Text-based Reasoning with Symbolic Memory Model

Kun Tu[1,2] **and Hava Siegelmann**[2]
[1] College of Computer Science and Engineering, South China University of Technology
[2] Department of Computer Science, University of Massachusetts Amherst
hava@cs.umass.edu

Abstract

In this paper, a sentence-based reasoning model is introduced for the prediction of new individual activities by means of memory reconsolidation that enables the integration of incoming evidence with related past experience. Both the evidence and previous experience are stored in extended semantic networks (ESN) as memory. They are then processed in Bayesian networks for inferring new and unified memory. Symbolic approaches, which focus on the structural aspect of language, ensure the correct extraction of the key information of words according to the context. Effective mechanisms for information propagation, Bayesian networks (BN) construction and combination are adopted to enable inference reasonable and adaptive to different scenarios based on the topic domain. Our model is compared to other reasoning systems through experiments. The results show that our model can both deduce more implicit information from texts, and avoid some incorrect reasoning caused by confusing data in the knowledgebase.

1 Introduction

Successful reasoning on activities from text should be more adaptive to different scenarios than other data-driven inference models that focus on structure and parameter leaning. For that, a preprocessing must include both the integration with commonsense knowledge and the representation of events after parsing the sentences [Klein and Manning, 2003].

Previous approaches have applied some language processing operations and built effective reasoning models. One method establishes "coreference mappings" of data in a memory system to reduce the number of ambiguous sentence interpretations [Livingston and Riesbeck, 2009]; another builds a large commonsense knowledgebase to make inference from key words [Liu and Singh, 2004]. These methods, although efficient in many cases, lack a systematic approach for disambiguation of word senses and scenario topics [Dahlgren, 1988], and hence may lead to partially incorrect inference.

Additionally, some data from commonsense knowledge are not always true given a specific scenario. For example, the assertion "*if someone is a lawyer, he practices law*" could sound correct, however, there are non-practice lawyers and lawyers are not likely to practice law during "their holidays" [Kathleen and Joyce, 1989].

To improve on the inference of individual activities from simple sentences, we propose a reasoning model with the characteristic of memory reconsolidation. Memory reconsolidation can be thought as an information-processing procedure where a recall of memory can be updated or strengthened as a result of integration of incoming information into the pre-existing "memory network"[Tulving and Thomson, 1973]. During this procedure, information similar in meanings, topics or scenarios within a domain is activated and selectively moved to working memory for cognitive process. Inspired by this phenomenon, we build a robust and scenario-adaptive memory system with extended semantic networks (ESN) as the symbolic representation of the language. Stanford Parser [Klein and Manning, 2003] is embedded to parse sentences. Key information such as subjects, verb and objects is extracted into working memory and referred to WordNet [Fellbaum, 1998] and VerbNet [Kipper *et al.*, 2006] for disambiguation of word senses and topics. Ontology categories of noun phrases, built from WordNet and stored in directed acyclic graphs, describe the features of entities of individual subjects. The relations of entities(representing activities, events or features of entities), are stored as sub-symbolic memory in multiple Bayesian networks (BN). The BNs are built from a)statistical analysis of information in ESN, b)sentences containing causal relation and c)other knowledge bases. Related BNs are adaptively selected by matching the keywords of topics in a scenario, and are combined together if they have nodes in common.

The selection and combination of BNs enable inference processes to avoid inaccurate conclusion and obtain more implicit information than previous models.

2 Related Work

We use the following programs to enable our memory model:

1) Stanford parser, a Java program, is used to analyze the grammatical structure of sentences and obtain subjects, objects and predicates [Klein and Manning, 2003]. 2) WordNet is an English lexical database that provides multiple meanings and topic domains of a word. It also groups words into sets of cognitive synonyms and indicate the relations of the words. (e.g. "dog" belongs to "canine", "canine" belongs to

"carnivore", "placental" and "mammal".) We use it to construct an acyclic graph for word disambiguation.

3) VerbNet classifies verbs into classes. Each class is assigned with thematic roles, selectional restrictions on the arguments and syntactic frames. For example, the Class "Hit-18.1" has a constraint on the syntax of the class members, showing the subject of "hit" should be a human or animal. Such constraints are used to disambiguate the meanings from WordNet and ConceptNet.

4) ConceptNet provides a commonsense knowledgebase that can describe concepts of nouns and provides causal relations between predicates. The knowledge in ConceptNet is expressed as five-tuple assertions ("relation type", "A", "B", "f", "I"), where "relation type" indicates the relation of "A" and "B", f is the number of times a fact is uttered in their training corpus, i counts how many times an assertion was inferred during the 'relaxation' phase. For example, (CapableOf "animal" "grow" "f=2; i=2;") has a relation type "CapabaleOf", which indicates the category of "animal" is capable to perform the activity "grow". Since some relation types provide causal relations of the verb phrases (VP), we build two-node Bayesian networks from each of the assertions and use them for inference on focused topics.

3 Information Representation

Predicates in sentences describe the features or activities of the subjects. Some predicates can break into a form of "joiner + object", where the joiner contains a verb, indicating the relations between the subjects and the objects (see Figure 1(a)) . From a graphical point of view, a vertex can symbolically represent a subject or an object while an edge can represent their relation. For this reason, our extended semantic networks(ESNs) are used as a memory that represents the information derived from sentences .

The ESNs focus on the relations between the subject and the object. Since the meaning of a word can be captured by the distribution of commonly co-occurring words or phrases [Landauer and Dumais, 1997]. The semantic roles of verbs have been characterized with nouns, and were shown to predict the brain activity associated with the meanings of nouns [Mitchell *et al.*, 2008]. The relation in an ESN contains a verb that helps to understand the subjects in a certain scenario.

3.1 Definition of Extend Semantic Network

Suppose $S = (s_1, s_2, \ldots, s_n)$ is a set of sentences, an extended semantic network to store S is a graph $G' = (V', E')$, where $V' = \{v_k | k = 1, 2, \ldots, m\}$ is a set of vertices representing the concepts of subjects or objects in the sentences in S; $E' = \{e_d(V_i, V_j) | d = 1, 2, \ldots;\ i, j = 1, 2, \ldots, m\}$ is a set of edges representing the relations between vertex V_i and V_j, which are indicated in the sentences in S.

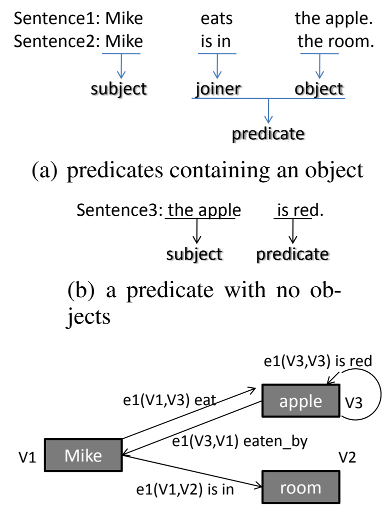
Notice that $e_d(V_i, V_j)$ is the d-th edge from vertex V_i to V_j, which means that there can be more than one relations between two concepts. Vertex V_i can be the same as V_j (that is, $i = j$), in which case it is a unary relation of V_i (e.g. edge $e_1(V_3, V_3)$ in Figure 1(b)) representing a intransitive verb or a property of a vertex.

In our model, the memory system has two kinds of memory: a long-term memory and a working memory. The long-term memory is an ESN that stores all the refined information from working memory. The working memory is a temporarily created ESN and receives information from the Stanford Parser, long-term memory and inferences from BNs. The working memory provides a disambiguation mechanism to ensure the accuracy of information by topic matching. Then, the disambiguated information would be either stored into the long-term memory as an update or propagated to BNs as evidence for inferences.

3.2 Parsing a Sentence

The subject and predicate of an input sentence are first extracted with the Stanford Parser.

Figure 1 shows the storing of sentences in an ESN. In Figure 1(a), the subject "*Mike*" and the objects "*apple*" and "*room*" are stored as vertices. The joiners "*eat*" and "*is in*" in the Verb Phases (VP) are stored as directed edges between vertices (Figure 1(c)).



(a) predicates containing an object

(b) a predicate with no objects

(c) symbolic representation in ESN

Figure 1: Parsing sentences into an ESN.

3.3 The Vertices

Vertices in an ESN denote noun phrases (NP) from subjects and objects in sentences. They represent the entities of people or things in the world. Features and constraints are added to these vertices to ensure the robustness of the memory system.

Ontology Categories

Ontology categories classify the vertices according to the word senses. WordNet provides such hierarchical category data with regard to the word meanings.

Our model uses WordNet to build directed acyclic graphs to represent these categories. Each category has descriptions of its features and activities. Figure 2 shows a sub-graph of the category graph generated from data in WordNet. Suppose "research professor" belongs to the categories of "professor" and "researcher". In WordNet, "professor" is a kind of "academician" who "works at a college or university". After extracting the subject and predicate in the annotation of "professor" with the Stanford Parser, "professor" can be assigned to the category of "academician" and the feature activities are "work at a college" or "work at a university". Similarly, "research professor" is also in the descendant categories of "educator" ("someone who educates young people") and "researcher" ("scientist who devotes himself to research").

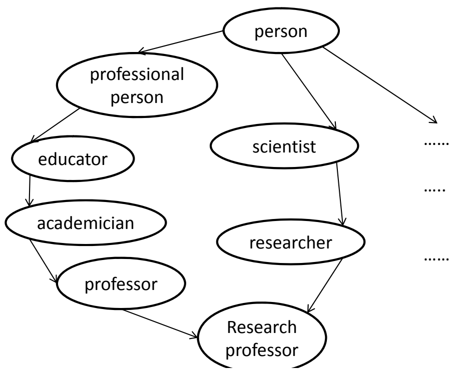


Figure 2: part of a directed acyclic graph for categories

In our category graph, a category can obtain the features or activities either from its parent categories or by importing data from the knowledgebase in ConceptNet.

The ontology categories are important to the entities in working memory because they can help to disambiguate the word sense and provide predicates to connect BNS(see next section).

Possible States of a ESN Vertex

During information processing in the human brain, some neurons are highly active, whereas others remain silent. This leads to the formation of neural circuits for specific memory. In analogy with biological neural memory, a vertex in an ESN has different states: 1)active, 2)semi-active and 3)inactive .

Table 1 shows the differences among the three states of a vertex. An active vertex represents an entity ready for inference. It comes from a subject or an object in a sentence or from the long-term memory. Its meaning is disambiguated and its renewed relations in the working memory with other entities will be directly updated to the long term memory. A semi-active vertex will become an active vertex if it is linked to another active vertex once there exists a new relation between them. An inactivated vertex is not for inference and remains in the long term memory. This can reduce the size of ESNs, which makes computation less complex.

state	Long-term memory	Working Memory	For inference	information update	assigned a category
active	Yes	Yes	Yes	Yes	Yes
semi-active	No	Yes	Yes	No	No
inactive	Yes	No	No	No	No

Table 1: Three states of a vertex in an ESN.

3.4 The Directed Edges

In a sentence, a predicate with an object is regarded as the form of "joiner + object". Edges in the ESNs represent the joiners. They denote the relations between vertices in ESNs and have features that are important to the inference in BNs:

1) **The indicator of the evidence**. Edges are indicated as evidence if they are generated from input sentences. These edges will be directly stored in both the long-term and working memory regardless of their probabilities.

2) **The probability of an edge**. The non-evidence edges are generated from inferences by BNs. Their probabilities can change after each inference. In a working memory, a non-evidence edge is added or removed depending on whether its probability is beyond or below a given threshold. After inferences in the working memory are finished, the non-evidence edges are transported to the long-term memory.

3) **Edges representing a passive voice**. As each BN in our model predicts the activities of only one subject, the interaction of two entities cannot be inferred in one BN. A reversed edge representing the passive voice of a verb is used in the ESN to ensure information can be propagated among the BNs for different entities. For example (Figure 1), "Mike eats the apple". An edge $e_3(V_3, V_1)$ representing "be eaten by" will be created, and a node representing "be eaten by Mike" in a BN if the activities of the apple need to be inferred.

4) **Unary and binary relation**. Edges connecting two different vertices indicate binary relations. Unlike standard SNs, loop edges are allowed as unary relations in our ESN when the predicate has no object(e.g. "stops"), or when the predicate only describes a property of its subject (e.g. "is red" $e_1(V_3, V_3)$ in Figure 1(b)).

5) **Multiple edges between two vertices**. Multiple relations between two vertices are represented as multiple edges in the ESN. In referring to Figure 1(c), suppose also "Mike likes the room", then another edge "like", denoted as $e_2(V_1, V_2)$, would connect "Mike" to the "room".

3.5 Disambiguation for Meanings

Words with multiple senses are classified into multiple categories in WordNet and cause ambiguities. We suggest operations that can disambiguate the word senses:

1) choosing categories with constraints in VerbNet :

For example, "a bat catches insets". The "bat" can be a kind of "mammal" or a kind of "club" for ball game. VerbNet constrain that "Catch" should have a human or animal as agent, so the meaning of "bat" should be a "mammal"(belonging to the "animal" category) not a "club" for sport.

2)selecting the word sense by matching the topics in the working memory and those in BNs:

E.g. "bank" has more than four meanings in its noun form in WordNet and thus can be related to topics such as "deposit", "depository financial institution", "flight maneuver", "slope" and so on. The sentence "he goes to the bank" can be confusing because the "bank" may relate to any of the topics. Two BNs containing conditional probability $P(depositmoney|gotobank)$ and $P(jumpintothewater|gotothebank)$ can limit the meaning to "bank building" and "sloping land beside water" within the topic domains "deposit" or "slope, water". Other information in the working memory (e.g. a vertex representing “river”, has the same topic “water” as “sloping land beside water”) can then decide that the “bank” should be related to a “slope”.

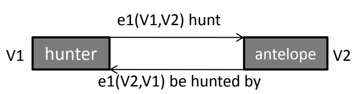
4 An Adaptive BN Mechanism for Reasoning

Our model focuses on update new activities of entities based on new information in memory. Each BN at a time infers activities of only one subject and takes in predicates from working memory to infer new ones.

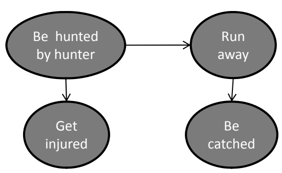
During inference, activities in the working memory are regarded as in the same scenario, and BNs are selected to the topic domain only.

4.1 The Bayesian Network Structure

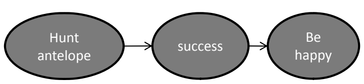
BNs are used in our model to update information in ESNs. Each BN is constructed with the causal relations between predicates. Thus, the agent of the predicate should be the same as the inferred ESN. Each BN has a vector containing words and representing a topic domain. Multiple BNs will be used for different entities in the working memory. Figure 3 shows two BNs for the hunter and antelope.



(a) hunting scenario in an ESN



(b) a BN for the antelope



(c) a BN for the hunter

Figure 3: Propagating information from an ESN to BNs.

The Nodes

We choose the predicates as the nodes of BNs. Figure 3 illustrates how the predicate “be hunted by hunter” is passed from the working memory to a BN as a node .

Evidence in BNs

In BNs, each node has a property called “evidence”. At the beginning of inference, nodes are set to “evidence” if they match the edges of evidence in the working memory. “be hunted by hunter”(Figure 3(b)) and “hunt antelope” (Figure 3(c)) are viewed as evidence.

4.2 BN Construction

We propose a mechanism for automaitc self-generation of BN nodes. There are two ways to construct a BN in our model.

The first way is to learn from experience. After an update in the long-term memory, if two predicates of a subject co-appear for a large number of times, they are considered to have causal relation and are extracted as a predicate pattern pair. The predicate pattern here consists of a verb phrase and the specific object category.

For example, an “alarm sounds” and 8 out of 12 “person”s hear it. The vertex “alarm” and the 12 vertices of different “person” are activated and moved to the working memory. The 12 “person”s are in the same category of human according to WordNet. So the predicate pattern is “be heard by + human”. “Sound” and “heard by human” are converted to BN nodes. Both predicates share “alarm” as their subject.

VerbNet has the syntax restriction to ensure the verb phrase “be heard by” to have a correct “agent” alarm and object “human”. The edge “be heard by human” only goes from the vertex “alarm” to the other 12 “human” vertices. When counting the frequency of “be heard by”, the maximum number should be 12. Thus $p(heardby = true|sound = true)$ = 8/12 =2/3. Another conditional distribution we can get from the data is $p(sound|heardby)$. The conditional distribution is initialized according to the frequency.

The second method is to extract information from other knowledge bases. For example, BN can be constructed from data in the eight of the twenty semantic relations in ConceptNet as described in Table 2.

relation type	probability function	probability value
(PrerequisiteEventOf “A”,“B”)	$P(A = true\|B = true)$	0.9
	$P(A = true\|B = false)$	0.2
(FirstSubeventOf “A”,“B”)	$P(A = true\|B = true)$	0.7
	$P(A = true\|B = false)$	0.3
(EffectOf “A”,“B”)	$P(A = true\|B = true)$	0.7
	$P(A = true\|B = false)$	0.2
(CapableOf “A”,“B”)	$P(A = true\|B = true)$	0.7
	$P(A = true\|B = false)$	0.1
(SubeventOf “A”,“B”)	$P(A = true\|B = true)$	0.9
	$P(A = true\|B = false)$	0.1
(MotivationOf “A”,“B”)	$P(A = true\|B = true)$	0.6
	$P(A = true\|B = false)$	0.1
(DesirousEffectOf “A”,“B”)	$P(A = true\|B = true)$	0.6
	$P(A = true\|B = false)$	0.4
(IsA “A”,“B”)	$P(B = true\|A = true)$	1
	$P(B = true\|A = false)$	0.6

Table 2: Integrate information to a BN from ConceptNet.

Since some commonsense knowledge bases do not provide the probabilities of causal relations, we initialize the probability distributions to default values(Table 2). The probabilities of nodes without a parent are set to 0.5.

BN Selection and Connection

BNs need to be selected and joined together as a new large Bayesian net for the activity prediction.

The BN selection is based on the predicates in the sentences and the topics in the working memory. The first step is to search BNs whose nodes represent the same predicate as in the sentences. E.g. for the sentence "He goes to the bank", BNs containing probability $P(depositmoney|gotobank)$ or $P(jumpintothewater|gotothebank)$ are all selected.

The second step is to remove BNs with a different topic from the working memory. E.g. if the working memory has entities in the categories related to "money" or edges related to financial operation, BNs without these topics are removed.

An adjacency matrix M is built for combining BNs. If BN i and BN j share a common predicate, then $M_{ij} = M_{ji} = 1$, otherwise, the element in the matrix is 0. By multiply the adjacency matrix, a path connecting the BNs can be found. This approach saves us from learning and inferring about unrelated data, and hence reduces much complexity.

The Joint Distribution Function

Suppose the shared node s has parents $s_k \in pa_i(s)$ in BN i and parents $y_l \in pa_j(s)$ in BN j, where $y_l \notin pa_i(s)$. the conditional distribution $p(s|pa_i(s))$ in BN i and $p(s|pa_j(s))$ in BN j are known and the new joint conditional distribution $p(s|pa_i(s), pa_j(s))$ should be initialized for the new network.

5 Making Inference on Evidence

There are several main steps to infer new information from new sentences:

1)A new entity from arriving sentences are automatically created as an active vertex in the working memory. If there are other entities have relation to the new entity, they are brought to the working memory as well.

2)Classify the entities into categories.

3)Predicates in the coming sentences are evidences. New predicates can be inferred from the evidence via BNs. BNs containing the to-be-inferred predicates are target BNs. They are selected for the next step.

4)Find a link path from the target BN to the BN containing evidence with the adjacency matrix. (e.g. fig.5) (If the link does not exist, then the evidence does not affect the activities in that target BN.)

5)Combine the BNs in the path to a large BN. Set the joint conditional distribution values of the share nodes.

6)Search a d-connecting path from the evidence node to the target node.

7)Calculate the probability of the variables in the BN

8)Update the working memory and the long-term memory

6 Case Study

In the following, we will compare our model and Direct Memory Access(DMAP) on the data provided by [Livingston and Riesbeck, 2009]. A scenario is built to compare the inference result of our system and ConceptNet.

6.1 Comparison with DMAP

DMAP uses a story of bombing attack at U.S. soldiers to test how well the model can understand the text. One of the stories is as follow: "An attack occurred in Afghanistan. The bombing was performed by Al Qaeda. The attack occurred on July 18, 2008. The attack targeted United States soldiers."

The aim of DMAP is to integrate the information of "bombing" and "attack" by using a language pattern to map the two words to the same reference in its memory system.E.g, the sentence "the bombing was performed by Al Qaeda" is represented as (performedBy, Bombing-54 Al-Qaeda). DMAP then searches if there are assertions such as (performedBy ?attack Al-Qaeda) to map the two words. DMAP successed in integrating the information only if "bombing" and "attack" share the same assertion pattern (performedBy xxx Al-Qaeda). However, if there is another keyword that share the assertion pattern as "bombing", DMAP would be confused. For example, if the "Al-Qaeda performed a celebration" after the attack, the three word "bombing", "attack" and "celebration" would share the same pattern "performed by Al-Qaeda". Thus, DMAP could falsely reason that "bombing" is equal to "celebration".

In our model, a disambiguation for the word sense is performed in the working memory by using WordNet. The meaning of "bombing" can be referred to "bomb", which has a meaning "throw bombs or attack with bombs". The keyword "attack" in the annotation is extracted by the Stanford Parser and matches the "attack" in the sentences. In this way, the word "celebration" can be omitted as an unrelated word.

6.2 Comparison with ConceptNet

Suppose there are three sentences in a scenario: "Mike swings a bat", "John throws baseball" and "Jones catches the ball". ConceptNet and our model will use the same knowledgebase to reason on the sentences. The following assertions can be found in the database in ConceptNet(explanation of the assertion can be seen in section 2):

1. *(CapableOf "batter" "hit ball" "f=2;i=1;")*
2. *(SubeventOf "play tennis" "hit ball" "f=7;i=1;")*
3. *(CapableOf "baseball player" "hit ball" "f=7;i=1;")*
4. *(CapableOf "bat" "hit ball" "f=2;i=0;")*
5. *(MotivationOf "play tennis" "hit ball" "f=3;i=0;")*
6. *(Isa "batter" "baseball player" "f=2;i=0")*
7. *(CapableOfReceivingAction "baseball pitcher" "throw baseball" "f=2;i=0;")*
8. *(CapableOf "baseball player" "throw ball" "f=3;i=1;")*
9. *(SubeventOf "play baseball" "throw ball" "f=2;i=1;")*
10. *(CapableOf " baseball pitcher" "throw ball" "f=4;i=1;")*
11. *(SubeventOf "play football" "throw ball" "f=2;i=0;")*
12. *(Capableof "cathcer" "catch ball")*
13. *(CapableOf "batter" "swing bat" "f=2;i=0;")*
14. *(Isa "catcher" "baseball player" "f=2; i=0")*

By matching the keywords, ConceptNet can get the following results about "Mike": "Mike might be a batter.(from assertion 13)" "Mike can hit ball.(from assertion 1)", "Mike might be a baseball player.(from assertion 3)" "Mike might play tennis(from assertion 2)" "Mike might be a bat.(from assertion 4)" Notice that "tennis" and "baseball" are different sports and "Mike" cannot play them at the same time. Additionally, ConceptNet cannot infer the exact activities of "John" because there are four different assertions regarding "throw ball"(from assertion 8 to 11).

In our model, the probability generated from data in ConceptNet is initialized as in Table 2. Joint conditional distribution is initialized as follow:$P(S|x_k, y_l) = 0.8$, $P(S|\bar{x_k}, y_l) = P(S|x_k, \bar{y_l}) = 0.55$,$P(S|\bar{x_k}, \bar{y_l}) = 0.1$.

BNs were constructed and combined according to the knowledge base(Figure 4).

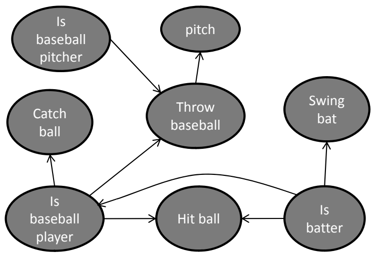


Figure 4: Constructed BN

Table 3 summaries the inferred result. We choose the threshold of probability for adding an edge in memory as 0.65, hence the first three row in table 3 are selected ass new inferred information and added to long-term memory. Notice that "is a baseball pitcher" is not in the D-connecting path given the evidence "swing bat". This means it is independent of "swing bat". Table 4 shows the inference about John's activities. Our concluded inferences, as shown in Table 3 and Table 4, are more reasonable than that in the ConceptNet.

Predicate for "Mike"	probability	value
Is batter	$P(Isbatter)$	0.77
Hit ball	$P(Hitball)$	0.67
Is baseball player	$P(Isbaseballplayer)$	0.86
Throw ball	$P(Throwball)$	0.62
Is baseball pitcher	$P(Isbaseballpitcher$	N/A
Catch ball	$P(Catchball)$	0.62

Table 3: probability of Mike's activities.

Predicate for "Mike"	probability	value
Is baseball player	$P(Isbaseballplayer)$	0.84
Pitch	$P(Pitch)$	0.70
Is baseball pitcher	$P(Isbaseballpitcher$	0.65
Hit ball	$P(Hitball)$	0.61
Catch ball	$P(Catchball)$	0.61
Is batter	$P(Isbatter)$	0.60
Swing bat	$P(Swingbat)$	0.50

Table 4: probability of John's activities.

7 Discussion and Future Work

In this paper, we built a reasoning model to represent and infer new information from texts. Experiments compare DMAP, ConceptNet and our model. Results show that our model is both robust and scalable because the disambiguation mechanism enables it to avoid inaccurate reasoning caused by confusing data. Through calculation, our model can obtain reasonable probabilities of the activities of entities.

The parameters of BNs are set to default values when BNs are first constructed. In the future, we will add a training mechanism of BNs to adjust the parameters after the BN combination. The threshold for adding edges should also be adjusted according to a specific scenario.

8 Acknowledgement

We thank Chinese Scholarship Council and the ONR grant "Symbolic and Continuous Representation in Adaptive Memory with Efficient Inference" number N00014-09-1-0069.

References

[Dahlgren, 1988] K. Dahlgren. *Using Commonsense Knowledge to Disambiguate Word Sense*. Natural Language Understanding and Logic Programming. Amsterdam, North Holland, 2 edition, 1988.

[Fellbaum, 1998] Christiane Fellbaum. *WordNet: An Electronic Lexical Database*. MIT Press, 1998.

[Kathleen and Joyce, 1989] Dahlgren Kathleen and McDowell Joyce. Knowledge representation for commonsense reasoning with text. *Computational Linguistics*, 5(3):149–170, 1989.

[Kipper *et al.*, 2006] Karin Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. Extending verbnet with novel verb classes. In *Fifth International Conference on Language Resources and Evaluation (LREC 2006)*, Genoa, Italy, 2006.

[Klein and Manning, 2003] Dan Klein and Christopher D. Manning. Fast exact inference with a factored model for natural language parsing. pages 3–10, 2003.

[Landauer and Dumais, 1997] T. K. Landauer and S. T. Dumais. A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychol. Rev*, 104:211–240, 1997.

[Liu and Singh, 2004] H. Liu and P. Singh. Conceptnet: A practical commonsense reasoning toolkit. *BT Technology Journal*, Volume 22, 2004.

[Livingston and Riesbeck, 2009] K. Livingston and C. K. Riesbeck. Resolving references and identifying existing knowledge in a memory based parser. In *the AAAI 2009 Spring Symposium*, Learning by Reading and and Learning to Read, California, USA, 2009.

[Mitchell *et al.*, 2008] Tom M. Mitchell, Svetlana V. Shinkareva, Andrew Carlson, Kai-Min Chang, Vicente L. Malave, Robert A. Mason, and Marcel Adam Just. Predicting human brain activity associated with the meanings of nouns. *Science*, 320(5880):1191–1195, 2008. 10.1126/science.1152876.

[Tulving and Thomson, 1973] E. Tulving and D. M. Thomson. Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80(5):352–373, 1973. ISI Document Delivery No.: Q9393 Times Cited: 1541 Cited Reference Count: 91 AMER PSYCHOLOGICAL ASSOC.