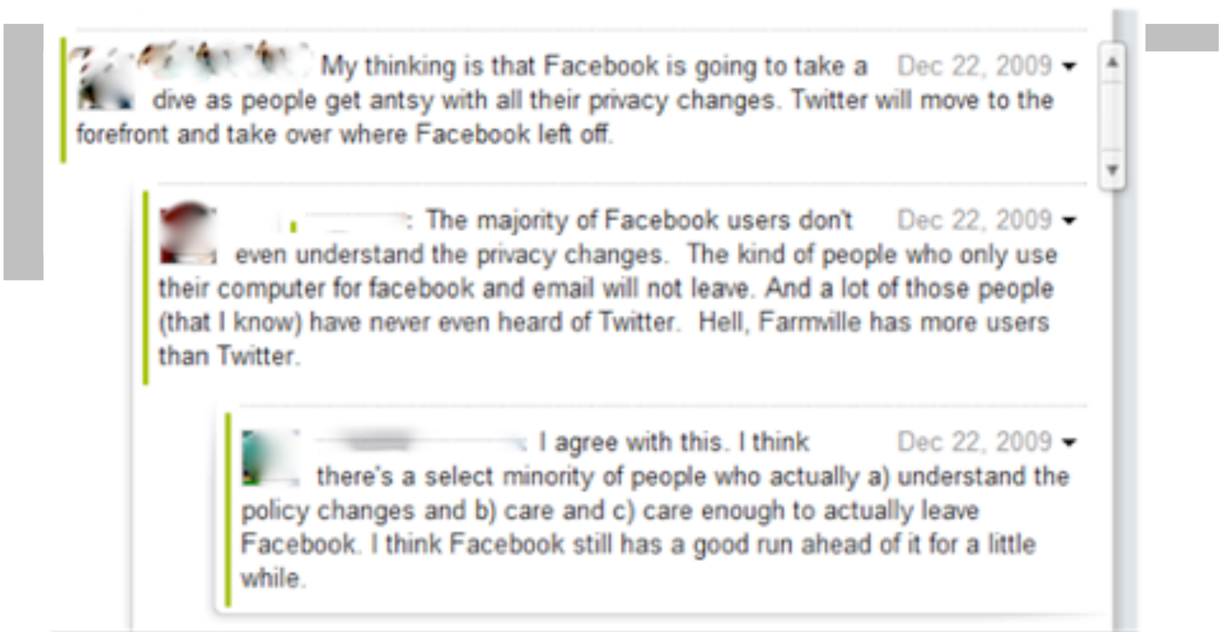

Figure 2 Excerpt of a public conversation on Google Wave

20 http://wave.google.com.

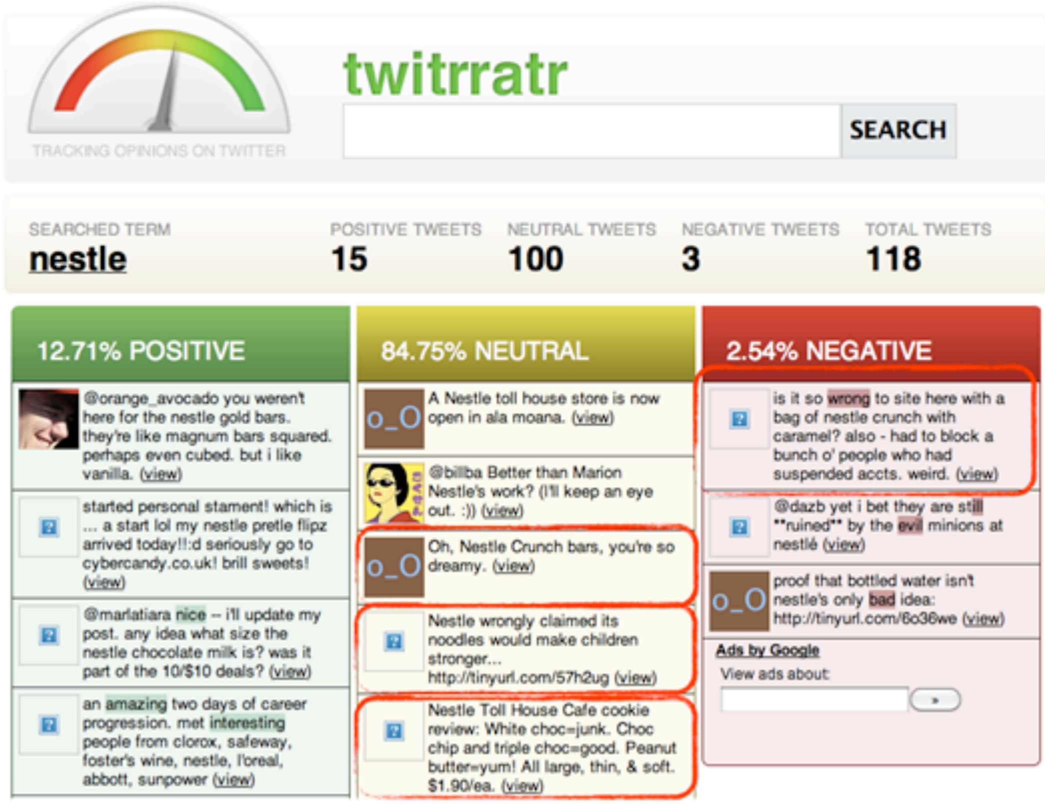

Figure 3 Twittratr's sentiment analysis of Néstlé search term

Similarly, casually searching for the company Nestlé on Twitrratr.com will provide a list of positive, neutral, and negative opinions on this company as shown in Figure 3. The percentages suggest a very exact result: 12.71% positive, 84.75 neutral, and 2.54% negative. However, 4 of the 12 opinions on the page are already wrong. A more careful analysis that we performed manually shows 34% positive, 40% neutral, 16% negative and 10% not clear. A clear example that existing tools cannot deal with these types of statements. In Figure 3, we circled in red those contributions that were misclassified

We recognize that sentiment analysis of Twitter can be extremely tricky due to the fact that tweets are difficult to put in context of a conversation. Moreover, Twitter users tend to engage very little in conversations (i.e. reply to other user's Tweets) but they use the system to post their opinions as a form of public speech. Nevertheless, when conversational structure is present, accounting for this context would prove beneficial in ruling out false positive cases.

There is an interesting trend in Opinion Mining that tries to move the focus from sentence-level towards discourse-level sentiment analysis. In (Somasundaran, Wiebe, & Ruppenhofer, 2008) work has been carried out to automatically label contributions in discussions (i.e. meetings) with discourse-level opinion categories. These categories highlight the role of contributions in the conversation with a polarity twist. While this is not sufficient to fully characterize the discussion process it is nevertheless a serious attempt in framing discussion from argumentative perspective, which we believe is one of the right ways to go if we want to capture the full understanding of conversations.

4 A new approach for Interaction Mining

So far we have demonstrated that in moving beyond Text Mining to Interaction Mining much is to be gained. We illustrated this with examples of online focus groups, contact centers and Opinion Mining. The key question in this paragraph is how can we close the gap between text business analytics and Interaction Mining? Our answer will point to automated argumentative analysis (Pallotta, 2006) as a way to include the richness of interaction into business analytics. In this paragraph we will first describe argumentative analysis including the type of information this reveals. Then we will take again the examples of online focus groups, contact centers and Opinion mining to illustrate what this argumentative analysis will add.

4.1 Argumentative analysis

Our proposal for a new approach to Interaction Mining is to leverage the pragmatic information (automatically extracted through deep linguistic processing of conversations) into structured actionable knowledge. In order to achieve this goal we advocate that the most difficult task is to choose the right level of representation of pragmatic information. This choice substantially affects the linguistic processing required to extract the relevant information.

We believe that a good starting point to address the Interaction Mining requirements is to focus on *argumentative analysis of conversations* as it has already been demonstrated being adequate in (Pallotta, Seretan, & Ailomaa, 2007) for post-meetings information retrieval. Argumentation is pervasive in conversations because people tend to defend their opinions through arguments. At one extreme of the types of conversations we find multi-party dialogs such as face-to-face meetings. These conversations show the highest level of interactivity in conversations and they can be considered as the most difficult case for Interaction Mining. We will make a case for an effective approach in Interaction Mining by providing a study of face-to-face meetings as a baseline for future developments.

It is important to note that looking at dynamics of conversations does not presuppose extensive knowledge of the domain of the conversation. This makes our approach very scalable and robust for dealing with heterogeneous conversational data. The only assumption is that conversations have purpose, which we are aimed at highlighting through argumentative analysis.

To better understand the impact of argumentative analysis we will provide in this section a real example of how it can help in solving outstanding problems in indexing and retrieval of conversational content. In our approach, we adopt a representation of conversational structure based on argumentation theory (Pallotta, 2006). The argumentative structure defines the different patterns of argumentation used by participants in the dialog, as well as their organization and synchronization in the discussion.

The argumentative structure of a conversation is composed of a set of topic discussion episodes (a discussion about a specific topic). In each topic discussion, there exists a set of issue episodes. An issue is generally a local problem in a larger topic to be discussed and solved. Participants propose alternatives, solutions, opinions, ideas, etc. in order to achieve a satisfactory decision. Meanwhile, participants either express their positions and standpoints through acts of accepting or rejecting proposals, or by asking questions related to the current proposals. Hence, for each issue, there is a corresponding set of proposal episodes (solutions, alternatives, ideas, etc.) that are linked to a certain number of related position episodes (for example a rejection to a proposed alternative in a discussed issue) or questions and answers.

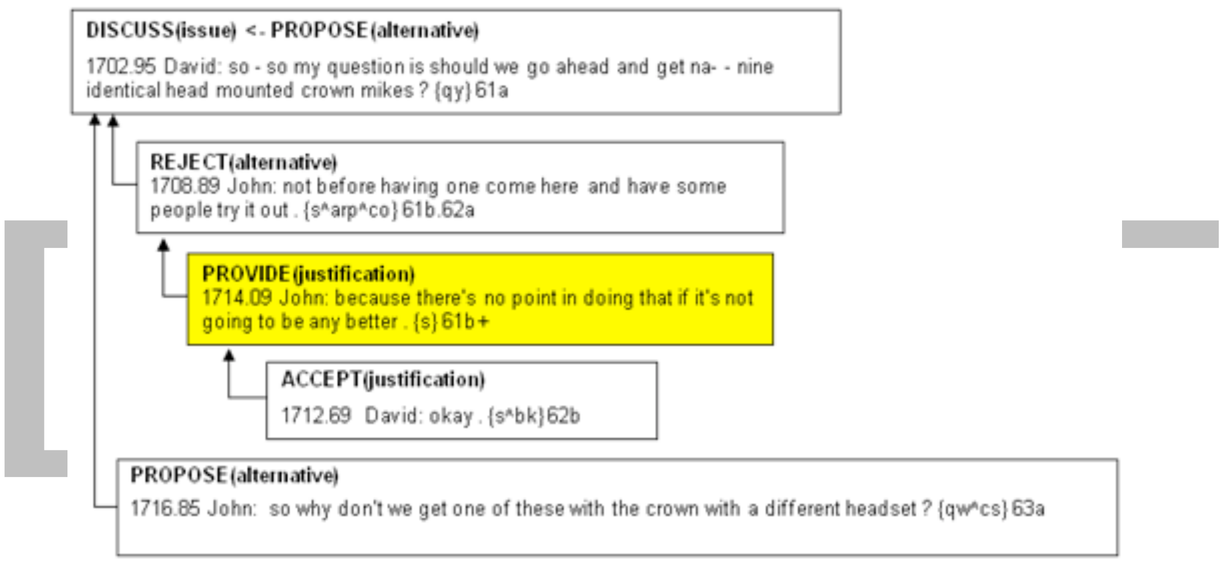

Figure 4. Argumentative structure of a conversation (excerpt).

In Figure 4, we show a diagram representing the argumentative analysis of an episode of a conversation on the topic “microphone”. Each box contains the argumentative role of the turn and its textual transcription as well as some additional metadata already contained in the corpus from which the excerpt has been taken (Janin, et al., 2003). This analysis would allow an analyst to answer a question such as: “*Why was the proposal on microphones rejected*?" by looking for a turn with an argumentative label, “justification”. Of course, finding a justification is not enough because the justification must have been provided for a rejection to a “proposal” (or “alternative”) made to an issue on the topic of “microphones”. This can be done by navigating back through the argumentative links up to the “issue” episode whose content thematically matches the “microphone” topic.

A system capable of identifying and linking argumentative elements of a conversation has been presented and evaluated in (Pallotta, Delmonte, & Bistrot, 2009). This system is also able to provide for each turn sentiment (e.g. positive, negative and neutral) and subjectivity (e.g. factual or subjective) analysis (Delmonte & Pallotta, 2010).

The core of our solution for argument extraction is based on adapting and extending GETARUNS (Delmonte. 2007; 2009), a natural language understanding system developed at the University of Venice. Automatic argumentative annotation is carried out by a special module of GETARUNS activated at the very end of the analysis of each conversation, taking as input its complete semantic representation.

To produce argumentative annotation, the system uses the following 21 discourse relations:

statement, narration, adverse, result, cause, motivation, explanation, question, hypothesis, elaboration, permission, inception, circumstance, obligation, evaluation, agreement, contrast, evidence, hypoth, setting, prohibition.

These are then mapped onto five general argumentative labels:

ACCEPT, REJECT/DISAGREE, PROPOSE/SUGGEST, EXPLAIN/JUSTIFY, REQUEST.

The subjectivity module is able to assign to each turn a combination of labels along several dimensions discussed in (Delmonte, 2007). For the purposes of the examples provided in this chapter, the labels are collapsed into three generic broad categories: FACTIVE, OPINION, and QUESTION.

4.2 The Interaction Mining dashboard

We provide in this section a number of visualizations that we see as good candidates for building Interaction Mining dashboards. The examples are only illustrative of what type of insights can be captured from information extracted from conversations using argumentative, opinion, and subjectivity analysis.

The visualizations are obtained by means of pivot tables and charts in Microsoft Excel. We have found that readers often assume the data for these charts "must have been manually generated". Hence, we emphasize that the data for the diagrams below were automatically generated without any analyst intervention.

The first diagram in Figure 5 highlights the cooperativeness of participants in a conversation. The X axis displays the level of cooperativeness of turns according to the mapping in Table 1, while the Y axis displays the number of turns of each participants falling in those categories. The chart also highlights the level of participation of each participant.

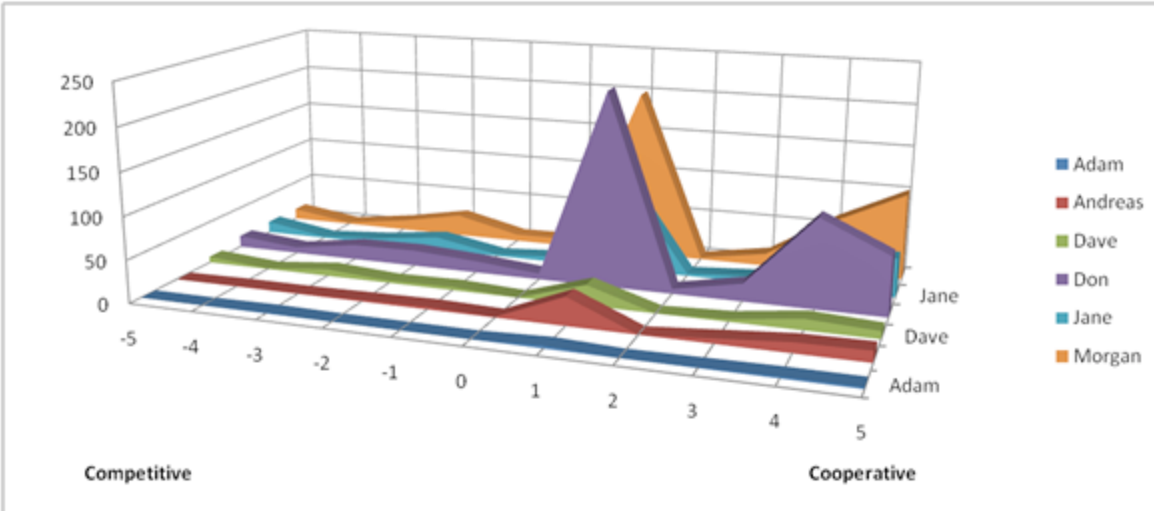

Figure 5 Cooperativeness levels of each participant

Note that from a strict business perspective “competitive” has a positive connotation. In this context, we consider “competitive behavior” as uncooperative. As shown in Table 1, uncooperativeness (i.e. negative scores) is linked to high level of criticism, which is not balanced by constructive contributions (e.g. suggestions and explanations). We acknowledge that this is a rough classification and a better mapping is needed. One possibility for improving the quality of group behavior assessment would be mapping the argumentative categories into Bales’s Interaction Process Analysis framework (Bales, 1950).

Argumentative Categories	Level of Cooperativeness
Accept explanation	5
Suggest	4
Propose	3
Provide opinion	2
Provide explanation or justification	1
Question	-1
Raise issue	-2
Request explanation or justification	-3
Provide negative opinion	-4
Disagree	-5
Reject explanation or justification	-5

Table 1 Mapping table for argumentative categories to levels of cooperativeness

Besides the level of participation the system also automatically highlights the number of participants, participants who talked the most, participants who has undergone the majority of overlaps (interruption) and who has done the majority of overlaps (dominance). This is highlighted in diagram of Figure 6 focusing on group social behavior in terms of both dominance and pair-wise interaction among participants. The size of nodes is proportional to the replies provided and the thickness of edges represents the proportion of turns exchanged between two participants.

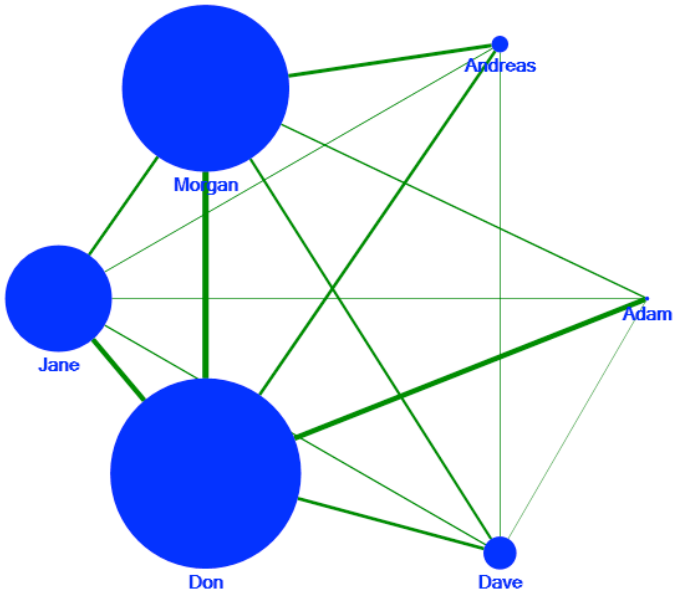

Figure 6 Group Social behavior

The diagram in Figure 6, not only confirms that most interaction was between Don, Jane and Morgan, but also that Morgan and Don dominated the conversation. Notably, Andreas and Adam never talked to each other. This type of analysis is only possible if the information extraction component can detect which participant is replying to a previous turn.

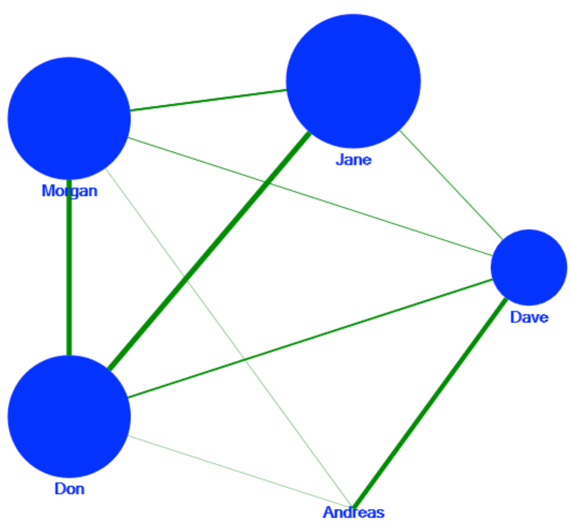

Figure 7. Social behavior for disagreement

If we restrict ourselves to the "disagreement" category, we can understand from Figure 7 that Jane shows proportionally more dissent than other participants and that only Don dares to argue with Morgan while others disagree substantially less with Morgan or between each other. This might also highlight the power relationships between members of a group.

The diagram in Figure 8 shows the attitudes of participants towards the top 10 themes of the discussion. The system automatically generates the discussion topics, who introduced these topics, and whether their attitudes where positive, negative or neutral.

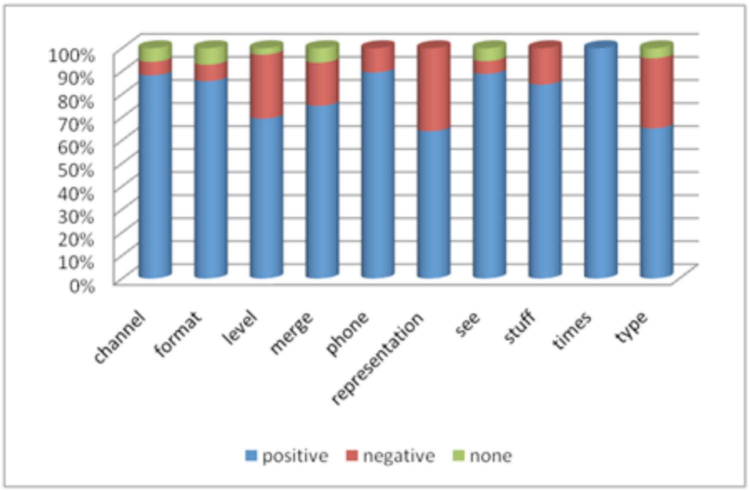

Figure 8 Participants' attitude towards top 10 topics

The diagram in Figure 9 shows a subjectivity analysis displaying the proportion of factual and subjective (opinion) statements (see Section 4.1) made by participants for the top 10 discussed topics. This type of analysis is interesting because it reveals how much of a perception is based on facts or 'mere opinion' and whether or not they have many doubts.

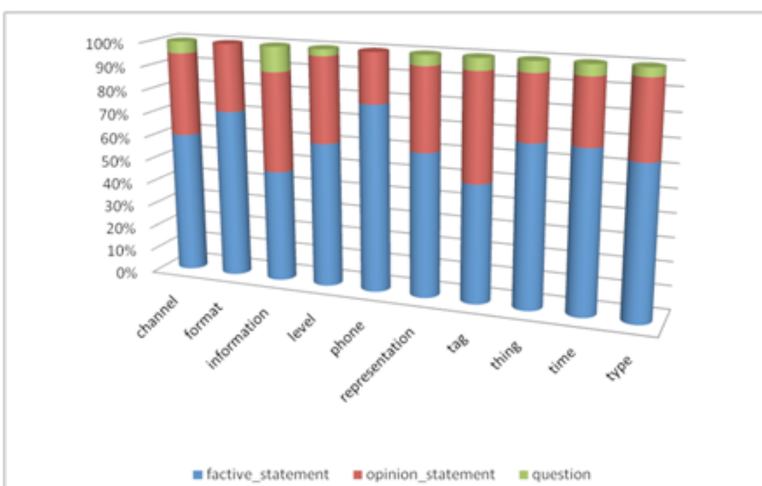

Figure 9 Subjectivity analysis for top 10 discussed topics

The spider diagrams in Figure 10 and Figure 11 show the consensus and dissent levels around the discussed topics.

Figure 10 highlights only the top 10 topics. This information can be helpful in retrospectively analyzing the decision process by looking at how much a decision was supported. If, hypothetically, a decision was made on “data”, although a consensus was present, it is clear that it was not a major one. If this revealed to be a wrong decision, one can track who was effectively responsible.

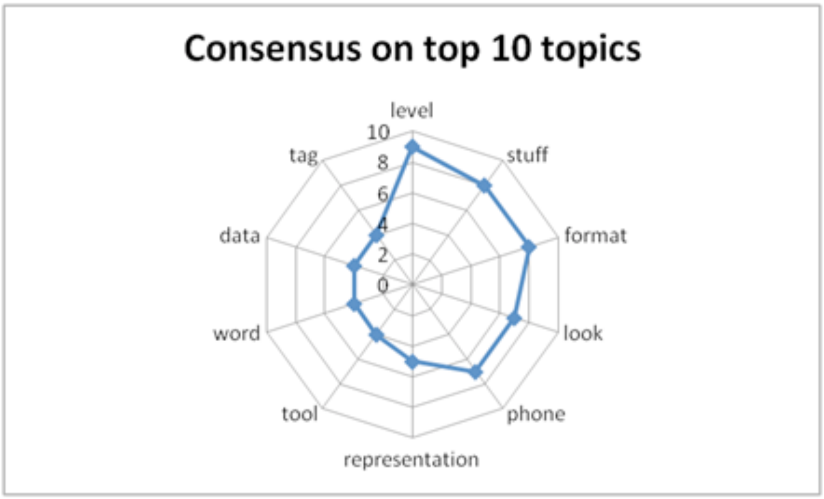

Figure 10 Consensus on top 10 topics

We also provide in Table 2 a breakdown analysis of consensus that highlights who agreed for each of the selected topic topics in terms of number of turns labeled with the Agree label. As one can check, Adam, Jane and Morgan did not explicitly agree on “data”. If they have also expressed dissent on that, they could be relieved from any responsibility on that decision.

Topics \ Speaker	Adam	Andreas	Dave	Don	Jane	Morgan
data		2	1	1		
format			1	4	1	2
level				1	1	7
look				1	3	3
phone					1	6
representation				3	1	1
stuff		1		3		4
tag				2		2
tool				2	1	1
word				1	2	1
Grand Total	**1**	**3**	**2**	**22**	**14**	**34**

Table 2 Breakdown analysis of consensus

In Figure 11, we have an overview of dissent for all the topics discussed in the meeting. In a dashboard containing this diagram, the user can drill down on the topic and visualize the turns where the dissent happens[21].

Figure 11 Dissent on all detected topics

In the following and last example, we show how Data Mining techniques can be used on top of data extracted from argumentative analysis. For this purpose we used a standard Data Mining tool for Excel, XLMiner[22] that we used to induce association rules. Table 3 shows the results.

Conf. %	Antecedent (a)	Consequent (c)
89.36	factive_statement, say=>	positive
80.30	positive, provide_expl_just=>	factive_statement
78.57	provide_expl_just=>	factive_statement
77.99	factive_statement, provide_expl_just=>	positive
76.32	provide_expl_just=>	positive
75.68	positive, say=>	factive_statement
71.18	factive_statement, positive=>	provide_expl_just
71.12	factive_statement=>	positive
68.81	Morgan=>	positive
64.91	factive_statement=>	provide_expl_just
63.09	positive=>	factive_statement
61.28	provide_expl_just=>	factive_statement, positive
56.06	Don=>	positive
55.92	positive=>	provide_expl_just
53.79	Don=>	factive_statement
52.17	factive_statement=>	positive
50.62	factive_statement=>	positive, provide_expl_just

Table 3 Associative Mining on argumentative data

[21] This is already possible in Excel. When clicking on an element of a pivot table or chart, a new sheet is created that contains the relevant rows.
[22] http://www.resample.com/xlminer/index.shtml

We can see that there are a few interesting facts emerging from data:

- We can confirm that Don and Morgan were consistently positive.
- We can observe that Explanations are provided as factual statements and in positive way.
- We can observe that Statements (i.e. those marked with the "say" predicate) are consistently positive.

These results may appear quite obvious. In fact the conversation we have analyzed does not present any particular issue. We might, however, expect radically different results in cases where the conversation is highly controversial.

These are merely some examples of what automated argumentative analysis can do with the conversations. As mentioned, these face-to-face conversations are the most difficult cases for Interaction Mining.

We conclude this section with applications of the Interaction Mining analyses in our three cases: Focus Groups, Contact Centers and Opinion Mining.

4.3 (Online) Focus-group

If we apply the above analysis to focus group interviews it would reveal the nature of the focus group, the level of interactivity, the different levels of contribution of the participants, the topics discussed, facts, opinions (positives, negatives, and neutrals), wishes, doubts, statements, consensus and dissent.

In the above example, the system would indicate that Adam, Andreas and Dave are not very helpful as focus group participants, and that, for example, Jane was quite competitive and not really open. Moreover, the system would list the main topics discussed (format, phone, representation) and the respective attitudes on these topics.

In short the system would analyze the focus group conversations in a consistent, systematic and objective manner.

4.4 Contact Centers

If we apply the analysis to contact centers and in particular to contact center operators the level of cooperativeness of some operators (Adam, Andreas, and Dave) would be 'flagged' for an evaluation conversation. We could compare the level of cooperation to the level of success of an operator. Furthermore, we could mine association rules between argumentative labels (e.g. reject, disagree, propose, suggest) and the conversion rate of operators. These rules would provide a list of specific statements beneficial to these conversion rates. If this would then be linked to an operator assistance system it could within milliseconds suggest 'helpful phrases' to better serve the customer. Indeed, the combination of argumentative analysis and data mining provides many very interesting opportunities for contact centers.

4.5 Opinion Mining

Even though the majority of current Opinion Mining systems are not technically conversational, argumentative and subjectivity analysis tools are much better equipped to handle these types of responses. The level of precision of qualifying opinions as positive, neutral of negative would be very high (about 90%). The real power of argumentative analysis, however, is only released when dealing with truly conversational data such as in the example of Google Wave in Section 3.3. These and other types of review conversations would provide very helpful and correct information on topics discussed.

In this section we introduced an effective approach to Interaction Mining. We demonstrated how argumentative analysis is essential for this approach and illustrated what kind of difference this can make in the area of focus group interviews, contact centers and opinion mining of user-generated content.

5 A roadmap for Interaction Mining

Interaction Mining is clearly in its infancy. The traditional works on Business Analytics, Business Intelligence and Text Mining have shown their intrinsic limitations to cope with conversational data. However, the power of recent NLP/NLU technology make a new generation of analytics tools possible that will eventually meet the requirements we detailed in Section 3.

We suggest a roadmap for research and development in this new area by highlighting the domains where work is still needed in order to solve outstanding problems.

First, we need to improve the quality of conversation analysis. For instance, the evaluation of accuracy of the state-of-the-art argumentative labeling system is still around 80%. To fully unleash the power of argumentative analysis, any system should also be able to compute the back-link between turns (e.g. the “replies to” and “elaborates” relations). This is fundamental if one wants to fully understand the conversation dynamics and detect relevant patterns of consensus/disagreement between participants.

Substantial work needs to be done in understanding, which Data Mining technique would help in discovering relevant patterns from conversation analysis data. We outlined a simple, but powerful technique, association mining, which might help the analyst in spotting issues in contact centers conversations. It might also help in generating a handy knowledge base from which contact center agents can look up during their interaction with clients. We believe that argumentative features can be leveraged to classification tasks as well, for instance in order to aggregate conversations as documents by their similarity (e.g. two conversations can be considered as similar regardless of the word uttered but on the basis of the type of interaction).

Another important aspect is related to visualization of results. We made our case by providing visualizations of analyzed data through off-the-shelf tools such as Microsoft Excel. Already with a low-end visualization tool we are able to provide relevant insights for Interaction Mining. We also did an exercise to imagine how an Interaction Mining dashboard would look like. In Figure 12, we provide a mock-up of a web-based Interaction Mining dashboard, which displays the conversation transcription on the left, as well as the video (if any) of the recorded conversation. This might look very similar to meeting browser such as those described in (AMI Consortium, 2007). However, we also present some other information that is missing in such tools. In the central part of the dashboard, we show a spider diagram, which displays the participants' attitudes. Below, we display a timeline where each participant contribution is plotted on a scale of cooperativeness. The diagram is similar to the conversation graph proposed in (Pallotta, Delmonte, & Ailomaa, 2010).

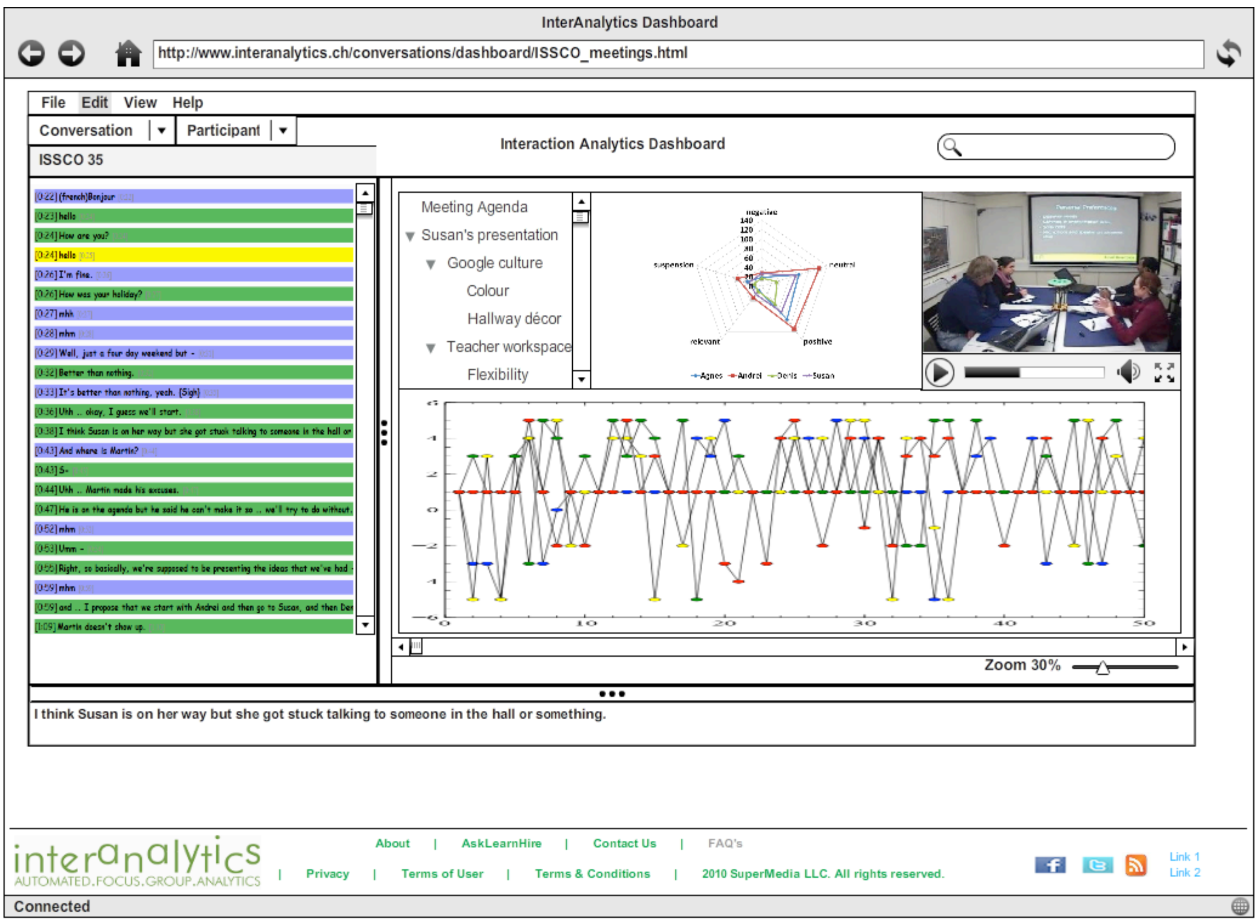

Figure 12 Dashboard for Interaction Mining applications

All the diagrams we have shown in Section 4 could be integrated in the dashboard at the analyst's convenience. Moreover, we also believe that effort should be done to ensure interoperability between Interaction Mining tools sources of conversations such as Social Media, VoIP systems, and qualitative research tools.

Finally, we believe that it is of fundamental importance to build a repository of sample data for experimenting new techniques and tools. At the current state, it is very difficult to put hands on significant data. On the one hand, corporate conversational data is typically confidential. On the other, conversations from the Web are sometimes difficult to extract due to lack of APIs[23]. As it is the case for the Google Wave, this platform provides an excellent conversational space on the web. We suggest that extracting[24] and analyzing conversations from Wave would be an excellent example of integration of a collaborative conversational platform with Interaction Mining tools.

6 Conclusions

Today there is a wealth of unstructured information available to businesses; a wealth that remains mostly untapped. So far, tapping into this information richness was costly, complex and time-consuming because it needed to be done manually. Automated analysis (Text Mining) would typically miss this richness altogether. Nowadays, much of our competitiveness is based on the robustness of our business analytics. Since traditional business analytics is quickly becoming a hygiene factor, Interaction Mining – we believe – will become the new differentiator.

Business Analytics typically aims at understanding business data using quantitative (statistical) methods resulting in actionable insights. Interaction Mining extends these standard approaches as it extracts rich information from unstructured customers' interactions. The difficulty is that conversations are difficult to process with standard Text Mining tools because it is unstructured data, which need to be understood within its pragmatic context.

We indicated that the user requirements for Interaction Mining are distinctly different and more demanding and illustrated this in three business domains (focus groups, contact centers, and opinion mining). We then took the hardest conversational data (face-to-face conversations) and build a case what argumentative analysis can contribute. The examples (again in focus groups, contact centers, and opinion mining) showed that automated argumentative analysis, when necessary combined with association mining, could unlock the wealth of unstructured information that so far remained untapped.

However, much research still remains to be done. We need to improve the quality of analysis, understand which Data mining technique aids the discovery of relevant patterns from conversation analysis data, and improve the visualization of results. Furthermore, there is an important need for sample data to facilitate the above.

---

[23] See (Pallotta, 2010) for more details on digital and online conversations.

[24] FerryBot (http://ferrybot.appspot.com/) is a simple Wave robot that exports a conversation into Google docs. At the current state it does not preserve the names of the speakers for each turn.

7 References

AMI Consortium. (2007). *State of the art report: Meeting browsing.* Retrieved July 15, 2010, from http://www.amiproject.org/ami-scientific-portal/documentation/annual-reports/pdf/D9_3_5.pdf

Autonomy. (n.d.). *Multichannel customer interaction strategy: Identify customer patterns to drive world class customer experience*.

Bales, R. (1950). *Interaction process analysis: A method for the study of small groups.* Cambridge, MA: Addison-Wesley.

Blakemore, D. (2002). *Meaning and relevance: The semantics and pragmatics of discourse markers.* Cambridge University Press.

Büttcher, S., Clarke, C., & Cormack, G. (2010). *Information retrieval: Implementing and evaluating search engines.* MIT Press.

Catterall, M., & Maclaran, P. (1997). Focus group data and qualitative analysis programs: Coding the moving picture as well as the snapshots. *Sociological Research Online , 2*(1), par. 4.6.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46.

Davenport, T. (2006). Competing on analytics. *Harward Business Review* (Special issue on Decision Making), 99-107.

Delmonte, R. (2007). *Computational linguistic text processing - Logical form, semantic interpretation, discourse relations and question answering.* New York, NY: Nova Science Publishers.

Delmonte, R., & Pallotta, V. (2010). Opinion mining and sentiment analysis need text understanding. *Proceedings of 4th International Workshop on Distributed Agent-based Information Retrieval,* Geneva.

Delmonte, R., Bristot, A., & Pallotta, V. (2010). Deep linguistic processing with GETARUNS for spoken dialogue understanding. *Proceedings of LREC 2010 Conference.* Malta: LREC Press.

Feldman, R., & Sanger, J. (2006). *The Text Mining handbook. Advanced approaches in analyzing unstructured data.* Cambridge University Press.

Grimes, S. (2010, January 21). *Breakthrough analysis: Two + nine types of semantic search.* Retrieved July 1st, 2010, from http://www.informationweek.com/news/showArticle.jhtml?articleID=222400100

Intertek. (2002). *Management report on leveraging unstructured data in investment management.* Paris, France: The Intertek Group .

Janin, A., Baron, D., Edwards, J., Ellis, D., Gelbart, D., Morgan, N., et al. (2003). The ICSI meeting corpus. *Proceedings of IEEE/ICASSP* (pp. 364-367). Hong Kong: IEEE Press.

Kirk, J. (2006 , February 7). *Analytics buzzword needs careful definition.* Retrieved October 29, 2010, from http://www.computerworld.com/s/article/108460/_Analytics_buzzword_needs_careful_definition

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes, 25*, 259-284.

Liu, B. (2010). Sentiment analysis and subjectivity. In N. Indurkhya, & F. J. Damerau (Eds.), *Handbook of natural language processing,* 2nd ed.

Mitkov, R. (2003). *Oxford handbook of computational linguistics.* Oxford University Press.

NICE. (n.d.). *Cross-channel interaction analytics.* Retrieved July 1st, 2010, from http://www.nice.com/solutions/enterprise/interaction_analytics.php

Pallotta, V. (2010). Content-based retrieval of distributed multimedia conversational data. In E. Vargiu, A. Soro, G. Armano, & G. Paddeu (Eds.), *Information retrieval and mining in distributed environments.* Springer Verlag series: Studies in Computational Intelligence.

Pallotta, V. (2006). Framing arguments. *Proceedings of the International Conference on Argumentation ISSA.* Amsterdam, NL.

Pallotta, V., Delmonte, R., & Ailomaa, M. (2010). Summarization and visualization of digital conversations. *Proceedings of the 1st Workshop on Semantic Personalized Information Management, part of LREC 2010 Conference.* Malta.

Pallotta, V., Delmonte, R., & Bistrot, A. (2009). Abstractive summarization of voice communications. *Proceedings of the LTC'09 Conference on Language Technology and Computers.* Poznan, PL.

Pallotta, V., Seretan, V., & Ailomaa, M. (2007). User requirements analysis for meeting information retrieval based on query elicitation. *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics.* Prague, Czech Republic: ACL Press.

Pang, B., & Lee, L. (2008). *Opinion mining and sentiment analysis*, vol. 2 (pp. 1-2). Now Publisher: Foundations and Trends in Information Retrieval.

Shmueli, G., Patel, N., & Bruce, P. (2010). *Data mining for business intelligence.* Wiley.

Somasundaran, S., Wiebe, J., & Ruppenhofer, J. (2008). *Discourse level opinion interpretation*. The 22nd International Conference on Computational Linguistics (COLING-2008).

Tapscot, D., & Williams, A. (2006). *Wikinomics.* Penguin Group.

Verint. (n.d.). *Customer interaction analytics.* Retrieved July 1st, 2010, from http://www.bps.nl/brochures/intelli/IntelliFind.pdf

Watt, D. (2003). *Six degrees: The science of a connected age.* New York, NY: W. W. Norton.

Key Terms

Business Intelligence: Intelligence as in "Intelligence Services" aimed at discovering relevant patterns from data so that they can be used for strategic or tactical purposes in the enterprise. Business Intelligence suites might include tools for data analysis, reporting and visualization as well as tools for supporting decision such as trend analysis and forecasting.

Text Business Analytics: All kind of computer tools that analyze unstructured data and turn qualitative information in textual form into measurable data. Also referred as Text Mining, it focus in extracting content from textual data and aggregate the content extracted from large number of similar documents (e.g. news, emails, web pages). Usually, standard Text Business Analytics tools do not take the context (e.g. location, author, time, relationships with other information) of the information pieces into account.

Natural Language Processing/Understanding: NLP/NLU technology is made of algorithms for processing and understanding natural language input and linguistic resources such as lexica, grammars, corpora and ontologies. Processing can be done at different linguistic levels such as syntax, semantics or pragmatics. Understanding of language happens when a system is capable to recognize user's intentions expressed through languages.

Syntax: A basic level of language structure that considers grammatical functions of the words and their aggregation into larger structures within the boundaries of the sentence. Syntax is usually used to check the "well-formedness" of a sentence (e.g. in orthographic checkers) and for determining what are the semantic relations between the sentence constituents (e.g. the subject, objects, predicate in a phrase).

Semantics: The study of language meaning. Semantics is relative to the language unit chosen. It can be word's semantics, sentence's semantics or discourse semantics. It usually refers to the link between expressions and objects (being real of fictive). In other words, semantics is about WHAT is referred by language.

Pragmatics: The study of pragmatics pertains to the use of language to perform actions. It is based on the notion of "speech act", namely the action that is performed through the production of a linguistic expression. The language unit of pragmatic analysis is the "utterance". An utterance can be just a word (e.g. "yes" as an agreement) or even a discourse made of several sentences (e.g. a monologue made as an appraisal). In the specific case of this chapter, one form of pragmatic analysis is "argumentative" analysis.

Argumentation: The study of how language is used to support claims. It is a type of pragmatic analysis. In it simplest form, it studies how people use language for agreeing and disagreeing. In our specific context, it also aims at modeling the process of decision-making and conflict resolution during multi-party discussion.

Focus Group: An intentionally orchestrated series of group discussions aimed to get perceptions on a certain subject (product, interest, etc.) in a relaxed, open environment.

Social Media: media supporting the production of content through social interaction. The content results as the byproduct of conversations between socially connected people that interact by using social communication channel. Social Media are possible because of i) a social connection infrastructure and ii) user-generated content. Social Media subvert the conventional publishing model, which is mediated by "editors". In Social Media, authors and readers are the same entity. Content can be (and is indeed) created by users. Social Media have larger reach for users as the content is often indexed by search engines. This means that topic of interest and their attached communities can be easily discovered by new users, who will eventually become members and (hopefully) contributors. Social Media foster dialogue over monologues.

Data Mining (Knowledge Discovery): It is part of Business Intelligence and it aims at discovering statistically relevant patterns from data. The most common types of analysis are: Classification, Clustering, Association Mining and Regression Analysis. Data Mining typically applies to structured data (e.g. data bases). In order to apply Data Mining to unstructured data (e.g. text or conversations) one has to transform it into structured data. Text Mining is one possible approach to turn textual data into structured data.