

Full Spectrum Opinion Mining: Integrating Domain, Syntactic and Lexical Knowledge

Daniel J. Olsher, Cognitive Science Program, Temasek Laboratories
National University of Singapore, Singapore 117411
dan@intmind.com // olsher@nus.edu.sg

I. INTRODUCTION

Imagine an automobile review containing the text ‘Some say the new Model 4000 has a large gas tank. It also has a large protective undercarriage to go with it.’ Is this a positive review? In off-road driving, a large gas tank might be desirable, and so might a large undercarriage providing protection from trail hazards. In such a context, the likely negative impact of ‘some say’ is overcome and the review represents a favorable one.

With respect to general driving, however, the first attribute is desirable and the second is not. Especially after ‘some say’, in the general driving context this represents a *negative* review.

If NLP systems could better simulate how people would evaluate various states of the world in contexts of interest, this would make it easier to accurately extract embedded sentiments and avoid being led astray by solely linguistic cues. Generally speaking, rarely is any particular thing wholly ‘good’ or ‘bad’; such judgments depend on context and the particular goals at hand. A very compact camera may be ‘good’ for packing in a backpack, but ‘bad’ if one seeks to attach heavy lenses.

If this knowledge could then be combined with ‘full-semantic’ linguistic processing capable of modeling the interplay between lexical and syntactic semantics and then interweaving these with domain knowledge, this would allow the use of important semantic information (including argument and, especially, valence structure) implicit in phrases such as ‘Critics say’ and ‘Despite this’.

The present paper seeks to implement these insights, employing domain models grounded in the INTELNET/COGVIEW ‘energy-based’ knowledge representation formalism ([1], [2]) and the Radical Construction Grammar-based COGPARSE parser ([3]), bringing together concepts, knowledge, language processing, and opinion mining.

Benefits include the ability to determine when apparently negative linguistic structures should be interpreted positively and vice versa, to model in detail the conceptual and information-flow substructures of an opinion, to understand complex conceptual interdependencies greatly affecting the link between words and sentiments, to show how opinions were reached, to precisely determine the conceptual entities

involved in particular sentiments, and to understand tradeoffs and better determine how arguments could be modified in order to generate different sentiments in future.

II. INTELNET ENERGY-BASED KR AND THE COGVIEW FRAMEWORK: INTRODUCTION

The processes described in this paper draw heavily on a novel ‘energy-based’ form of knowledge representation (titled INTELNET) and the associated COGVIEW cognitive-conceptual representation framework (developed in depth in [1] and applied in [2]). As realized here, energy-based KR represents complex concepts (and, critically, the larger systems these concepts underpin) by setting up pathways upon which information (conceptualized as *energy*) may flow between various semantic elements. Rather than use symbolic representations, the core concept is that complex representations can be built up from simpler pieces by connecting them together via energy flows. Each element reached by a certain quantum of energy flow participates in and becomes part of the wider concept representation. Through this mechanism, conceptual connections between simple elements deeply affect the modeling of larger systems. This technique is optimal for modeling domains characterized by nuanced, interconnected semantics (including most socially-oriented AI modeling domains).

Each quantum of energy possesses a scalar magnitude, a valence (binary positive/negative), and an edge history, defined as a list of the edge labels that a particular quantum of energy has traversed in the past. These three elements, taken together, describe the semantics indicated by that quantum of energy.

A. States of Affairs Networks

When modeling dialogue and argumentation semantics, it is important to differentiate between states of the world that are merely *potential*, that is, raised solely for argumentation purposes, and those actually claimed to exist in the real world. COGVIEW handles potential world states through *States of Affairs (SOA) Networks*, used to represent ‘snapshots’ of the world as it is at a particular point in time. Energy arising from States of Affairs is different from that arising from assertions: if the correctness of a world state is refuted, the valences of energies arising from that state will be reversed, generating proper downstream semantics.

III. RELATED WORK

While some previous work has sought to use domain knowledge in the opinion mining process ([4], [5], [6]), this has been generally accomplished via the use of keywords. Past work has employed linguistic structure to group keywords ([7]), but has not directly integrated this linguistic structure with deep syntax-embedded semantics or with nuanced domain knowledge.

The approach here 1) uses rich, interconnected domain models instead of keywords, and 2) employs a linguistic model connecting syntactic semantics with domain model semantics, allowing each to affect the other. This allows significantly more fine-grained sentiment detection and correctly determines sentiment in the face of conflicting linguistic clues, complex constructions negating specific parts of previous arguments, and references to/negation of past arguments *implicit* in input texts.

Keyword models do not permit one concept to affect another in an interconnected manner (i.e. arguments affecting *medical treatment choice* in turn affect both *treatments* and *doctors*) and do not allow the precise calculation of just *why* particular keywords tend to lead to specific sentiments (important if those keywords participate in linguistic constructions in complex ways or if concepts interact). The specific impact of conceptual clashes is also difficult to determine (i.e. is 'small camera' worse than 'small jet' in some domain? How much? In relation to which concepts?).

In political domains, keyword-spotting models do not have access to the *opposite* of particular policy choices and cannot easily determine when the structure of argumentation involves implicit reference to such policy choice opposites (as occurs in the example given in this paper). Under keyword approaches it is also difficult to identify references to policy choice components and to relate those components, if identified, to the larger whole of the policy in which they reside, capabilities that real-world rhetorical strategies demand.

Finally, the present approach offers a path for modeling *emotion* within sentiment analysis and addresses the ways in which *cognitive processes* contribute to human interpretation of expressive texts. (cf. [8], [9]).

IV. SYSTEM COMPONENTS

This section provides a brief overview of the components underpinning the capabilities described above.

A. Constructions

In order to build a bridge between opinions and semantics, the system draws on meanings embedded within natural language syntax. This information is stored within a corpus of *constructions*, defined as form-meaning pairings [10]. Parsing is accomplished via the COGPARSE engine [3].

Constructions are structured as a series of *slots*, each of which may be a lexical item (a word) or a semantic category ({animate}, {action}, and so on). Constructions

may be nested within other constructions, permitting a single construction to cover significant stretches of text.

Each construction is labeled with an overall type describing its general semantics. An example would be the construction ACTION-DIRECTION-OBJECT, covering texts such as "pull up the chain", "push down the red lever", and so on, possessing an overall type of {action}.

As employed in opinion mining, COGPARSE constructions also contain information about their inherent semantics in the form of COGVIEW network fragments attached to those constructions. These COGVIEW fragments link to each slot of the construction, specifying how these slots work together to create the semantics inherent to the construction.

Constructions also indicate whether they advance new arguments and how these relate to earlier ones, and whether arguments are being asserted or presented merely as background.

Construction corpora need only have coverage for those text snippets containing the most relevant semantic information expected to appear within a particular domain, and need only be generated once per language.

B. Knowledge Base

In normal operation, the COGPARSE parser sits on top of a knowledge base containing information on general category membership of those lexical items most likely to be found in input texts. The use of semantic information allows textual content to be precisely matched with those constructions most likely to reflect the semantics that were intended by the language user. If such knowledge bases are not available, the system will still function, though the quality of results will depend on the overall level of semantic specificity used to generate the construction corpus.

C. Basic Lexical Valence Database

In order to properly determine proper negativity/positivity for lexical item-derived energy flows, the system draws on a database containing simple lexical positive/negative valence information. This database should only contain words which generally (and unambiguously) tend to present specific valences (i.e. *restrict*: fairly negative, *love*: very positive). Words which are likely to hold different valences in different contexts should not be included in the valence database, as the construction and other mechanisms described here provide a better venue for calculating the actual valence signified by those words as they are used in context. Databases such as these may be automatically induced (see, for example, [11]).

D. Domain Model

A key contribution of the present system is the use of domain models (DM), expressed in the COGVIEW formalism. DMs embody the insight that what is 'good' in one domain may be 'bad' in another, representing minimal subsets of key concepts present in important domains

(particular political debates, commercial domains, and so on). DMs place concepts into relation with one another, specify interrelationships, make clear how information flows between various concepts, and clearly specify how various concepts combine to create final sentiments. In political domains, DMs critically allow the system to reason about policy choices and their opposites, as well as determine some information about the likely outcomes of each policy choice.

Domain models are highly reusable and need only be specified once per domain. DMs may be generated by anyone familiar with a domain by first identifying the core concepts operant in a particular domain and then determining how these concepts interrelate with one another. Correctness may be checked through ‘mini-simulations’, whereby energy is introduced into a particular node and then flows are analyzed for semantic and practical validity.

E. Discourse Stack

Certain constructions signal that new arguments are being introduced to the ongoing discourse. In order that future constructions may be able to refer to the semantics of previously-made arguments, a *discourse stack* is maintained recording the arguments made so far at each point in a given input text.

V. TECHNICAL APPROACH

The system performs the following steps for each sentence in the input text:

A. Construction Identification(Parsing) / Valence Lookup

Sentence text is submitted to the COGPARSE parsing engine, which identifies 1) known constructions and 2) the lexical items and nested constructions that fill those constructions’ slots. COGPARSE outputs *paths*, each representing one construction matching one span of text. Path output includes information on precisely how constructions are nested within one another.

Once matches are found, COGPARSE identifies those paths best covering the most text (ideally one construction per sentence), which then act as the base for the next stage. Valence database lookups are also performed for each lexical item at this stage.

B. Network Binding

As previously noted, COGPARSE constructions contain information about their own inherent semantics in the form of COGVIEW network fragments attached to those constructions. COGVIEW networks link to each slot of the construction and specify how these slots work together to create the semantics inherent to the construction.

Construction COGVIEW networks are bound together according to the syntactic structure of the text. If word W fills slot S in construction C, for example, then W will be bound into C’s COGVIEW network at the location associated with S. If slot S in construction C is filled by another construction C2 (that is, C2 is nested within C),

C2’s COGVIEW network will be bound into C’s COGVIEW network at location S. A detailed example of the binding process is provided in section VII below.

During this stage, construction slots are bound to the lexical items and constructions that fill them and, where necessary, bindings are made to previous arguments on the discourse stack. If constructions are encountered that introduce new arguments, these arguments are pushed onto the discourse stack.

C. Network Traversal

This stage introduces *energy* into the COGVIEW networks of the main construction(s) covering each sentence, which then flows outwards to other linguistic components’ networks, crossing links created in the previous stage.

Traversal both 1) identifies possible paths for energy flow within constructions and to points within other COGVIEW networks, and 2) follows the flow of those energies, recording the network edges and concept nodes encountered along the way, as well as updating the state of specific energy quanta as they traverse relevant graph edges.

Also identified at this stage are COGVIEW *clashes*, which occur when energy arriving at a node meets conflicting (opposing energy valence, for example) energy entering the same node. Clashes and their importance are described in more detail in section VII.

Once energy begins flowing it tends to become more and more widely distributed across the COGVIEW network, such that individual concepts exert complex effects on the operation of the network as a whole. This property is key to the system’s ability to model complex sentiments.

D. Calculation

At this stage, the total amount of energy received by specific DM nodes and the importance of clashes are determined. In order to maintain properly meaningful semantics, energy that arises from potential states of the world (and is therefore subject to potential negation during argumentation) is treated differently than that arising from direct assertions (statements which describe the world as it already is and are not likely to be reversed).

E. Sentiment Determination

Based on 1) the amount and valence of energy reaching particular DM nodes and 2) the COGVIEW graph edges traversed by those energies, the system constructs a composite opinion. Final sentiment values are calculated *pro* and *con* the arguments represented by specific concepts, and the cumulative effect the argument is expected to have on the listener is also determined. Key metrics are also calculated here, including relative argument strength and likelihood of persuasion.

After processing is complete, post-hoc analysis of energy flows permits the identification of ‘hot spots’ - active areas of contention within the problem domain, among other interesting quantities.

VI. ARGUMENTATION: THEORY AND COGVIEW IMPLEMENTATION

In order to demonstrate how the system takes linguistic structure into account and to provide foundation for the example in section VII, this section provides a preliminary COGVIEW-based theory of a common mode of argumentation in which an argument is advanced and then another is made relative to the outcomes claimed for the previous argument in hopes of refuting it. In section VII's example, the reader is told that opponents of public health care say it restricts choice, and an argument is then made that private providers already restrict such choice. The intent is to show that, since the claimed consequences have already taken place without the existence of the claimed necessary conditions (public health care), significant doubt must be cast on the original argument as well as the courses of action it calls for.

In the pattern just mentioned, the semantics of both the first and second *explicit* arguments also include significant *implicit* arguments. As an example, if one states that X will lead to negative consequence Y, this is an implicit argument that X is not desirable. The present system fully takes implicit arguments into account and is able to model their effects on overall argument semantics.

Arguments of the form described here are represented in COGVIEW using up to six main nodes, each linked to a key component of the argument being made. These nodes are IF, THEN, THEREFORE, and, optionally, DEMONSTRATED, FAVORED, and DISFAVORED. The nodes are employed as follows: if the state of affairs (SOA) linked to the IF node should ever be realized, the consequences attached to the THEN node are likely to take place. The next part of the argument suggests that, in order to obtain the most positive outcome, the SOA linked to the THEREFORE node should be realized, pointing to suggested courses of action.

The DEMONSTRATED node is employed when a particular state of affairs has previously been threatened to occur, but has now actually been realized. The fact that the threat has become real acts as support for further argumentation.

FAVORED and DISFAVORED point to the practical choices implied by the argument being made. In the example in this paper, those who use the construction OPPONENTS-OF-X-SAY-IT-WILL-Y are implicitly arguing against whatever X is; therefore, X is bound to DISFAVORED. At the same time, the use of this construction also implicitly argues for whatever the *opposite* policy choice is, which is bound in turn to FAVORED.

All six of these node types are demonstrated in the following example.

VII. PUTTING IT ALL TOGETHER: HEALTHCARE DEBATE EXAMPLE

This section provides a detailed, step-by-step example of the system presented in this paper. The example is drawn from Post #4 of the Healthcare section of the Somasundaran

and Wiebe ideological debate corpus [12]. The text of the post is as follows (with header information removed):

Signe Wilkinson. "Unhealthy arguments against public option" Philadelphia Daily News. July 14, 2009: "Opponents of a public option say it will deny Americans the right to choose their own doctors. It's true that some doctors and dentists will choose not to participate in a government-run plan. But many private insurance plans already restrict your choice of doctors by refusing to pay for treatment provided 'out of network.' "

During evaluation, this text was presented to COGPARSE, as is, separated on space boundaries with punctuation removed. A prototype construction corpus was generated without specific reference to the example beyond consideration of which constructions would be minimally necessary.

The semantic information attached to each construction was generated solely by reference to the semantics and pragmatics attached to the corpus constructions in general American English. All constructions are generic and could apply to any English text. The goal was to create construction semantics compatible with any context the constructions might be used within, capable of generating correct results when embedded in (or containing) any other overlapped construction.

Though all sentences were successfully handled by the parser and domain model, due to space limitations this paper analyzes the headline and the first and third sentences (referred to as Sentences 1 and 2, respectively) of the text, as these were identified as most illustrative.

In the following figures, constructions and COGVIEW network fragments associated with each sentence are presented in a series of graphs. The color black is used for fixed, static elements (DMs, constructions and lexical items) with red denoting dynamic elements (such as links between construction slot fillers and COGVIEW nodes) generated at runtime based on input text.

A. Headline

The text of the headline reads: "Unhealthy arguments against public option".

As shown in Figure 1, COGPARSE identifies the constructions DESCRIBED-ARGUMENTS and ARGUMENT-AGAINST-TOPIC as covering the entire text of the headline. The lexical item 'UNHEALTHY' fills the {Description} slot of DESCRIBED-ARGUMENTS, which then fills the {Argument} slot of ARGUMENT-AGAINST-TOPIC.

Via the lexical database, the lexical item 'UNHEALTHY' is associated with negative energy at the standard strength of 10 units. Energy is channeled from the negative energy source to the first subnode of the construction WORD 'UNHEALTHY', which then moves to the last subnode,

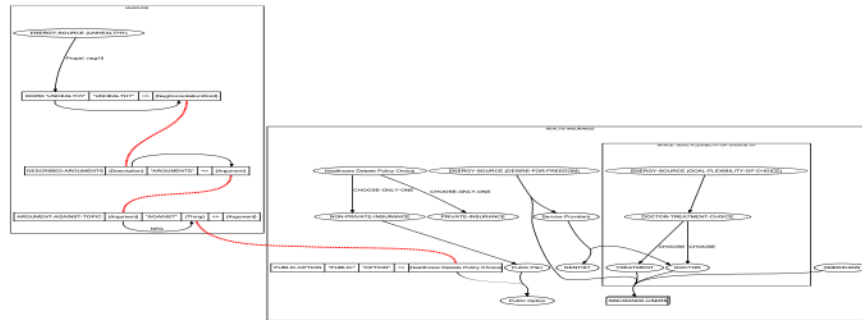


Figure 1. Headline and Domain Model

{NegConnotationWord}, representing the overall semantic type of the construction. From that subnode there is a red (dynamic) binding link to the first slot of DESCRIBED-ARGUMENTS ({Description}), representing the fact that "WORD 'UNHEALTHY'" fills the first slot of the DESCRIBED-ARGUMENTS construction within the input text.

A similar process then occurs with ARGUMENT-AGAINST-TOPIC. The edge spanning the central subnodes of ARGUMENT-AGAINST-TOPIC includes the label 'NEG', indicating that the valence of any energy crossing this edge will be reversed (i.e. -10 becomes +10). Thus, starting from the top left corner, -10 energy traverses through WORD 'UNHEALTHY' to DESCRIBED-ARGUMENTS to ARGUMENT-AGAINST-TOPIC, where its valence is reversed by traversing the 'NEG' edge, becoming +10. This energy then flows across the red binding link to PUBLIC-OPTION and then into the node PUBLIC OPTION.

Moving to interpretation, we note that the final result of energy traversal was the deposition of +10 energy in the node PUBLIC OPTION. As the energy reaching the final node was positive, the system concludes that the overall thrust of the headline was *for* some phenomenon. Moreover, this positive energy was found deposited in the PUBLIC OPTION node in the domain model, suggesting that this was the focus of the positive argument. Overall, then, the sentiment expressed by the headline is understood by the system as a positive argument (of unexceptional strength, however, as energy reached only 10 units - a nominal amount) for the Public Option policy choice, a wholly correct interpretation.

Perhaps most interestingly, *even though the headline contained two words with negative valence and the verbatim phrasing 'against the public option', the system is able to use semantics to determine that the headline in fact represents a positive argument for this policy choice.*

The system determines also that the overall semantic thrust of the headline was {argument} - the overall semantic type of the construction best capable of covering all the words in the headline. Given that construction matching is based in lexical semantic category membership, even though they are generic, constructions that match particular spans

of text are likely to be highly semantically representative of those spans.

B. Sentence 1

Sentence 1 of the text reads as follows: "Opponents of a public option say it will deny Americans the right to choose their own doctors."

The entire sentence is covered by the construction OPPONENTS-OF-X-SAY-IT-WILL-Y (termed the *core construction*). Semantically, this construction introduces a new argument, requiring the addition of a new argument frame onto the stack.

There are three paths for energy flow for the core construction, one beginning at {Healthcare Debate Policy Choice}, one beginning at the construction's start subnode, traversing ARG1-THEN, and ending at AMERICANS, and one beginning at the start, going through ARG1-THEN, and ending at DOCTOR.

Generally, if no energy sources are attached to the first meaningful slot/component of a construction (as is the case for the two paths beginning at the start subnode), a weak positive default source of magnitude 2 is inserted. This default energy is transported to the {Argument} slot, which is linked to ARG1-THEN, indicating that the filler of the {Argument} slot is the linguistic realization of the THEN semantic component of this particular argument.

Energy then travels from ARG1-THEN to the starting subnode of the construction this node is bound to - DENY-X-Y, the first overlapped construction. DENY-X-Y's semantics are structured within an SOA network (denoted by a rectangular box), suggesting that this construction describes a potential state of the world as opposed to something directly asserted. +2 energy flows through the SOA start node, reaching DENY-X-Y's {Animate} slot across a NEG link which flips the valence of that energy. Thus, from here, -2 energy is propagated.

The first subpath from {Animate} sends energy to AMERICANS and then INSURANCE-USERS, a node which 'stands in' for the audience of the opinion piece within the domain model.

The second subpath transfers energy from {Animate} to {Thing}, bound in turn to THE-RIGHT-TO-SOA. Via the

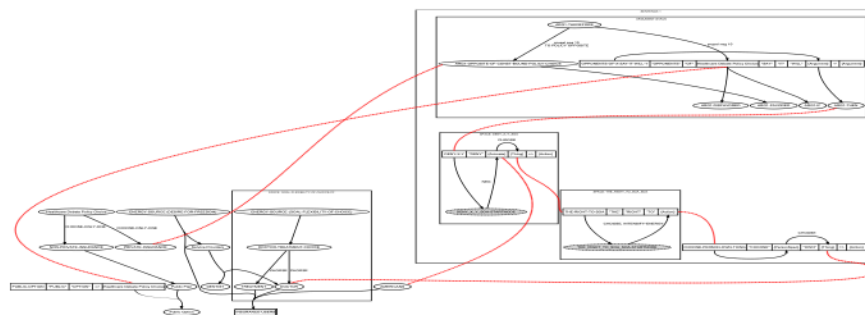


Figure 2. Sentence 1 and Domain Model

SOA start node energy reaches the {Action} slot, crossing an edge marked with the two labels CHOOSE and INTENSIFY-ENERGY. INTENSIFY-ENERGY multiplies energy magnitude by 10, representing the insight that the use of the phrasing ‘right to X’ suggests that X must be very important indeed.

Energy leaving this subnode now has the edge history “CHOOSE” with its magnitude multiplied by 10, yielding a magnitude of -20, which flows to the start subnode of CHOOSE-PERSON-OWN-THING. That energy traverses multiple edges, including a CHOOSE edge, reaching DOCTOR through the binding from the slot {Thing}. The energy then continues on through the network, eventually reaching INSURANCE-USERS.

Energy of -20 with the edge history “CHOOSE CHOOSE” has now reached the node DOCTOR. At this juncture it is key to note that DOCTOR also receives energy through another path; the energy source GOAL-FLEXIBILITY-OF-CHOICE sends positive energy through DOCTOR-TREATMENT-CHOICE and TREATMENT, crossing a CHOOSE edge along the way. Explicit energy sources send magnitude 20 energy by default; thus, energy of +20, with edge history “CHOOSE”, also reaches this node.

In COGVIEW, when energies with opposing valences meet at the same node, this results in a *clash*. Clashes generally always hold highly significant meanings, but in this case, directly opposing energy valences and magnitudes (-20 and +20) with nearly identical edge histories are involved. Matching edge histories evidence incompatibilities between identical semantics in two different portions of the graph, making such clashes especially important.

This particular clash contributes significantly to final sentiment determination; its semantic interpretation suggests that the outcome threatened by opponents (the loss-of-choice state of affairs expressed thus far in the dialogue) would be perceived highly negatively.

The necessity of the domain model is quite evident here; without this, it would be impossible to determine the effect of any particular set of semantics. As an example, the system would not otherwise know that the ability to choose one’s doctor is especially important in the healthcare domain. What about the ability to choose brand-name vs. generic

drugs (important for some) or obtain treatment on Tuesdays (generally less so)? Without domain models, such questions are impossible to answer.

OPponents-OF-X-SAY-IT-WILL-Y provides further useful information by sending +10 SOA-derived energy from ARG1-THEREFORE to the policy choice opposite of that found in the {Healthcare Debate Policy Choice} slot (PRIVATE-INSURANCE) and -10 to the filler of that slot, PUBLIC OPTION. This energy arises from construction semantics implicitly suggesting that the mentioned component (PUBLIC OPTION) is undesirable (disfavored) and favoring the reverse (PRIVATE-INSURANCE). ARG1-FAVORED and ARG1-DISFAVORED act as placeholders for these two components.

By a separate pathway, energy travels from the slot {Healthcare Debate Policy Choice} to ARG1-IF and ARG1-DISFAVORED; these nodes increase in importance when the next construction is processed.

If the dialogue stopped here, the combination of these two highly negative energy flows would lead to a final, highly negative verdict. Sentence 2, however, changes things considerably.

C. Sentence 2

Sentence 2, reading “But many private insurance plans already restrict your choice of doctors.”, is processed through the core construction BUT-ANIMATE-ALREADY-ACTION-THING, the semantics of which act to help refute a previously-offered argument.

This construction contains a semantic consistency test expressed through two edges labeled IDENTITY-ALL-SAME (dotted to indicate they are not energy-conveying). All IDENTITY-ALL-SAME links from the same origin should eventually lead to the same object. In the present example, the first such link leads to PREVIOUS-FAVORED-CHOICE, which is then bound to ARG1-FAVORED, passes through ARG1-OPPOSITE-OF-CONST-BOUND-POLICY-CHOICE and ends up at PRIVATE-INSURANCE. The second link passes through the overlapped MANY-X, and then to PRIVATE-INSURANCE-PLANS and PRIVATE-INSURANCE. Since both paths ultimately end up at PRIVATE-INSURANCE, the test is passed.

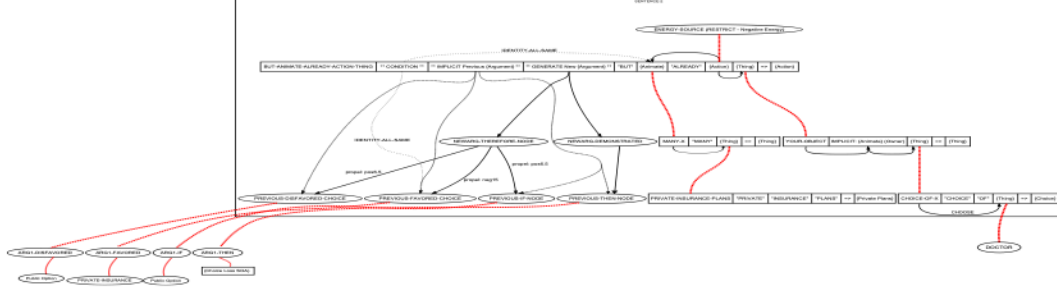


Figure 3. Sentence 2 (with Argument Stack Links)

The passing of such tests, generated *a priori* based on construction semantics and then tested based on the actual word/construction semantics of input texts, serve as significant indicators of model and construction correctness. The fact that the test was passed via such a circuitous route, involving nodes belonging to the *previous argument*, strongly suggests that strong and proper integration exists between the semantics of the two sentences and that the constructions selected by COGPARSE to match and represent them are correct.

The semantics of BUT-ANIMATE-ALREADY-ACTION-THING generate the following argument: Because the consequences threatened by the previous argument (denoted by NEWARG-DEMONSTRATED) have already taken place, even though the posited necessary (IF-) conditions have not been met, consider the original argument as unpersuasive, and reverse the original allocation of FAVORED and DISFAVORED policy options.

BUT-ANIMATE-ALREADY-ACTION-THING references and seeks to refute aspects of an argument which has come before. This reference is accomplished through binding to the argument stack via PREVIOUS-IF-NODE, PREVIOUS-THEN-NODE, PREVIOUS-FAVORED-CHOICE, and PREVIOUS-DISFAVORED-CHOICE, inserted as part of the construction semantics and bound to the corresponding nodes within the most recent appropriate argument.

Energy flow for Sentence 2 begins with BUT-ANIMATE-ALREADY-ACTION-THING. This construction participates in three separate energy flow paths, two originating with the negative energy source auto-generated via the slot filler 'RESTRICT' in the {Action} slot, which then flow to {Animate} and {Action}, and one set of paths originating from the new argument advanced by the construction.

From {Animate}, -18 energy flows to MANY-X, to PRIVATE-INSURANCE-PLANS, and then to PRIVATE-INSURANCE. From {Action}, the same -18 energy flows to YOUR-OBJECT and CHOICE-OF-X, crossing a CHOOSE edge and ultimately ending up at DOCTOR, generating a clash similar to that observed in the first sentence, but with only one CHOOSE edge crossed, no energy intensification, and significantly weaker energy magnitude.

The new argument binds NEWARG-DEMONSTRATED to

a series of constructions denoting restriction of choice (the putative consequences that were successfully demonstrated). The construction then sends +10 energy to the previously disfavored choice (the public option), and -10 energy to the previously favored choice (private insurance), reversing the previous argument.

VIII. CALCULATING FINAL EXPRESSED OPINION: THEORY

Once processing is complete, energy is tallied and final opinions are generated.

A. Calculating the Contribution Made By Clashes

Clashes participate in final energy accounting through *significators*, which estimate the relative energy contribution of a clash. Significators accrue to those nodes responsible for generating the energy flows resulting in a clash, signaling these nodes' semantic incompatibilities with states of affairs located elsewhere in the network.

SOA valence reversal applies to significators whose energy arises from arguments that are ultimately refuted. If a particular SOA results in a clash with a -20 significator associated with node X, for example, upon SOA reversal (that is, refutation) X will receive +20 energy.

Significators are calculated as follows (assuming two-path clashes for ease of exposition):

$$\text{Significator} = \text{Sign} * (SF + EHF)(MF)$$

where Sign=original sign of energy passing through originator node.

$$\text{Sign Factor}(SF) = \begin{cases} 1.7 & : \text{Neg} \rightarrow \text{Pos} \\ 1.5 & : \text{Pos} \rightarrow \text{Neg} \\ 1 & : \text{Same Signs} \end{cases}$$

The sign factor is higher in cases where negative energy impacts positive as, semantically, this signals situations where desired states of affairs are truly threatened. Results from social psychology (i.e. [13]) suggest that loss and threat are counted as somewhat more significant in opinion generation than is energy that merely blunts the effects of otherwise negative states of affairs.

$$\text{Magnitude Factor}(MF) = \sum_i |\text{Clash Magnitude}_i|$$

$$\text{Edge History Factor}(EHF) = \frac{\text{MaxRep} * \text{CommonEdges}}{1.33}$$

where MaxRep=Max # of times any edge is repeated and CommonEdges= # of common edges.
(ex.: CHOOSE CHOOSE vs. CHOOSE = 1.5).

IX. CALCULATING FINAL EXPRESSED OPINION: DATA

Table I summarizes the energies arising from various paths, with Table II providing final energy tabulations.

Various interesting metrics may be calculated from these tables; for example, the spread between the Public and Private energy values is 300.1, suggesting that a significant amount of energy indeed was involved in generating the opinion that ultimately resulted.

A relative strength measure (of one argument versus another) can be calculated as $\frac{||a|-|b||}{avg(|a|,|b|)}$, yielding a value of .15, suggesting that the argument against Private Insurance was 15% stronger than the argument for the Public Option.

The sum amount of energy reaching the Insurance Users node, representing the ‘opinion maker’, including clash signifiers, is 23.9. This number can be understood as an ‘affectedness score’, suggesting that the target is likely to have taken notice of the argument and considered it important, but not to have been inordinately swayed.

The above metrics likely do not exhaust the potential space of what can be calculated, and further metrics are under active development.

Table I
IDENTIFICATION OF ENERGIES

Headline	Path	Energies		Clashes		SOA	Gen Node	Edge History	MF	SF	EHF	Sign	Signif
		Public	PrHC	Am	Ins Users								
Sentence 1 HC Deb Pol Ch Arg-CHOOSE Arg-Americans	Sf-10	Sf10	Sf10	Sf-20	-20x20 (@DOCTOR)	ARG1-THEN	PubHC	2xCHOOSE x CHOOSE	40	1.7	1.5	-1	Sf-128
Sentence 2 RISTR-MANY-X RISTR-YOUR-OBJ New Argument	A/13	A/18	A/18	A/18	-18x20 (@DOCTOR)	Note (Direct)	PrHC	CHOOSE x CHOOSE	38	1.7	0.75	-1	A/-93.1

Table II
ENERGIES SUMMARY

Node	Asserted Energy	SOA Energy	SOA Reversed	Asserted Clash	SOA Clash	SOA Cl Rev	Total
PubHC	23	-10	10		-128	128	161
PrHC	-36	10	-10	-93.1			-139.1
Am		-2	2				2
I-U	-18	-20	20				23.9

X. CONCLUSION

This paper has put forth a novel method, drawing on domain models, linguistic constructions, specific world knowledge, the COGVIEW framework, and INTELNET energy-based KR for determining the opinions expressed in text. Through an in-depth example, the method has been demonstrated and shown to accurately and correctly analyze the opinions expressed within a particular randomly-chosen text.

Further steps include the development of methods for domain model creation and validation, building standard construction corpora, and the further development of methods for evaluating the semantic correctness of induced opinions.

REFERENCES

[1] D. Olsher, “COGVIEW & INTELNET: Nuanced Energy-Based Knowledge Representation and Integrated Cognitive-Conceptual Framework For Realistic Culture, Values, and Concept-Affected Systems Simulation,” 2012, submitted, available at <http://danielolsher.com/cogview>.

[2] —, “Changing discriminatory norms using models of conceptually-mediated cognition and cultural worldviews,” in *Proceedings, Cognitive Science Society Annual Meeting*. Cognitive Science Society, 2012.

[3] —, “COGPARSE: Brain-Inspired Knowledge-Driven Full Semantics Parsing: Radical Construction Grammar, Categories, Knowledge-Based Parsing & Representation,” in *Advances in Brain Inspired Cognitive Systems*, ser. LNCS, vol. 7366. Springer, 2012, p. 1.

[4] C. Lin and Y. He, “Joint sentiment/topic model for sentiment analysis,” in *Proceedings of the 18th ACM conference on Information and knowledge management*, ser. CIKM ’09. New York, NY, USA: ACM, 2009, pp. 375–384.

[5] K. Dave, S. Lawrence, and D. M. Pennock, “Mining the peanut gallery: opinion extraction and semantic classification of product reviews,” in *Proceedings of the 12th international conference on World Wide Web*. New York, NY, USA: ACM, 2003, pp. 519–528.

[6] Q. Miao, Q. Li, and D. Zeng, “Mining fine grained opinions by using probabilistic models and domain knowledge,” in *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 01*, ser. WI-IAT ’10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 358–365.

[7] Y. Wu, Q. Zhang, X. Huang, and L. Wu, “Phrase dependency parsing for opinion mining,” in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3 - Volume 3*, ser. EMNLP ’09. Stroudsburg, PA, USA: Association for Computational Linguistics, 2009, pp. 1533–1541.

[8] E. Cambria, D. Olsher, and K. Kwok, “Sentic activation: A two-level affective common sense reasoning framework,” in *Proceedings, AAAI-12*. AAAI Press, 2012, pp. 186–192.

[9] —, “Sentic panalogy: Swapping affective common sense reasoning strategies and foci,” in *Proceedings, Cognitive Science Society Annual Meeting*. Cognitive Science Society, 2012, pp. 174–179.

[10] A. Goldberg, *Constructions at Work: The Nature of Generalization in Language*. Oxford, 2006.

[11] E. Cambria and A. Hussain, *Sentic Computing: Techniques, Tools, and Applications (SpringerBriefs in Cognitive Computation)*. Springer, 2012.

[12] S. Somasundaran and J. Wiebe, “Recognizing stances in ideological on-line debates,” in *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*. Los Angeles, CA: Association for Computational Linguistics, June 2010, pp. 116–124. [Online]. Available: <http://www.aclweb.org/anthology/W10-0214>

[13] A. Tversky and D. Kahneman, “Loss aversion in riskless choice: A reference-dependent model,” *The Quarterly Journal of Economics*, vol. 106, no. 4, pp. 1039–1061, 1991.