Improving Trust in Automation of Social Promotion

Boris Galitsky, Dmitri Ilvovsky, Nina Lebedeva and Daniel Usikov

Knowledge-Trail Inc. San Jose CA 95127 USA
bgalitsky@hotmail.com; dilv_ru@yahoo.com; nina.lebedeva_kt@gmail.com; usikov@hotmail.com

Abstract

We build a conversational agent performing social promotion (CASP) to assist in automation of interacting with friends and managing other social network contacts. This agent employs a domain-independent natural language relevance technique which filters web mining results to support a conversation with friends and other network members. This technique relies on learning parse trees and parse thickets (sets of parse trees) of paragraphs of text such as Facebook postings. We evaluate the robust intelligent features and trust in behavior of this conversational agent in the real world. It is confirmed that maintaining a relevance of automated posting is essential in retaining trust and efficient social promotion of CASP.

Introduction

Simulated human characters and agents are increasingly common components of user interfaces of applications such as interactive tutoring environments, eldercare systems, virtual reality-based systems, intelligent assistant systems, physician-patient communication training, and entertainment applications (Cassell et al., 2000, De Rosis et al., 2003; Dias and Paiva 2005; Lisetti 2008, among others). While these systems have improved in their intelligent features, expressiveness, understanding human language, dialog abilities and other aspects, their social realism is still far behind. It has been shown (Reeves and Nass 1996) that users consistently respond to computers as if they were social actors; however most systems don’t behave as competent social actors, leading to user loose of trust and alienation. Most users used to distrust conversational agent who has shown poor understanding of their needs in the areas such as shopping, finance, travel, navigation, customer support and conflict resolution. To restore trust in automated agents, they have to demonstrate robust intelligence features on one hand and operate in a domain where users are more tolerant to agent’s misunderstanding of what they said.

In this study we build a conversational agent enabled with robust intelligence for social promotion. It is presented as a simulated human character which acts on behalf of its human host to facilitate and manage her communication for her. The agent relieves its human host from the routine, less important activities on social networks such as sharing news and commenting on messages, blogs, forums, images and videos of others. Unlike the majority of application domains for simulated human characters, its social partners do not necessarily know that they exchange news, opinions, and updates with an automated agent. We refer to this agent as a conversational agent which assists its human host with social promotion (CASP). We will experiment with CASP in a number of Facebook accounts and evaluate its performance and trust by human users communicating with it.

To be trusted, a conversational character operating in a natural language must produce the content relevant to what the human peers share. To do that, it needs to implement robust intelligence features (NSF 2013, Lawless et al 2013):

1. Flexibility in respect to various forms of human behavior, and human mental states, such as information sharing, emotion sharing, and information request by humans.
2. Resourcefulness, being capable of finding relevant content in emergent and uncertain situations, for a broad range of emotional and knowledge states.
3. Creativity in finding content and adjusting existing content to the needs of human peers.
4. Real-time responsiveness and long-term reflection on how its postings being perceived. CASP builds a taxonomy of entities it used for relevance assessment and reuses it in future sessions (Galitsky 2013).
5. Use of a variety of reasoning approaches, in particular based on simulation of human mental states.
6. Ability to learn and adapt performance at a level of intelligence seen in humans and animals.

7. Awareness of and competence in larger natural, built, and social contexts.

For a conversational system, users need to feel that it properly reacts to their actions, and that what it replied makes sense. To achieve this in a limited, vertical domain, most effective approaches rely on domain-specific ontologies. In a horizontal domain, one needs to leverage linguistic information to a full degree to be able to exchange messages in a meaningful manner. Once we do not limit the domain a conversational agent is performing in, the only available information is language syntax and discourse, which should be taken into account in a full scale linguistic relevance filter (Galitsky et al. 2012).

Social promotion is based on
• involvement (living the social web, understanding it, going beyond creation of Google+ account);
• creating (making relevant content for communities of interest);
• discussing (each piece of content must be accompanied by a discussion. If an agent creates the content the market needs and have given it away freely, then you will also want to be available to facilitate the ensuing discussions);
• promoting (the agent needs to actively, respectfully, promote the content into the networks).

CASP acts in the environments subject to constant changes. As news come, political events happen, new technologies are developed and new discoveries are made, CASP needs to be able to find relevant information using new terms or new meanings of familiar terms and entities. Hence it needs to automatically learn from the web, expanding its taxonomy of entities and building links between them (Galitsky 2013). These taxonomies are essential when CASP needs to match a portion of text found on the web (as a candidate message) against a message posted by a human user. By growing these taxonomies, CASP learns from the web, adapts its messages to how the available information on the web is evolving. Also, CASP applies accumulated experience from user responses to its previously posted messages to new posting new problems may be better suited to respond to novel environmental challenges.

Paragraphs of text as queries appear in the search-based recommendation domains (Montaner et al., 2003; Bhasker and Srikumar 2010; Thorsten et al., 2012) and social search (Trias et al., 2012). Recommendation agents track user chats, user postings on blogs and forums, user comments on shopping sites, and suggest web documents and their snippets, relevant to a purchase decisions. To do that, these recommendation agents need to take portions of text, produce a search engine query, run it against a search engine API such as Bing or Yahoo, and filter out the search results which are determined to be irrelevant to a purchase decision. The last step is critical for a sensible functionality of a recommendation agent, and poor relevance would lead to lost trust in the recommendation engine. Hence an accurate assessment of similarity between two portions of text is critical to a successful use of recommendation agents.

Parse trees have become a standard form of representing the syntactic structures of sentences (Abney 1991; de Salvo Braz et al 2005; Domingos and Poon 2009). In this study we represent a linguistic structure of a paragraph of text based on parse trees for each sentence of this paragraph. We will refer to the set of parse trees plus a number of arcs for inter-sentence relations between nodes for words as Parse Thicket. A PT is a graph which includes parse trees for each sentence (Collins and Duffy 2002), as well as additional arcs for inter-sentence relationship between parse tree nodes for words.

We will rely on the operation of generalization of text paragraphs via generalization of respective parse thickets to assess similarity between them.

The domain of social promotion

On average, people have 200-300 friends or contacts on social network systems such Facebook and LinkedIn. To maintain active relationships with this high number of friends, a few hours per week is required to read what they
post and comment on it. In reality, people only maintain relationship with 10-20 most close friends, family and colleagues, and the rest of friends are being communicated with very rarely. These not so close friends feel that the social network relationship has been abandoned.

However, maintaining active relationships with all members of social network is beneficial for many aspects of life, from work-related to personal. Users of social network are expected to show to their friends that they are interested in them, care about them, and therefore react to events in their lives, responding to messages posted by them. Hence the users of social network need to devote a significant amount of time to maintain relationships on social networks, but frequently do not possess the time to do it. For close friends and family, users would still socialize manually. For the rest of the network, they would use CASP for social promotion being proposed.

The difficulty in solving this problem lies mainly in the area of relevance. A robust intelligence paradigm is expected to come into play to solve this problem in real world environment with different cultures, broad range of user interests and various forms of friendship and professional connections in social networks.

Messages of the automated agent must be relevant to what human agents are saying. These messages are not always expected to be impressive, witty, or written in style, but at least they should show social engagement, it should show that the host of CASP cares about the friend being communicated with.

The opportunity to automate social promotion leverages the fact that overall coherence and exactness of social communication is rather low. Readers would tolerate worse than ideal style, discourse and quality of content being communicated, as long as overall the communication is positive and makes sense. Currently available commercial chat bots employed by customer support portals, or packaged as mobile apps, possess too limited NLP, text understanding and overall robust intelligent capabilities to support conversations to support social promotion.

For the purpose of promoting social activity and enhance communications with the friends other than most close ones, CASP is authorized to comment on postings, images, videos, and other media. Given one or more sentence of user posting or image caption, this agent issues a Bing Web and Bing Blogs APIs search request and filters the results for relevance.

**CASP architecture**

CASP inputs a seed (a posting written by a human) and outputs a message it forms from a content mined on the web and adjusted to be relevant to the input posting. This relevance is based on the appropriateness in terms of content and appropriateness in terms of mental state: for example, it responds by a question to a question, by an answer to a recommendation post seeking more questions, etc.

![Fig. 2: CASP is about to post a message about his “experience” at lake Tahoe, having his host’s friend image caption as a seed.](image)

![Fig. 3: Three main components of CASP.](image)

CASP includes the following components (Fig. 3):

- Web mining component, which forms the web search queries from the seed.
- Content relevance component, which filters out irrelevant portions of candidate content found on the web, based on syntactic generalization operator. It functions matching the parse forest for a seed with the parse forest for a content found on the web.
- Mental state relevance component, which extracts
mental states from the seed message and from the web content and applies reasoning to verify that the former can be logically followed by the latter.

Fig. 4: Generalizing parse thicket to assess relevance of the seed text and candidate mined portion of text (the second component in Fig. 3).

In (Galitsky 2013) we developed a generic software component for computing consecutive plausible mental states of human agents, which is an important part of a robust intelligence and is employed by CASP. The simulation approach to reasoning about mental world is based on exhaustive search through the space of available behaviors. This approach to reasoning is implemented as a logic program in a natural language multiagent mental simulator NL_MAMS, which yields the totality of possible mental states few steps in advance, given an arbitrary initial mental state of participating agents. Due to an extensive vocabulary of formally represented mental attitudes, communicative actions and accumulated library of behaviors, NL_MAMS is capable of yielding much richer set of sequences of mental state than a conventional system of reasoning about beliefs, desires and intentions would deliver. Also, NL_MAMS functions in domain-independent manner, outperforming machine learning-based systems for accessing plausibility of a sequence of mental states and behaviors of human agents in broad domains where training sets are limited.

The system architecture serves as a basis of OpenNLP – similarity component, which is a separate Apache Software foundation project, accepting input from either OpenNLP or Stanford NLP. It converts parse thicket into JGraphT objects which can be further processed by an extensive set of graph algorithms (Ehrlich and Rarey, 2011). The code and libraries described are available at http://code.google.com/p/relevance-based-on-parse-trees http://svn.apache.org/repos/asf/opennlp/sandbox/opennlp-similarity/. The web mining and posting search system is pluggable into Lucene library to improve search relevance. Also, a SOLR request handler is provided so that search engineers can switch to a parse thicket-based multi-sentence search to quickly verify if relevance is improved in their domains.

Relevance assessment operation

Modern search engines are not very good at tackling queries including multiple sentences. They either find a very similar document, if it is available, or very dissimilar ones, so that search results are not very useful to the user. This is due to the fact that for multi-sentences queries it is rather hard to learn ranking based on user clicks, since the number of longer queries is practically unlimited. Hence we need a linguistic technology which would rank candidate answers based on structural similarity between the question and the answer. In this study we build a graph-based representation for a paragraph of text so that we can track the structural difference between these paragraphs, taking into account not only parse trees but the discourse as well.

To measure the similarity of abstract entities that are expressed by logic formulas, a least-general generalization, or anti-unification, was proposed for a number of machine learning approaches, including explanation-based learning and inductive logic programming. We extend the notion of generalization from logic formulas to the sets of syntactic parse trees of these portions of text. Rather than extracting common keywords, the generalization operation produces a syntactic expression that can be semantically interpreted as a common meaning shared by two sentences.

In this section we briefly describe a contribution to Apache OpenNLP Similarity, which takes the constituency parse trees and finds a set of maximal common sub-trees as expressions in the form of a list \{<lemma, part-of-speech>, ..., \}.

Let us represent a meaning of two NL expressions by using logic formulas and then construct the generalization \(^\star\) of these formulas: camera with digital zoom and camera with zoom for beginners.
To express the meanings, we use the predicates `camera(name_of_feature, type_of_users)` and `zoom(type_of_zoom)`. The above NL expressions will be represented as follows: `camera(zoom(digital), AnyUser) ^ camera(zoom(AnyZoom), beginner)`, where the variables (uninstantiated values that are not specified in NL expressions) are capitalized. The generalization is `camera(zoom(AnyZoom), AnyUser)`. At the syntactic level, we have the generalization of two noun phrases as follows: \{NN-camera, PRP-with, NN-zoom\}.

We use (and contribute to) the (OpenNLP 2013) system to derive the constituent trees for generalization (chunker and parser). The generalization operation occurs on the levels of Paragraph/Sentence/Phrases (noun, verb and others)/Individual word. At each level, except for the lowest level (individual words), the result of the generalization of two expressions is a set of expressions. In this set, expressions for which less-general expressions exist are eliminated. The generalization of two sets of expressions is a set of the sets that are the results of the pair-wise generalization of the expressions in the original two sets.

![Image of constituency parse trees](image)

**Algorithm for generalizing two constituency parse trees**

**Input:** a pair of sentences  
**Output:** a set of maximal common sub-trees  

1. Split sentences into sub-trees that are phrases for each type: verb, noun, prepositional and others. These sub-trees are overlapping. The sub-trees are coded so that the information about their occurrence in the full tree is retained.
2. All of the sub-trees are grouped by phrase types.
3. Extend the list of phrases by adding equivalence transformations (Galitsky et al 2012).
4. Generalize each pair of sub-trees for both sentences for each phrase type.
5. For each pair of sub-trees, yield an alignment and generalize each node for this alignment. Calculate the score for the obtained set of trees (generalization results).
6. For each pair of sub-trees of phrases, select the set of generalizations with the highest score (the least general).
7. Form the sets of generalizations for each phrase type whose elements are the sets of generalizations for that type.
8. Filter the list of generalization results: for the list of generalizations for each phrase type, exclude more general elements from the lists of generalization for a given pair of phrases.

**Evaluation of content relevance**

Having described the system architecture, we proceed to evaluation of how generalization of PTs can improve multi-sentence search, where one needs to compare a query as a paragraph of text against a candidate posting as a paragraph of text (snippet). We conducted evaluation of relevance of syntactic generalization – enabled search engine, based on Yahoo and Bing search engine APIs. For an individual query, the relevance was estimated as a percentage of correct hits among the first thirty, using the values: \{correct, marginally correct, incorrect\}. Accuracy of a single search session is calculated as the percentage of correct search results plus half of the percentage of marginally correct search results. Accuracy of a particular search setting (query type and search engine type) is calculated, averaging through 40 search sessions. For our evaluation we use user postings available at author’ Facebook accounts. The evaluation was conducted by the authors. We refer the reader to (Galitsky et al 2012) for the further details on evaluation settings.
Query type & Query complexity & Relevance of baseline Bing search, %, averaging over 100 searches & Relevance of single-sentence phrase-based generalization search, %, averaging over 100 searches & Relevance of thicket-based phrase generalization search, %, averaging over 100 searches & Relevance of parse thicket-based graph generalization search, %, averaging over 100 searches
---
Facebook friend agent support search
1 compound sent & 74.5 & 83.2 & 85.3 & 87.2
2 sent & 72.3 & 80.9 & 82.2 & 83.9
3 sent & 69.7 & 77 & 81.5 & 81.9
4 sent & 70.9 & 78.3 & 82.3 & 82.7
Table 1: Content relevance evaluation results

To compare the relevance values between search settings, we used first 30 search results and re-ranked them according to the score of the given search setting. We use three approaches to verify relevance between the seed text and candidate posting:

1) Pair-wise parse tree matching, where the tree for each sentence from seed is matched with the tree for each sentence in candidate posting mined on the web;
2) The whole graph (parse thicket) for the former is matched against a parse thicket for the latter using phrase-based approach. In this case parse thickets are represented by all paths (linguistically, phrases)
3) The match is the same as 2) but instead of phrases we find a maximal common subgraph (Galitsky et al 2013).

One can observe that unfiltered precision is 58.2%, whereas improvement by pair-wise sentence generalization is 11%, thicket phrases – additional 6%, and graphs – additional 0.5%. One can also see that the higher the complexity of sentence, the higher the contribution of generalization technology, from sentence level to thicket phrases to graphs.

**Evaluation of trust**

In this section we conduct evaluation of how human users lose trust in CASP and his host in case of both content and mental state relevance failures.

Table 2: Evaluation results for trust loosing scenarios.

Trust in the CASP can be lost in the following scenarios:

1) The content CASP is posted is irrelevant to the content of original post by a human friend.
2) It is relevant to the content, but irrelevant in terms of mental state.
3) Both of the above holds: it is irrelevant to the content and the message is in a wrong mental state.
We measure the failure score of an individual CASP posting as the number of failures (between zero and two). In Table 2 we show the results of tolerance of users to the CASP failures. After a certain number of failures, friends lose trust and complain, unfriend, shares negative information about the loss of trust with others and even encourage other friends to unfriend an friend who is enabled with CASP. The values in the cell indicate the average number of postings which failed relevance when the respective event of lost trust occurs. These posting of failed relevance occurred within one months of the experiment run, and we do not access the values for the relative frequency of occurrences of these postings. On average, 100 postings were done for each user (1-4 per seed posting).

One can see that in various domains the scenarios users loose trust in the system are different. For less information-critical domains like travel and shopping, tolerance to failed relevance is relatively high. Conversely, in the domains taken more seriously, like job related, and with personal flavor, like personal life, users are more sensitive to CASP failures and the loss of trust in its various forms occur faster.

For all domains, tolerance slowly decreases when the complexity of posting increases. Users’ perception is worse for longer texts, irrelevant in terms of content or their expectations, than for shorter, single sentence or phrase postings by CASP.

Now we compare relevance in Table 1 and failed relevance in Table 2. Out of hundred posting per user total, averaging between 2 and 3 for each human user posting, failures occur in less than 10 different user postings. Therefore most users did not land to the lost trust. The friends who were lost due to the loss of trust would be “inactivated” anyway because of a lack of attention, and the majority of friends will be retained and “active”.

**Conclusion**

We proposed a problem domain of social promotion and build a conversational agent CASP to act in this domain. CASP maintains friendship and professional relationship by automatically posting messages on behalf of its human host. We demonstrated that a substantial intelligence in information retrieval, reasoning, and natural language-based relevance assessment is required to retain trust in CASP agents by human users.

We performed the evaluation of relevance assessment of the CASP web mining results and observed that using generalization of parse thickets for the seed and candidate message is adequate for posting messages on behalf of human users.

We also evaluated the scenarios of how trust was lost in Facebook environment. We confirmed that it happens rarely enough for CASP agent to improve the social visibility and maintain more friends for a human host than
being without CASP. Hence although some friends lost trust in CASP, the friendship with most friends was retained by CASP therefore its overall impact on social activity is positive.

The robust intelligence is achieved in CASP by integrating linguistic relevance based on parse thickets and mental states relevance based on simulation of human attitudes. As a result, preliminary evaluation showed that CASP agents are trusted by human users in most cases, allowing CASPs to successfully conduct social promotion.

The content generation part of CASP can be used at www.facebook.com/RoughDraftEssay. Given a topic, it first mines the web to auto build a taxonomy of entities (Galitsky 2014) which will be used in the future comment or essay. Then the system searches the web for these entities to create respective chapters. The resultant document is delivered as DOCX email attachment.

In the interactive mode, CASP can automatically compile texts from hundreds of sources to write an essay on the topic (Fig. 6). If a user wants to share a comprehensive review, opinion on something, provide a thorough background, then this interactive mode should be used. As a result an essay is automatically written on the topic specified by a user, and published.

References
National Science Foundation. 2013. Robust Intelligence (RI); www.nsf.gov/funding/pgm_summ.jsp?pims_id=503305&org=IIS
de Salvo Braz, R., Girju, R., Punyanakon, V., Roth, D. and Simmons, M. 2005 An Inference Model for Semantic Entailment in Natural Language, Proc AAAI-05


