Building a Concept-Level Sentiment Dictionary Based on Commonsense Knowledge

Angela Charng-Rurng Tsai, *National Taiwan University*

Chi-En Wu and Richard Tzong-Han Tsai, *Yuan Ze University*

Jane Yung-jen Hsu, *National Taiwan University*

Sentiment dictionaries are essential for research in the sentiment analysis field. A two-step method integrates iterative regression and random walk with in-link normalization to build a concept-level sentiment dictionary.

The rise in social media use has changed the role of users from information receivers to information providers. As increasing numbers of people share their ideas, experiences, and opinions on the Web, sentiment analysis has become a popular topic for those who wish to understand public opinion from online data.

A critical step in most approaches to sentiment analysis is using a sentiment dictionary to identify the sentiment units in a document.[1–3] These dictionaries should have a high coverage of words, phrases, and concepts (referred to as *elements*), and their related sentiment information (for example, sentiment polarity and value).

Given the effectiveness of the concept-level sentiment analysis carried out by Erik Cambria and his colleagues,[4] we decided to develop a novel, two-step method that propagates sentiment values based on Concept-Net, a semantic network of commonsense knowledge. Our approach integrates two value-propagation methods: iterative regression and random walk.[5] Unlike previous studies that use mean error for evaluation, we use polarity accuracy, Kendall τ distance, and a new metric (the average-maximum ratio) for large-scale and independent evaluation of sentiment dictionaries.

Proposed Method

Manually annotated sentiment dictionaries such as the *Affective Norms for English Words* (ANEW)[6] contain only about 1,000 elements, which has proven insufficient for several sentiment analysis tasks.[3,4,7] Other approaches[7,8] have used sets of manually annotated synsets and syntactic pair relations between words on WordNet[9] to propagate sentiment values from seed synsets to synsets without values. Cambria and his

Common Approaches to Building a Sentiment Dictionary

Researchers have been building sentiment dictionaries for more than 40 years. One common sentiment dictionary compilation approach has two main steps: sentiment seed collection and sentiment value propagation. In the first step, seeds with accurate sentiment values are collected. Usually, these seeds are manually annotated or collected from existing dictionaries. In the second step, an existing word, phrase, and concept graph is used as the foundation. Sentiment values are propagated from seeds to the remaining parts of the foundation graph.

Thesauruses and commonsense knowledge bases are two commonly used propagation foundations. In a thesaurus, sentiment values can be propagated between *synsets* (synonymous sets of words) using syntactic relations. Unlike a thesaurus, a commonsense knowledge base (such as ConceptNet[1]) maps relations among concepts. Each concept aggregates all possible surface word and phrase expressions. Therefore, when employed in sentiment analysis, concept-based matching has higher recall than word-based matching. In addition, commonsense knowledge bases offer a richer variety of relation types among elements. Therefore, sentiment values can be propagated via not only syntactic relations but also additional relations. For these reasons, we use the commonsense knowledge base ConceptNet as our propagation foundation.

ConceptNet

ConceptNet is a semantic network of commonsense knowledge made up of more than a million simple sentences—such as "A banana is a fruit"—contributed by volunteers on the Open Mind Commons website.[2] The sentences are parsed into assertions such as "banana / IsA / fruit." Each assertion is composed of two concepts ("banana" and "fruit") connected by a relation ("IsA"). There are 31 different types of relations in ConceptNet, including "IsA," "PartOf," and "UsedFor." In total, ConceptNet contains more than 150,000 concepts (nodes) and more than 800,000 assertions (edges). Users can vote for the assertions, and their votes are compiled into a numerical quality score for each assertion. Assertions with higher scores are more likely to be true.

Existing Sentiment Dictionaries Using ConceptNet

Hugo Liu and his colleagues carried out sentiment propagation using the spreading activation approach on ConceptNet. In their approach, each concept contains a floating-point value for six basic sentiments (happy, sad, anger, fear, disgust, and surprise).[3] A bag of affective concepts is chosen, which are preclassified into the six emotions. The propagation begins from these seeds. Each concept with a sentiment value propagates its value discounted by a factor d to its neighboring concepts. When an assertion's polarity is negative, the propagated value is multiplied by (−1). Liu and his colleagues' approach has three passes to increase the coverage of their sentiment dictionary.

SenticNet was developed by Erik Cambria and his colleagues in 2010 and contains 5,273 English concepts.[4] Each concept is labeled with a sentiment value between −1 and +1, which represents the concept's degree of pleasure (the pleasantness-to-unpleasantness dimension).[5] Cambria and his colleagues use the sentiment keywords and the corresponding values[6] as seeds and derive the sentiment values of other concepts on the AffectiveSpace.[7]

The first step of building the AffectiveSpace is to blend WordNet-Affect and ConceptNet using the overlap between them. Then, a matrix C, which describes the relationships among concepts, is built. If C is applied to a single concept, that concept's value is spread to its neighboring concepts. Similarly, applying C^n to a concept spreads the concept's value to all concepts connected by n links. Cambria and his colleagues apply singular value decomposition (SVD) to C before the spreading process to greatly decrease computing time.

References

1. C. Havasi, R. Speer, and J.B. Alonso, "ConceptNet 3: A Flexible, Multilingual Semantic Network for Common Sense Knowledge," *Proc. Recent Advances in Natural Language Processing,* AAAI, 2007; www.media.mit.edu/~jalonso/cnet3.pdf.
2. P. Singh, "The Public Acquisition of Commonsense Knowledge," *Proc. AAAI Spring Symp. Series on Acquiring Linguistic Knowledge for Information Access,* AAAI, 2002; web.media.mit.edu/~push/AAAI2002-Spring.pdf.
3. H. Liu, H. Lieberman, and T. Selker, "A Model of Textual Affect Sensing Using Real-World Knowledge," *Proc. 8th Int'l Conf. Intelligent User Interfaces,* ACM, 2003, pp. 125–132.
4. E. Cambria et al., "SenticNet: A Publicly Available Semantic Resource for Opinion Mining," *Proc. AAAI Fall Symp. Series on Commonsense Knowledge, Assoc. for the Advancement of Artificial Intelligence* (AAAI), 2010; http://aaai.org/ocs/index.php/FSS/FSS10/paper/viewFile/2216/2617.
5. K.R. Scherer, *Psychological Models of Emotion,* Oxford Univ. Press, 2000, pp. 137–164.
6. E. Cambria, A. Livingstone, and A. Hussain, *The Hourglass of Emotions: Cognitive Behavioural Systems,* LNCS 7403, Springer, 2012, pp. 144–157.
7. E. Cambria and A. Hussain, *Sentic Computing: Techniques, Tools, and Applications,* Springer, 2012.

colleagues[4] further enhanced these approaches by using sentic computing[10] to propagate sentiment values based on ConceptNet.[11] The "Common Approaches to Building a Sentiment Dictionary" sidebar further describes these methods.

Our proposed solution is a two-step method. First, we use iterative regression to give each concept on ConceptNet a sentiment value. We then use the values as starting values for our random-walk method with in-link normalization. Figure 1 presents the system flow chart.

We break the process into two separate steps to deal with a particular weakness of random-walk methods: if not enough concepts have nonzero sentiment values, most of the concepts end up with small sentiment values after a few iterations and few retain the
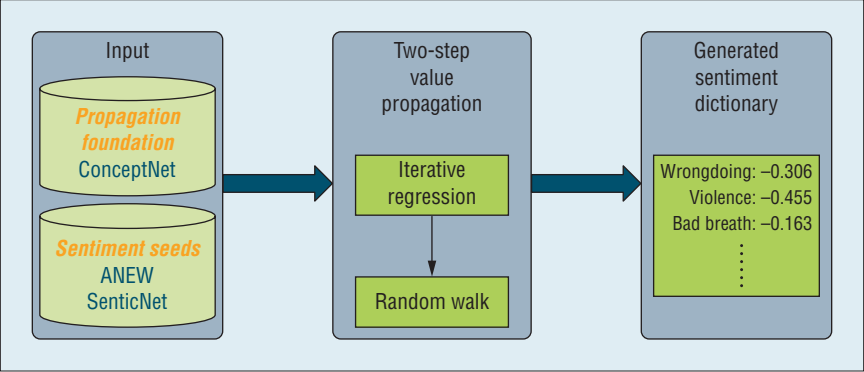
Figure 1. The proposed two-step method. First, iterative regression assigns a sentiment value to ConceptNet's sentiment values. In the second step, these sentiment values are used as starting values for a random-walk method.
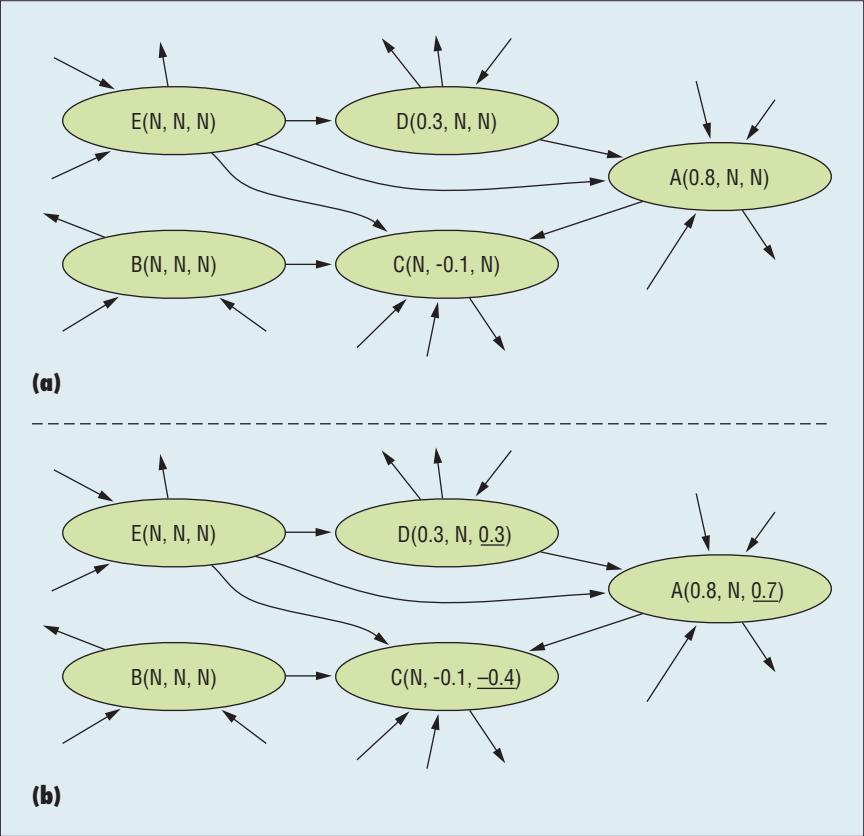

Figure 2. Illustration of value propagation in iterative regression. (a) Concepts A, B, C, D, and E and their Affective Norms for English Words (ANEW), SenticNet, and current value (in parenthesis). (b) Iterative regression step for the concepts.

other nodes' values. In our approach, we have two feature types: features of the concept itself, and features of neighboring concepts. Features of the concept itself contain the concept's value and polarity in the existing sentiment dictionaries. The features of neighboring concepts reflect their statistical distribution.

We must use distributive representation, because if we represented each neighboring concept individually, then each concept would have a different number of features. We generate features using the relation type, polarity, direction, and sentiment value range of an assertion's neighboring concept. For each iteration, we take the value predicted in the previous iteration as the concept's starting sentiment value.

As a hypothetical example to illustrate our iterative regression method's value propagation, Figure 2 presents five highlighted concepts or nodes: A, B, C, D, and E. Each concept letter is followed by parentheses enclosing three possible values: ANEW, SenticNet, and the current value. In Figure 2a, we can see that concepts A and D have ANEW values, while C has a SenticNet value. The surrounding neighboring nodes (linked to, but not shown) are assumed to have empty ANEW values. In each iteration of the iterative regression step (see Figure 2b), concepts B, C, and E receive values.

Random-Walk Step

Random-walk methods are commonly used to spread values on a network. Other researchers have used similar methods for sentiment value propagation when building a sentiment dictionary.[4,12] The equation for random walk[5] is

$$S_{t+1} = C * S_t,$$

where C is the weighted adjacency matrix of the ontology, and S_t is the

sentiment values in the original range. We refer to this as the *insignificant low-value problem*.

Iterative Regression Step

To avoid the insignificant low-value problem, we use a regression model to predict a starting value for each ConceptNet concept. Unlike ordinary regression, we propose using an iterative regression process to capture ConceptNet's graph structure.

We use nodes with values to construct a regression model to predict

value matrix of the tth iteration. Random walk is an iterative process, and after n iterations each concept spreads its value to the concepts that are n links distant from it.

In random walk with restart methods, the value matrix of the $t + 1$th iteration, S_{t+1}, is a linear combination of the value matrix of the tth iteration S_t and the initial value:

$$S_{t+1} = 1(1 - \alpha) * C * S_t + \alpha * S_0,$$

where α is the restarting weight and S_0 is the initial value matrix.

Out-link normalization. The adjacency matrix of a standard random walk is an out-link normalized matrix. That means the value of each node is distributed equally among its neighbors in each iteration. This will cause out-link normalization to underestimate the influence of nodes that have more neighbors in ConceptNet. Furthermore, the propagation process does not limit the final sentiment values within a given range (for example, $[-1, +1]$).

In-link normalization. As we discussed earlier, the random walk with out-link normalization method is unsuitable for spreading sentiment values on ConceptNet. Therefore, we propose using the in-link normalization method. In in-link normalization, each node's new sentiment value in the $t + 1$th iteration is the average of all its neighbors in the tth iteration.

Evaluation

We use three evaluation metrics to analyze the collected data: polarity accuracy, Kendall τ distance, and average-maximum ratio. Polarity accuracy is the polarity correctness of the dictionary, while Kendall τ distance measures the distance between the

dictionary sentiment values and those in the gold standard (where lower is better). The average-maximum ratio for a dictionary D is the ratio of the average absolute value of the sampled sentiment concepts to the maximum absolute value of concepts in D.

To collect ground truth data, we designed two types of human intelligence tasks (HITs) and submitted them to Amazon Mechanical Turk. In addition, we compiled two datasets: our Sentic dataset for comparing with SenticNet, and our Outside dataset for evaluating concepts without sentiment values from other sentiment dictionaries.

Evaluation Metrics

Building a sentiment dictionary with a set of sentiment seeds can be regarded as a value-prediction problem. For such problems, the most common evaluation metric is the mean error metric:

$$MeanError(D,M) = \frac{\sum_{c \in M} |v_c(D) - v_c(M)|}{|M|},$$

where D is the sentiment dictionary, M refers to the concepts in the gold standard dataset (a dataset of concepts with sentiment values), |M| is the size of the gold standard dataset, $v_c(D)$ is the sentiment value of concept c in the dictionary D, and $v_c(M)$ is the sentiment value of concept c marked in the gold standard dataset.

However, a mean error metric requires a gold standard dataset in which the concepts are marked with sentiment values, and this type of dataset is difficult to compile. Therefore, instead of mean error metric, previous work has used polarity accuracy and Kendall t distance to evaluate sentiment dictionaries.[3] Polarity accuracy measures only whether the dictionary concepts' sentiment polarity

matches the gold standard, as the following formula shows:

$$PolarityAccuracy(D,N) = \frac{n_c(D,N)}{|N|},$$

where N is the gold standard dataset (a dataset of concepts with sentiment polarity), $|N|$ is the size of the gold standard dataset; and $n_c(D, N)$ is the number of concepts in the gold standard dataset N whose sentiment polarity is correctly predicted by the dictionary D.

Kendall τ distance evaluates the ranking distance between the dictionary and the gold standard dataset as follows:

$$Kendall\tau(D,Z) = \frac{n_d(D,Z) + p * n_u(D,Z)}{|Z|},$$

where Z refers to the paired gold standard dataset (a dataset composed of randomly selected concept pairs of the same polarity, in which each pair member is labeled as either larger, smaller, or indeterminate); $|Z|$ is the size of the paired gold standard dataset; $n_d(D, Z)$ refers to the number of pairs in which the order (larger to smaller or smaller to larger) is one way in the paired gold standard dataset Z, and the other way in the dictionary D; $n_u(D, Z)$ refers to the number of pairs in the gold standard dataset Z with the same order as in the dictionary D; and p is the weight factor.

In addition, we apply a two-sample t-test to examine whether one system is better than the other with statistical significance. The null hypothesis, which states that there's no difference between the two systems, is given by

$$H_0 : \mu_A = \mu_B,$$

where μ_A is the true mean score of system A, μ_B is the mean of system B, and the alternative hypothesis is

$$H_0 : \mu_A > \mu_B.$$

We apply a two-sample *t*-test, because we assume the samples are independent. Because the number of samples is large and the samples' standard deviations are known, the following two-sample *t*-statistic is appropriate in this case:

$$t = \frac{(\overline{\chi}_A - \overline{\chi}_B)}{\sqrt{\dfrac{S_A^2}{n_A} + \dfrac{S_B^2}{n_B}}},$$

where $\overline{\chi}_A$ is the mean score of system A, S_A is the sample standard deviation of system A, and n_A is the number of datasets for system A. If the resulting *t*-score is equal to or less than 1.67 with a degree of freedom of 29 and a statistical significance level of 95 percent, the null hypothesis is accepted.

According to our findings, polarity accuracy plus Kendall *t* distance is still insufficient for evaluating an automatically constructed sentiment dictionary, mainly because our approach doesn't consider the distribution of sentiment values in the dictionary. Instead, we propose using the average-maximum ratio. Given a dictionary D, the average-maximum ratio is defined as the ratio of the average absolute value of the sampled sentiment concepts to the maximum absolute value of concepts in D. Because the average absolute value is calculated from a set of sentiment concepts, the ratio value should not be too small. Otherwise, that would mean that most sentiment concepts have a smaller sentiment value in scale compared to the maximum value of the dictionary.

Evaluation Data Collection

The Amazon Mechanical Turk platform gives developers access to on-demand crowdsourcing. Requesters can post HITs to the system, which workers can then choose to complete for a small fee paid by the requester.

We proposed two data collection tasks to collect the data for the evaluation metrics. The first task is concept polarity evaluation data collection. We use the positive and negative concepts in the collected data and their polarities in the polarity accuracy metric. We also apply the concepts when calculating the average-maximum ratio. The second task is concept-pair ranking evaluation data collection, which collects the ranking of pairs of concepts for the Kendall τ metric.

To ensure data quality, Mechanical Turk lets requesters set necessary qualifications for workers, and requires requesters to pay only when satisfied with the results. Requesters can set an assignment number (the number of workers required) for each HIT. We requested workers located in the US with more than 50 approved HITs and a 95 percent assignment approval rate. We also hid gold standard data in our assignment data to double check whether the submitted work was accurate.

The gold standard data comes from the aforementioned ANEW,[6] a manually compiled dictionary of 1,034 English words that has been widely used in the sentiment analysis area since its release in 1999. Each word in ANEW is rated in terms of pleasure, arousal, and dominance, with a value ranging from 1 to 9.

Concept polarity evaluation data collection. The HITs collect concept polarity. There are 10 concepts in each HIT and workers are asked to annotate the sentiment polarity of each concept. The workers annotate each concept as either "positive," "negative," "neutral," or "I don't recognize the concept."

Among the 10 concepts in each HIT, we hid one gold-standard positive concept and one gold-standard negative concept. We do this to check

assignment quality. We select the positive concepts from ANEW words with a pleasure value larger than 6, while the negative concepts are from ANEW words with a pleasure value lower than 4. We don't use concepts with values between 4 and 6 because annotators tend to disagree over their sentiment value.

We approve the assignments only if at least one of the gold-standard concepts is labeled correctly and the other concept is not annotated as the opposite of the gold standard. The last dataset consists of a list of positive concepts and a list of negative concepts. We collect three approved assignments for each HIT. Concepts with three positive labels are considered positive concepts, and concepts with three negative labels are considered negative concepts.

Concept-pair ranking evaluation data collection. These HITs are designed to collect the ranking of pairs of concepts for Kendall *t* evaluation. There are 10 pairs of concepts in each HIT. Workers are asked to rate which of the two concepts in each pair is more positive: "the former," "the latter," or "not sure."

We randomly pick the concepts from the dataset of collected polarity HITs. Only concepts with the same polarity are paired. Among the 10 pairs of concepts in each HIT, we randomly place one gold-standard pair from ANEW, in which the former concept is more positive and one in which the latter is more positive to verify assignment quality. The pleasure-value difference between the two concepts in each pair is greater than 2. We approve assignments in which at least one pair is correctly labeled and the other is not labeled in reverse order.

We collect three approved assignments for each HIT. We consider pairs with three "former" labels as

positive-former pairs, and pairs with three "latter" labels as positive-latter pairs.

Evaluation datasets. To evaluate different parts of the dictionary, we collect two datasets: the Sentic and the Outside datasets. The purpose of the Sentic dataset is to compare our dictionary with SenticNet. Therefore, the evaluation dataset is composed of concepts in both SenticNet and our dictionary.

We use the Outside dataset to evaluate the concepts that aren't in SenticNet and ANEW. Because we use only the sentiment information provided in SenticNet and ANEW, and the Sentic dataset contains only concepts from SenticNet, the values of the remaining concepts in each dictionary must be verified when checking each method's effectiveness. The concepts in the Outside dataset are picked from concepts in ConceptNet that are not found in SenticNet or ANEW.

The Sentic and Outside datasets contain 800 and 400 randomly selected concepts with sentiment polarity, respectively. In addition, the datasets each contain 800 pairs in the concept-pair ranking evaluation data.

To perform a *t*-test, we repeated the dataset generation and evaluation processes 30 times and obtained the average scores (polarity accuracy and Kendall τ distance) and standard deviation of scores.

Experiments

The experiment results contain both the single-step and the two-step methods. For the two-step methods, we use the resulting values of iterative regression as the starting values of the ConceptNet concepts. We use concepts in SenticNet or ANEW sentiment seeds. We normalize the pleasure value in ANEW to [−1, 1], the same range of values in SenticNet.

Experiment Settings

The details of our implementation for the two steps are illustrated as follows.

Iterative regression step. We use support vector regression in the Library for Support Vector Machines (LIBSVM) for the iterative regression experiments. We use concepts that occur in both ANEW and ConceptNet for training. We use all other concepts in ConceptNet for testing.

Each concept has two features related to itself: its current sentiment value and its polarity in SenticNet. The sentiment values of concepts in ANEW are normalized to [−1, 1] and fixed at those values in all iterations. For each iteration, the current sentiment values of concepts not in ANEW are set to their predicted sentiment values from the previous iteration.

In addition to self-features, each concept c_i has three groups of neighborhood features: ANEW, SenticNet, and previous iteration statistics (PIS). Our approach generates these three groups of features using the following procedure: For each c_i neighbor c_{ij}, we calculate three 4-tuples (relation type, relation polarity, relation direction, and sentiment interval)— one for ANEW, another for SenticNet, and still another for PIS. The first three 4-tuple elements are set at their original values. If c_{ij} appears in ANEW, SenticNet, and PIS, the value of the fourth element of the ANEW, SenticNet, and PIS tuple is set to c_{ij}'s normalized values for these features. To calculate the normalized value of the aforementioned fourth elements, we map the original sentiment value into one of the 11 intervals. If c_{ij} doesn't appear in ANEW, the fourth component of the ANEW tuple will be set to another interval, and likewise for SenticNet and PIS.

Each of c_{ij}'s three tuples is then converted to a vector of sets. Because there are 31 relation types, two relation polarities, two relation directions on ConceptNet, and 12 intervals, there are $31 \times 2 \times 2 \times 12 = 1,488$ possible tuples, corresponding to 1,488 sets. So, if c_{ij}'s ANEW tuple is (a, b, c, d) and the score for the assertion between c_i and c_{ij} is k, k will be placed in $Set_{(a-1)*2*2*12+(b-1)*2*12+(c-1)*12+d}$.

After converting each of c's three tuples to a vector of sets, we have three vectors: **VS-ANEW**, **VS-SenticNet**, and **VS-PIS**. We sum up all values in each set Set_i and store the summation in an integer vector with 1,488 dimensions. The resulting integer vectors corresponding to **VS-ANEW**, **VS-SenticNet**, and **VS-PIS** are referred to as **VI-ANEW**, **VI-SenticNet**, and **VI-PIS**, respectively. If the m_{th} set in **VS-ANEW** contains the values m_1, m_2, and m_3, the m_{th} dimension of **VI-ANEW** is $m_1 + m_2 + m_3$.

Then, for each vector, each dimension's value is divided by the sum of all its dimensions. In total, we have $1,488 \times 3$ neighbor features.

Random-walk step. For the random-walk methods, we set the restarting weight α to 0.5. When the sum of the square of the change of the sentiment value of all concepts is lower than 0.001, the convergence condition is met.

The seed set of the single-step experiment is made up of concepts that appear in both ANEW and ConceptNet, plus concepts that appear in both SenticNet and ConceptNet. If a concept is in both ANEW and SenticNet, we use the normalized ANEW pleasure value as the concept's value. The proposed two-step method uses the in-link normalization described earlier. However, we also run the out-link normalization experiments for comparison.

Experiment Results

Although both iterative regression and random walk are iterative processes,

Table 1. The *t*-test on polarity accuracy of the Sentic dataset.

Method	Mean	Standard deviation (%)	T	t > 1.67
Iterative regression	0.872	0.61	23.78	Yes
Random-walk out-link	0.879	0.56	19.38	Yes
Random-walk in-link	0.874	0.60	22.40	Yes
Two-step out-link	0.870	0.52	27.56	Yes
Proposed two-step in-link	0.904	0.42	–	–
SenticNet	0.874	–	–	–

Table 2. The *t*-test on Kendall τ distance of the Sentic dataset.

Method	Mean	Standard deviation (%)	T	t > 1.67
Iterative regression	0.413	0.85	26.01	Yes
Random-walk out-link	0.442	1.04	26.84	Yes
Random-walk in-link	0.422	0.74	31.80	Yes
Two-step out-link	0.413	0.87	25.56	Yes
Proposed two-step in-link	0.355	0.88	–	–
SenticNet	0.450	–	–	–

Table 3. The *t*-test on polarity accuracy of the Other dataset.

Method	Mean	Standard deviation (%)	T	t > 1.67
Iterative regression	0.761	1.04	29.12	Yes
Random-walk out-link	0.809	0.97	9.15	Yes
Random-walk in-link	0.830	0.84	–	–
Two-step out-link	0.700	0.91	57.27	Yes
Proposed two-step in-link	0.818	0.76	5.45	Yes

Table 4. The *t*-test on Kendall τ distance of the Other dataset.

Method	Mean	Standard deviation (%)	T	t > 1.67
Iterative regression	0.406	0.85	12.01	Yes
Random-walk out-link	0.485	0.66	5.86	Yes
Random-walk in-link	0.390	0.78	5.54	Yes
Two-step out-link	0.415	0.74	16.54	Yes
Proposed two-step in-link	0.377	1.02	–	–

Table 5. The average-maximum ratios of the Sentic and Outside datasets.

Method	Sentic ratio	Outside ratio
Iterative regression	0.366	0.277
Random-walk out-link	0.149	0.022
Random-walk in-link	0.397	0.041
Two-step out-link	0.005	0.009
Proposed two-step in-link	0.387	0.307
SenticNet	0.485	–

we found that the results change insignificantly after the third iteration. Hence, here we only show the results of iteration three.

Tables 1 and 2 list the result of the *t*-test on polarity accuracy and Kendall τ distance of the Sentic dataset. Because the Sentic dataset's purpose is to compare the quality of our dictionary with SenticNet, we also list the polarity accuracy and Kendall τ distance of SenticNet. The proposed two-step method with in-link normalization has the best polarity accuracy and lowest Kendall τ distance.

Tables 3 and 4 show the result of the *t*-test on polarity accuracy and Kendall τ distance of the Outside dataset. The proposed two-step method with in-link normalization has the lowest Kendall τ distance, and its polarity accuracy is only lower than that of the best method, single-step random walk with in-link normalization.

In Table 5, we list the average-maximum ratios on the Sentic and Outside datasets for each method. The higher the ratio, the fewer concepts have insignificantly low sentiment values. In general, ratios in the Sentic dataset are smaller than those in the Outside dataset. Our proposed two-step in-link method achieves the highest ratios in both datasets.

Discussion

The problem with the single-step random-walk methods is that the sentiment concepts in the Outside dataset have smaller sentiment values in scale compared to the maximum value in the dictionaries. This is why we proposed two-step methods with in- and out-link normalization. The out-link normalization method doesn't get a better result on the average-maximum ratio metric, while in-link normalization and the iterative regression

methods achieve rather reasonable results on the average-maximum ratio metric.

In both the single- and two-step methods, in-link normalization outperforms out-link normalization (see Tables 1–4) on both metrics. These results are consistent with our statements in the "Proposed Method" section. Our proposed two-step method with in-link normalization achieves the top performance in both datasets. Although the single-step method with in-link normalization comes close to our proposed method on some metrics, this approach suffers from the insignificant low-value problem. In addition, the two-step method achieves a much lower Kendall τ distance than the single-step approach. These comparisons demonstrate the two-step method's robustness under various conditions.

The two-step method also outperforms the single-step iterative regression method. We believe that this is because structural information about ConceptNet's graph can't be captured by iterative regression, which only has neighborhood features.

Our two-step method combines iterative regression and random walk with in-link normalization to build a concept-level sentiment dictionary using commonsense knowledge. Compared to single-step iterative regression and random-walk methods, our method achieved the best result. Moreover, our two-step method outperforms the state-of-the-art sentiment dictionary in terms of both polarity accuracy and Kendall τ distance. In particular, Kendall τ distance decreases 22 percent relatively.

Our proposed method still has much room for improvement. Because different relations have different effects on propagating sentiment values, assigning different weights for different relations might be helpful for predicting the sentiment values of unknown concepts more accurately. In addition, the ConceptNet contains a large number of invalid nodes and relations. If we can operate our proposed method on a polished version of the ConceptNet, we believe we can achieve a higher level of accuracy. □

THE AUTHORS

Angela Charng-Rurng Tsai is an engineer at KKBOX, a provider of digital music subscription services in Asia. Her research interests include sentiment analysis, music information retrieval, and applications of commonsense knowledge bases. Tsai has an MS in computer science and information engineering from National Taiwan University. Contact her at r99922050@csie.ntu.edu.tw.

Chi-En Wu is a graduate student at National Taiwan University. His research interests include sentiment analysis, Web interface design, and the development of commonsense knowledge bases. Wu has a BS in computer science and engineering from Yuan Ze University. Contact him at s973309@mail.yzu.edu.tw.

Richard Tzong-Han Tsai is an associate professor in the Computer Science and Engineering Department at Yuan Ze University. His main research interests include natural language processing, cross-lingual information access, sentiment analysis, and context-based recommendation. Tsai has a PhD in computer science and information engineering from National Taiwan University. He is a member of the Association for Computational Linguistics (ACL), Association for Computational Linguistics and Chinese Language Processing (ACLCLP), and the Taiwanese Association for Artificial Intelligence (TAAI). Contact him at thtsai@saturn.yzu.edu.tw.

Jane Yung-jen Hsu is a professor and the department chair of Computer Science and Information Engineering at National Taiwan University (NTU). She is also the director of the Intel-NTU Connected Context Computing Center, featuring global research collaboration among NTU, Intel, and the National Science Council of Taiwan. Her research interests include multiagent systems, intelligent data analysis, commonsense knowledge, and context-aware computing. Hsu has a PhD in computer science from Stanford University. She is a member of AAAI, IEEE, ACM, and Phi Tau Phi. Contact her at yjhsu@csie.ntu.edu.tw.

References

1. G. Mishne, "Experiments with Mood Classification in Blog Posts," *Proc. 2005 ACM SIGIR Workshop Stylistic Analysis of Text for Information Access*, ACM, 2005; http://web.media.mit.edu/~jalonso/cnet3.pdf.

2. F. Keshtkar and D. Inkpen, "Using Sentiment Orientation Features for Mood Classification in Blogs," *Proc. Int'l Conf. Natural Language Processing and Knowledge Engineering* (NLP-KE), IEEE, 2009, pp. 1–6.

3. M. Taboada et al., "Lexicon-Based Methods for Sentiment Analysis," *Computational Linguistics*, vol. 37, no. 2, 2011, pp. 267–307.

4. E. Cambria et al., "SenticNet: A Publicly Available Semantic Resource for Opinion Mining," *Proc. AAAI Fall Symp. Series on Commonsense Knowledge*, Assoc. for the Advancement of Artificial Intelligence (AAAI), 2010; http://aaai.org/ocs/index.php/FSS/FSS10/paper/viewFile/2216/2617.

5. K. Pearson, "The Problem of the Random Walk," *Nature*, vol. 72, no. 1865, 1905; www.nature.com/nature/journal/currentdecade.html?decade=1900&year=1905.

6. M.M. Bradley and P.J. Lang, *Affective Norms for English Words (ANEW) Instruction Manual and Affective Ratings*, tech. report C-1, Center for Research in Psychophysiology, 1999.

7. S. Baccianella, A. Esuli, and F. Sebastiani, "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment

Analysis and Opinion Mining," *Proc. 7th Int'l Conf. Language Resources and Evaluation*, European Language Resources Association (ELRA), 2010; www.lrec-conf.org/proceedings/lrec2010/pdf/769_Paper.pdf.

8. C. Strapparava and A. Valitutti, "WordNet-Affect: An Affective Extension of Word-Net," *Proc. 4th Int'l Conf. Language Resources and Evaluation*, European Language Resources Assoc., 2004, pp. 1083–1086.

9. G.A. Miller, "WordNet: A Lexical Database for English," *Comm. ACM*, vol. 38, no. 11, 1995 pp. 39–41.

10. E. Cambria and A. Hussain, *Sentic Computing: Techniques, Tools, and Applications*, Springer, 2012.

11. C. Havasi, R. Speer, and J.B. Alonso, "ConceptNet 3: A Flexible, Multilingual Semantic Network for Common Sense Knowledge," *Proc. Recent Advances in Natural Language Processing*, AAAI, 2007; www.media.mit.edu/~jalonso/cnet3.pdf.

12. H. Liu, H. Lieberman, and T. Selker, "A Model of Textual Affect Sensing Using Real-World Knowledge," *Proc. 8th Int'l Conf. Intelligent User Interfaces*, ACM, 2003, pp. 125–132.

cn *Selected CS articles and columns are also available for free at http://ComputingNow.computer.org.*

IEEE Intelligence and Security Informatics 2013
Seattle Washington, USA
June 4-7, 2013

Seattle ISI 2013

THEME: Big Data, Emergent Threats and Decision-Making in Security Informatics

HOST: Pacific Northwest National Laboratory
Conference website: www.ISIconference2013.org

Keynote Speakers:
Dr. Joseph Kielman – Department of Homeland Security.
Dr. CAPT Dylan Schmorrow – US Office of the Secretary of Defense.
Dr. William Burns – Decision Science Research Institute Inc.
Dr. Pietro Michelucci – Strategic Analysis, Inc.

Workshops:
Evaluating Visualization in Support of Analytical Reasoning & Decision Making for Cybersecurity
Signature Discovery for Intelligence and Security
Social Media Analytics

Panels:
Open Source Intelligence and Predicting the News
Cyber Security and Information Assurance
Funding Opportunities in Intelligence and Security Informatics