A Semantic Approach to Enhance HITS Algorithm for Extracting Associated Concepts using ConceptNet

Madhat Alsoos, Ammar Kheirbek
Faculty of Information Technology Engineering
Damascus University
Syrian Arab Republic
msoossoos@gmail.com, ammar.kheirbek@gmail.com

ABSTRACT: *Common sense knowledge base creation and usage is an active field of research, many researches in this field are trying to invent new methods to make machines more intelligent. ConceptNet is one of the most popular freely available common sense knowledge bases with millions of common concepts and assertions. In this paper, we propose a novel algorithm called SMHITS to extract associated concepts which have conceptual connections with a group of input concepts using ConceptNet knowledge base. We modify HITS algorithm to take into account the semantic meaning of concepts in the context by defining weights for concepts. We also eliminate the domination of the famous concepts that have huge number of predecessors and successors by using weights for hub and authority values. Evaluation shows the superiority of SMHITS over current state of the art methods, with about 10% improvement in performance comparing with AnalogySpace which is the current ConceptNet association method. Evaluation also confirms that this improvement is statistically significant. We use two versions of ConceptNet for evaluation, results show that SMHITS method performance will remarkably be increased with the coming on versions of ConceptNet.*

Categories and Subject Descriptors
I.2.6 [Learning]: Concept learning; I.2.4 [Knowledge Representation Formalisms and Methods]: Semantic networks.

General Terms: Associated Concepts, Semantic Networks

Keywords: HITS, ConceptNet, Associated Concepts, Common Sense, Knowledge Base, Linked Data

1. Introduction

Common sense knowledge is a collection of facts about the everyday words that are possessed by all people. Common sense knowledge represents a huge portion of human experience, encompassing knowledge about the spatial, physical, social, temporal and psychological aspects of typical everyday life. Common sense knowledge creation and usage is a challenging field because of the enormous breadth and details of common sense knowledge. Many tasks like scene understanding, object recognition, machine translation and text mining require the machine to reach the human level of understanding to be done as well as a human being does it. This means that the machine should appear as intelligent as a human being.

Current well-known common sense knowledge bases are: WordNet [1], Cyc [2] and ConceptNet [3]. WordNet is one of the most popular and widely used semantic resource in the computational linguistics community today. It is a database of words that are categorized as nouns, verbs, adjectives and adverbs. WordNet contains only a lexical structural relations such as synonym relation, which makes it semantically poor. Cyc project is one of the oldest active AI project owned by Cycorp. It tries to formalize common sense knowledge in an upper knowledge base for free as a carefully designed ontology named OpenCyc. OpenCyc can be downloaded in OWL format or directly

be accessed as a RDF store using web services from Cycorp. To use Cyc reasoning engine, it is necessary to map the knowledge to its proprietary logical representation using CycL language which is a quite complex process[5]. The difficulty of this mapping and the present unavailability of Cyc full knowledge base to the general public, make Cyc an avoided choice in most cases. ConceptNet is a freely available multilingual semantic network designed for common sense contextual reasoning. ConceptNet has about 23 kinds of semantic relations between concepts which make it a semantically rich knowledge base.

Many researches use common sense knowledge bases in wide different fields like scene understanding [6], search engines [7], natural language processing [8] and ontology matching [9]. Other researches use common sense knowledge base for just a detailed task like finding sentiment orientation of linguistic unit [10]. To have better results, researchers try to use more than one common sense knowledge base like enriching WordNet with ConceptNet knowledge base for word sense disambiguation [11].

In this paper, we use ConceptNet because it is semantically rich and freely available.

2. ConceptNet

ConceptNet is a freely available multilingual common sense knowledge base, representing words and phrases that people use (concepts) and the common sense relationships between them (assertions). The knowledge base is a semantic network presently consisting of about 10 million assertions of common sense knowledge encompassing space, sport, physics, society, psychology and many other domains. The knowledge in ConceptNet is collected from a variety of resources including Open Mind Common Sense Project, WordNet, DBPedia, OpenCyc and Wikipedia[12].

ConceptNet is mainly a network (weighted directed graph) of concepts (nodes) and assertions (edges) between them. Each assertion comes from a particular knowledge source, giving us a reason to believe it. The source also assigns a weight to the assertion, indicating how important and informative that assertion is. ConceptNet has about 23 kind of assertions, the most popular ones are: IsA, PartOf, MemberOf, RelatedTo, HasA, UsedFor, CapableOf, Synonym, Antonym and TranslationOf. Almost all kinds of assertions can be prefixed with "Not" to express a negative assertion such that: NotIsA, NotHasA and NotCabableOf [3,9].

ConceptNet can be downloaded freely from MIT website or can be directly used by ConceptNet Web API that provides a way to execute a lookup query about any concept, search for a list of assertions that match certain criteria and execute an assertion query to find concepts similar to a list of concepts.

Even though ConceptNet is compatible with Semantic Web vision, it is not published as RDF, because edges simply exist in RDF while they have many properties in ConceptNet. Representation of ConceptNet as RDF would be difficult to create and even more difficult to work with. Much of the Semantic Web community has shifted its focus away from particular standards such as RDF and OWL, and now focuses on a broader ecosystem of "Linked Data". ConceptNet supports Linked Data because it meets its view. Linked Data view is that datasets should be represented in a convenient accessible form, which may differ from one dataset to another, and that we should strive to build useful connections between pairs of datasets that have overlapping information and this is an important part of how ConceptNet is built[12].

3. Association in ConceptNet

ConceptNet is different from other knowledge bases like WordNet or OpenCyc even that it depends on them as sources. ConceptNet is based more on context because of its natural language knowledge representation beside its investment in association knowledge. ConceptNet has a high percentage of assertions dedicated to different sorts of generic conceptual connections called knowledge-lines. This allows computers to understand novel or unknown concepts by employing structural analogies to situate them within what is already known [9,11].

ConceptNet has an online web API allows users to directly query the knowledge base with a group of input concepts to get list of ranked associated concepts. Association ConceptNet API depends on AnalogySpace algorithm which makes rough conclusions based on similarities and tendencies of ConceptNet data. ConceptNet is converted to a concept\feature matrix where rows are concepts, columns are features and the value corresponds to the truth value of an assertion. A feature is defined as an assertion with one concept left blank, so each assertion has two features. The dot product of rows represents the similarity between two concepts. It is possible that a concept doesn't have a truth value for a given feature. In this case, AnalogySpace would look up if similar concepts have a truth value for that feature. If they do, it can infer that the original concept probably has that feature too. This makes AnalogySpace inferences about information that is not explicitly in the common sense knowledge base and predicates new assertions.

Concept\feature matrix is a sparse matrix because of the sparse knowledge in ConceptNet. AnalogySpace uses Singular Value Decomposition (SVD) to reduce the dimensionality of concept\feature matrix. The key idea is that semantic similarity can be determined using linear operations over the resulting vectors[13].

4. Review of HITS algorithm

Hyperlink-Induced Topic Search (HITS) algorithm is based on the hyperlinks importance of web algorithm [14]. It is

too popular in Information Retrieval domain, and mainly used to rank webpages depending on their in-links and out- links. Two main terms are used in this algorithm: hubs and authorities. Hubs are webpages that link-to many webpages while authorities are webpages that many webpages are linked-to [15]. Authorities and hubs are illustrated in Figure 1.
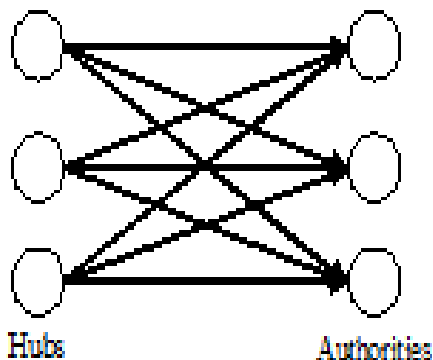

Hubs   Authorities

Figure1. Hubs and Authorities

For each page n let a_n represents authority score and h_n represents hub score. First HITS algorithm set $a_n = h_n = 1$ for all pages, then hub and authority scores are updated in mutually reinforcing way using iteratively process depends on the following equations:

$$a_n = \sum_{m \in P(n)} h_m \qquad (1)$$

$$h_n = \sum_{m \in S(n)} a_m \qquad (2)$$

Where $P(n)$ is the set that has all predecessors pages for page n (pages refer to n) and $S(n)$ is the set that has all successors pages for page n (pages than n refers to).

HITS algorithm is used by search engines to rank crawled webpages based on their similarity to a given query. It ignores page textual content and focuses only on the structure of links between pages [16]. We can roughly say that ConceptNet graph has the same structure with crawled webpages where each concept is like a webpage and each assertion is like a link from webpage to another one. Applying HITS algorithm to ConceptNet graph will not make use of assertions weights nor semantic meaning of them and will also cause a problem of domination of famous concepts because they have a huge number of in-links and out-links.

5. SMHITS Algorithm

In this paper we propose a novel method based on HITS algorithm. Our algorithm is a semantic modification of HITS, we called it SMHITS. The input of the algorithm is a group of concepts which acts as a query of terms in normal search engines where normal HITS algorithm is used. The main goal of SMHITS is to rank ConceptNet concepts based on their association with the input group of concepts. The proposed algorithm takes into account the semantic of relations between concepts and the weights of them during calculating hubs and authorities. To

formalize this let us give the following definitions.

5.1 Definitions
5.1.1. Definition1
The popularity of concept n is measured by two parameters: number of concepts that n refers to (successors) and number of concepts that refer to n (predecessors), we called them O_n and l_n respectively. They are denoted as:

$$O_n = count\,(S\,(n)) \qquad (3)$$

$$I_n = count\,(P\,(n)) \qquad (4)$$

5.1.2 Definition2
Each concept n has a weight C_n to a group of input concepts *seed* which is defined as the sum of the relations types weights R_{Type} for all relations between n and any concept m in *seed* or 1 if this sum is zero. This is denoted as:

$$CSum_n = \sum_{m \in P(n) \cap seed} R_{Type}(m,n) + \sum_{m \in S(n) \cap seed} R_{Type}(n,m) \qquad (5)$$

$$C_n = \begin{cases} CSum_n & if\ CSum_n <> 0 \\ 1 & if\ CSum_n = 0 \end{cases} \qquad (6)$$

Where $R_{Type}(m, n)$ is the weight type for relation between concept m and concept n.

R_{Type} function is used to measure the semantic mean of a relation. R_{Type} weights types of relations that reflect high contextual connection with high positive values while weights others with less values. It also weights "*not*" relations with negative values because they reflect inconsistency between the two connected concepts. In our evaluation we use this function for R_{Type} :

$$R_{Type}(m,n) = \begin{cases} 2 & if\ Type\,(m,n) \in PositiveSet \\ -2 & if\ Type\,(m,n) \in NegativeSet \\ 0 & Otherwise \end{cases} \qquad (7)$$

Where:
PositiveSet = {Synonym, HasA, PartOf, HasSubevent, InstanceOf, MemberOf, SimilarTo}
NegativeSet = {Antonym, NotIsA, NotCapableOf, NotCauses, NotDesires, NotHasA, NotHasProperty, NotMadeOf}

The weight C_n for concept n reflects the semantic of the concept in the context of the group of input concepts, because it depends on type of relations between m and input concepts which is calculated using R_{Type} function.

5.1.3 Definition3
Let us travel over ConceptNet graph starting from concept m. If we are on m that has a successor m, there is a probability $PO\,(m,n)$ to follow relation (m, n) and going to n, which is equal to the percentage between concept m weight C_n and sum of all m successor concepts weights.

Also if we are on m that has a predecessor n, there is a probability $PI(m,n)$ that we followed relation (n,m) and came from n to m, which is equal to the percentage between concept n weight C_n and sum of all m predecessor concepts weights. $PO(m,n)$ and $PI(n,m)$ are denoted as:

$$PO(m,n) = \frac{C_n}{\sum_{k \in S(m)} C_k} \quad (8)$$

$$PI(n,m) = \frac{C_n}{\sum_{k \in P(m)} C_k} \quad (9)$$

Famous concepts like (human, tree,…) have a huge number of in-links and out-links which make them always show up as a good hubs or authorities for almost any group of input concepts. $PO(m,n)$ and $PI(n,m)$ help us removing the domination of these famous concepts by dividing concept hub and authority by outgoing or incoming probability instead of just adding the same value to all concept predecessors or successors.

5.2 Semantically Modified HITS Algorithm

To find the associated concepts for specific group of input concepts, we first extract a related subgraph of ConceptNet graph. The subgraph contains input concepts and all concepts that are connected with any input concept by at least one in-link or out-link. Because ConceptNet is highly interlinked this subgraph contains in average about 2000 concepts for a group of just 3 or 4 input concepts.

We use the same HITS iteratively process to mutually update hubs and authorities for concepts in related subgraph but with new equations. New equations should take into account the semantic of concepts in the context of input concepts and also should divide concept hub and authority values in weighted way between concept predecessors and successors rather than in equality way as in standard equations. This leads to get higher hub and authority values for really associated concepts and solves the problem of famous concepts domination that just have huge number of in-links and out-links without be related to the input concepts context.

Our equations to calculate hubs and authorities are:

$$a_n = C_n \cdot \left| \sum_{m \in P(n)} h_m \cdot \frac{|C_m|}{O_m} \cdot PO(m,n) \cdot W(m,n) \right| \quad (10)$$

$$h_n = C_n \cdot \left| \sum_{m \in S(n)} a_m \cdot \frac{|C_m|}{I_m} \cdot PI(n,m) \cdot W(n,m) \right| \quad (11)$$

Where $W(m,n)$ is ConceptNet weight for assertion (m,n).

In these equations concept weight C reflects the semantic of the concept in the context of input concepts, PO and PI functions values depend on the input concepts because they are calculated using C. To get rid of famous concepts nomination, we divide the hub and authority values of a concept by its number of input or output links (I or O) and multiply them with the probability value to follow the relation between it and the calculated concept (PI or PO).

ConceptNet weight function W allows us to take the assertion source reliability into account. The more positive the weight, the more solidly we can conclude from its source that the assertion is true.

Hub\authority and concept weight both have negative values for out of context concepts. We add the absolute function to avoid getting a positive value from multiplication two negative values, which will lead to a wrong assumption of considering the concept as associated while it is totally out of the context.

After apply our equations in iteratively process to mutually update hubs and authorities for subgraph concepts, we use descending sort for the sum of hub and authority values to rank subgraph concepts. Figure 2 depicts the result of SMHITS comparing with AnalogySpace algorithm for input concepts: car, blood, crowd, scream using ConceptNet 5.2. We can see that our method finds concepts that are more contextually related than AnalogySpace.

6. Evaluation

To evaluate our method we use two versions of ConceptNet, ConceptNet 5.2 which is the current latest version of ConceptNet and ConceptNet 4 which was published in 2012. We use only English concepts of these two versions, ConceptNet 5.2 has more than 2 million English concepts and about 6.4 million assertions while ConceptNet 4 has only about 300,000 English concepts and 1.3 million assertions. Table 1 shows some statistical information about these two ConcetpNet versions.

Property	ConceptNet 4	ConceptNet 5.2
Year of publish	2012	2014
English concepts	313,380	2,049,077
Assertions between English concepts	1,308,436	6,463,725

Table 1. A statistical comparison between ConceptNet 4 and ConceptNet 5.2

Using different versions of ConceptNet allows us to measure the improvement of SMHITS performance while human knowledge in ConceptNet increase in size. To get over incorrect ConceptNet concepts, which is coming from natural language processing errors, we eliminate any English concept that has only one or two assertions. We convert the flat JSON version of each of these versions to normal databases to query them easily.

To evaluate the success of SMHITS we asked users to assess the relevancy of its associated concepts comparing to other state of the art methods. We use 6 sources in our evaluation, first source is ConceptNet 5.2 association method based on AnalogySpace technique which is directly available to use as an online service on: http://conceptnet5.media.mit.edu/data/5.2/assoc.

Second source is SMHITS tested on ConceptNet 5.2 named SMHITS on v5.2 while third source is SMHITS tested on ConcpetNet 4 named SMHITS on v4. We add a fourth source depends on ConceptNet relations, which ranks concepts based on their relations weights with input concepts. We also add a random concepts generator source and manually suggested concepts by people source as two more sources. Table 2 shows some examples of input concepts and top 10 results for AnalogySpace, SMHITS on v5.2, SMHITS on v4 and relations algorithms.

We asked 10 users between 22 and 60 years old toevaluate proposed associated concepts from different 6 sources without any compensation. We provided each user with 20 different groups of input concepts (3 to 4 concepts in each group) covered 20 different contexts. We also provided for each group the top 20 associated concepts for 6 sources shuffled together after removing any redundancy. Users can rate each associated concept with one of the choices: strongly associated with the context, weakly associated with the context, context independent, out of the context. We get from each user one rating for each of 20 top associated concepts from 6 sources for each of 20 input groups of concepts which means 2400 ratings. To describe the validity and reliability of the collected ratings data, we use Intra-Class Correlation Coefficient (ICC). ICC for our users collected ratings data is 0.89 (p<0.0001) which reflects almost perfect consistency and agreement of ratings.
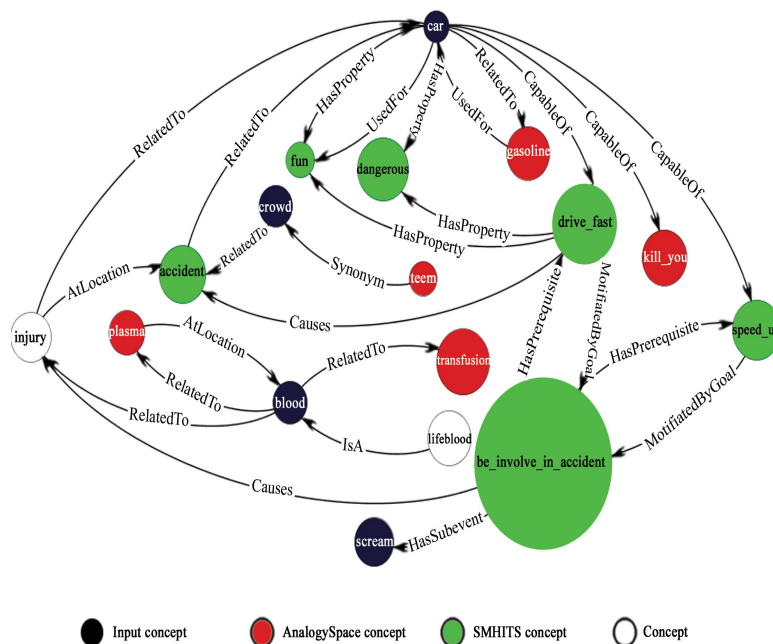

Figure 2. AnalogySpace and SMHITS top results for input concepts: car, blood, crowd, scream using ConceptNet 5.2. Circles represents concepts while arrows represents assertions.

Figure 3 shows the all 24000 ratings collected from all users for 6 sources. As expected the random source gets the worst performance with about 3% strongly associated concepts and 7% weakly associated concepts. Users rated only about 85% of concepts suggested manually by people as strongly associated with concepts, that confirms the diverse understanding of concepts semantic meaning in contexts between different persons. Current ConceptNet association method (based on AnalogySpace) and relations method show almost the same performance with about 44% strongly associated concepts. SMHITS on v4 shows about 42% strongly associated concepts, while SMHITS on v5.2 shows noticeable enhancement with about 54% strongly associated concepts.

In order to quantitatively compare these rating, we map the ratings choices to scores where strongly associated with the context worth 2, weakly associated with the context worth 1, context independent worth 0, out of the context worth -1. Different users used choices very differently that makes some users scores always higher than others regardless the algorithm. To make results independent of that we recalibrate the average score for each user for each source. We linearly map each user's average score to a scale from 0 to 100 where 0 represents that user's average score of random associated concepts and 100 represents that user's average score of manually associated concepts by people.

Table 3 shows the mean and standard deviation for users calibrated average score for all sources except random and manual ones.

Analogy Space and Relations methods have almost the same value (63.0±5.9 and 62.96.3) which is slightly lower than SMHITS on v4 (about 65.5±5.5). The best score is achieved by SMHITS on v5.2 (about 73.1±4.0).

To make sure that the difference between these means is statistically significant we use dependent t-test (paired t-
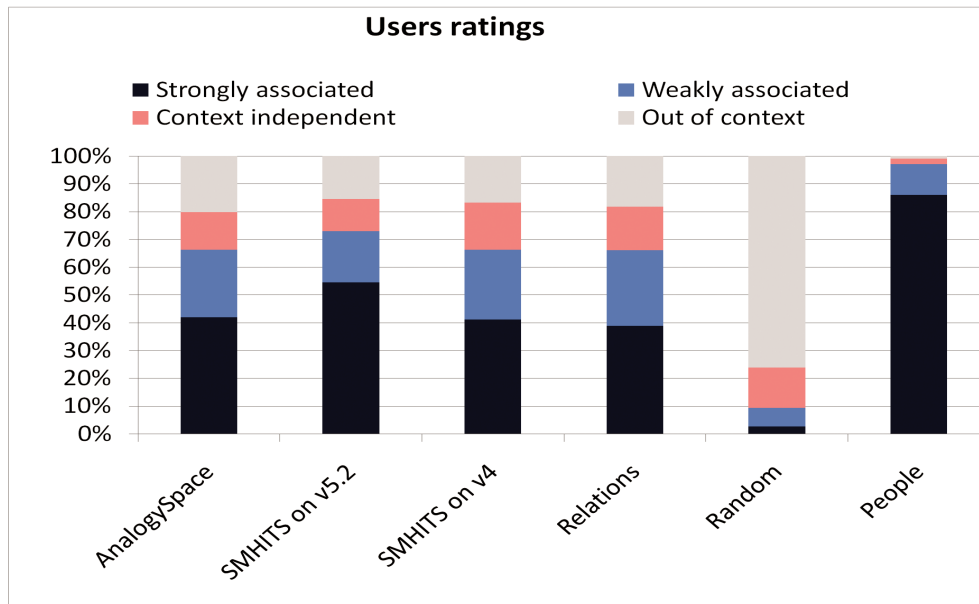
Figure 3. Ratings for AnalogySpace, SMHITS on v5.2, SMHITS on v4, Relations, Random and People sources

Input concepts groups	AnalogySpace	SMHITS on v5.2	SMHITS on v4	Relations
car blood crowd scream	plasma transfusion teem lubricate kill_you gasoline lifeblood restrainer disable red_cell	be_involve_in_accident drive_fast speed_up fun dangerous accident surprise_someone kill car attend_rock_concert	be_involve_in_accident drive_fast speed_up fun dangerous person car water drive window	car_show city crash drive expensive four_wheel freeway go_fast group headlight
run scream shoot somebody human	pump_action somebody gunlock target_practice be_shoot police_station kill_someone high_speed hide_evidence jeer	fight_enemy kill_person stupidity war hate fear soldier danger advance_into_battle gun	fight_enemy kill_person stupidity fear war attack danger fire soldier hate	die_only_once sweat school country home love dog animal die eat
elephant crowd clown	clown person/neg crowd congestion lot many all_together alone/neg large_person lot_person	clown harmonica big_foot carnival delight_child act_silly at_carnival entertain_person fair circus	memory person crowd audience many lot nose many_person group fan	act_silly africa animal big carnival circus city delight_child forget gather

Table 2. Examples of input concepts with top 10 results for AnalogySpace, SMHITS on v5.2, SMHITS on v4 and relations algorithms

test), results are shown in Table 4. Comparing the average calibrated scores for each user for SMHITS on v5.2 and AnalogySpace using paired t-test showed that the 10.0±2.8 difference in their means is statistically significant at p=0.001(<0.05) with a 95% confidence interval of (6.5, 13.5) which gives SMHITS additional inferential power in extract associated concepts over current state of the art. Also applying paired t-test on SMHITS on v5.2 and SMHITS on v4 showed that the 7.5±1.9 difference in their means is statistically significant at p=0.001(<0.05) with a 95% confidence interval of (5.1, 9.8) which indicates that SMHITS results will be improved when knowledge base size is increased. This shows that SMHITS performance can be improved when applying it on coming-on ConceptNet versions that will have much more concepts and assertions.

7. Discussion

Unlike AnalogySpace algorithm which can find similar or synonym concepts to a group of input concepts, we focus on making better inferences to find associated concepts that have conceptual connections with input concepts using the current ConceptNet assertions. This makes our algorithm go behind similar concepts and figure out the context itself and any context related concepts that are not necessary just synonyms. Even that, SMHITS can be tuned to find only similar concepts or to totally eliminate them by changing R_{Type} function. This extends its application domains from text similarity and ontology matching to scene understanding and intelligent robots.

Our evaluation for AnalogySpace on ConcetptNet 5.2 shown in Figure 3 is almost identical with the original AnalogySpace results reported in [17], this shows that AnalogySpace performance doesn't improve when ConceptNet knowledge base is increased in size. In the other hand, testing SMHITS on ConceptNet 4 and ConceptNet 5.2 shows statistically significant improvement of 7.5% in performance, which indicates that it has auto-improvement power when ConceptNet is increased in size.

	Mean	Standard Deviation
AnalogySpace	63.07%	5.98%
SMHITS on v5.2	73.10%	4.05%
SMHITS on v4	65.59%	5.52%
Relations	62.99%	6.34%

Table 3. Comparison between AnalogySpace, SMHITS on v5.2, SMHITS on v4 and relations algorithm using users calibrated average score

	Paired Differences						
				95% Confidence Interval of the Difference			
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	p value
SMHITS on v5.2 − AnalogySpace	10.02%	2.83%	1.26%	6.50%	13.54%	7.90	0.001
SMHITS on v4 − AnalogySpace	2.51%	1.18%	0.52%	1.05%	3.98%	4.76	0.009
SMHITS on v5.2 − Relations	10.11%	2.43%	1.08%	7.09%	13.13%	9.30	0.001
SMHITS on v4 − Relations	2.60%	1.14%	0.51%	1.18%	4.02%	5.10	0.007
SMHITS on v5.2 − SMHITS on v4	7.50%	1.91%	0.85%	5.12%	9.88%	8.76	0.001

Table 4. Comparison between AnalogySpace, SMHITS on v5.2, SMHITS on v4 and relations algorithm using dependent t-test (paired t-test)

SMHITS can be used to push the knowledge base to fill in gaps of missing contextual assertions. This can be done by asking users to evaluate the associated concepts for certain input concepts and use their evaluation to add new assertions in the knowledge base. This enriches ConceptNet knowledge base with new assertions that may not be added by depending only on current sources of ConceptNet. Adding these new assertions will enhance the results not only for SMHITS but also for any other ConceptNet inference method like AnalogySpace.

8. Conclusion

In this paper we propose a novel method named SMHITS

to extract associated concepts with a group of input concepts using ConceptNet. SMHITS is a semantically modified version of HITS algorithm to rank ConceptNet concepts taking into accounts the semantic means of them in the current context. Unlike current association method of ConceptNet which focuses on finding similar concepts, we focus on finding concepts that belong to the same context in people mind. We compare our method with the current association method of ConceptNet which depends on AnalogySpace algorithm. Our evaluation shows that SMHITS has statistically significant improvement of performance over current association method of ConceptNet. We also show that SMHITS has auto-improvement power when knowledge base size increase and its performance can be improved when applying on coming on versions of ConceptNet.

References

[1] Fellbaum, C. (1998). *WordNet: an electronic lexical database*: MIT Press

[2] Guha, R. V., Lenat, D. B., Pittman, K., Pratt, D., Shepherd, M. (1990). Cyc: A Midterm Report. *Communications of the ACM, 33* (8) 33-59.

[3] Liu, H., Singh, P. (2004). ConceptNet — A Practical Commonsense Reasoning Tool-Kit. *BT Technology Journal,* 22 (4) 211-226.

[4] Lenat, D. B. (1995). CYC: a large-scale investment in knowledge infrastructure. *Commun. ACM,* 38 (11) 33-38.

[5] Conesa, J., Storey, V. C., Sugumaran, V. (2010). Usability of upper level ontologies: The case of ResearchCyc. *Data & Knowledge Engineering,* 69 (4) 343-356.

[6] Bicocchi, N., Lasagni, M., Zambonelli, F. (2012). Bridging vision and commonsense for multimodal situationrecognition in pervasive systems. *In:* Proceeding of the IEEE International Conference on Pervasive Computing and Communications (PerCom).

[7] Conesa, J., Storey, V., Sugumaran, V. (2006). Using Semantic Knowledge to Improve Web Query Processing. In Kop, C., Fliedl, G., Mayr, H., Métais, E. *Natural Language Processing and Information Systems*, 106-117

[8] Curtis, J., Cabral, J., Baxter, D. (2006). On the application of the Cyc ontology to word sense disambiguation. *In:* Proceeding of the Proceedings of the 19th International Florida Artificial Intelligence Research Society Conference.

[9] Keshavarz, M., Lee, Y.-H. (2012). Ontology matching by using ConceptNet. *In:* Proceeding of the Asia Pacific Industrial Engineering & Management Systems Conference.

[10] Hung, C. (2008). A personalized word of mouth recommender model. *Webology,* 5 (3) Article 61.

[11] Chen, J., Liu, J. (2011). Combining ConceptNet and WordNet for Word Sense Disambiguation. *In:* Proceeding of the Proceedings of the 5th International Joint Conference on Natural Language Processing. Chiang Mai, Thailand

[12] Speer, R., Havasi, C. (2013). ConceptNet 5: A Large Semantic Network for Relational Knowledge. In I. Gurevych & J. Kim, *The People's Web Meets NLP*, 161-176

[13] Havasi, C., Speer, R., Pustejovsky, J., Lieberman, H. (2009). Digital Intuition: Applying Common Sense Using Dimensionality Reduction. *Intelligent Systems, IEEE,* 24 (4) 24-35.

[14] Desheng, M., Weibo, L., Pin, H. (2013). Optimization of WEB Data Collection Technology Based on HITS Algorithm. *In:* Proceeding of the International Conference on Computer Sciences and Applications (CSA).

[15] Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM, 46* (5) 604-632.

[16] Zhang, X., Yu, H., Zhang, C., Liu, X. (2007). An Improved Weighted HITS Algorithm Based on Similarity and Popularity. *In:* Proceeding of the Second International Multisymposiums on Computer and Computational Sciences.

[17] Speer, R., Havasi, C., Lieberman, H. (2008). AnalogySpace: reducing the dimensionality of common sense knowledge. *In:* Proceeding of the Proceedings of the 23rd National conference on Artificial intelligence. Chicago, Illinois