

International Journal of Multimedia Information Retrieval manuscript No.  
(will be inserted by the editor)

# Semantics-Based Selection of Everyday Concepts in Visual Lifelogging

Peng Wang · Alan F. Smeaton

Received: date / Accepted: date

**Abstract** Concept-based indexing, based on identifying various semantic concepts appearing in multimedia, is an attractive option for multimedia retrieval and much research tries to bridge the semantic gap between the media's low-level features and high-level semantics. Research into concept-based multimedia retrieval has generally focussed on detecting concepts from high quality media such as broadcast TV or movies, but it is not well addressed in other domains like lifelogging where the original data is captured with poorer quality. We argue that in noisy domains such as lifelogging, the management of data needs to include semantic reasoning in order to deduce a set of concepts to represent lifelog content for applications like searching, browsing or summarisation. Using semantic concepts to manage lifelog data relies on the fusion of automatically-detected concepts to provide a better understanding of the lifelog data. In this paper, we investigate the selection of semantic concepts for lifelogging which includes reasoning on semantic networks using a density-based approach. In a series of experiments we compare different semantic reasoning approaches and the experi-

mental evaluations we report on lifelog data show the efficacy of our approach.

**Keywords** Concept selection · Lifelogging · Semantics · Everyday concepts · Life log

## 1 Introduction

The idea of digitally recording our everyday lives for our own personal and private use is not a new phenomenon. The writing diary has been used to record the experiences of everyday happenings and has been handed down from generation to generation. With the near-pervasive application of computing technology, the form we can use to record our daily experiences is changing. The earliest motivation for automatic generation of personal *digital* archives can be traced back to 1945 when Vannevar Bush expressed his vision [7] that our lives could be recorded with the help of technology and that access could be made easier to these 'digital memories' via the principle of what became known as hypertext/hypermedia. Automatically generating autobiographies has become more realistic recently with advances in lightweight computing devices and sensors though it is recognised that being able to digitally record everything about our lives doesn't then imply we have perfect recall of everything from our past or that that is even a good idea [36]. Mobile devices have low cost and small and lightweight embedded sensors including cameras, GPS, Bluetooth, accelerometers and gyroscopes, etc. which make computing devices portable thus enabling life recording to be done unobtrusively. Lifelogging is the term describing the notion of digitally recording aspects of our daily lives, where the recorded content in multiple media is a reflection of activities which we subsequently use to obtain an insight into our

---

This work is supported by Science Foundation Ireland under grant number 07/CE/I1147 (CLARITY CSET) and by the China Scholarship Council (File No. 2008611102).

---

Peng Wang  
CLARITY: Centre for Sensor Web Technologies  
School of Computing, Dublin City University,  
Dublin 9, Ireland  
E-mail: pwang@computing.dcu.ie

Alan F. Smeaton  
CLARITY: Centre for Sensor Web Technologies  
School of Computing, Dublin City University,  
Dublin 9, Ireland  
E-mail: alan.smeaton@dcu.ie

daily lives by browsing, searching, or querying. In some cases the motivation for this is work-related, sometimes it is to record special events, and sometimes it is to record for purposes as yet unknown.

Conventional content-based methodologies for retrieval of images or video try to map low-level features to high-level semantics without bridging the semantic gap. This approach has limitations because of the lack of coincidence between low-level features and query semantics. This makes concept-based high-level semantic reasoning an attractive option, where concepts are first detected via a mapping from low-level features using generic methods from training data, then fused together to reason or to infer a final set of concepts which may be used as a representation for whatever the application. Progress in the development of semantic concept detection for video can be seen in the annual TRECVID benchmark [37]. As reported in [22], automatically detected concepts in the TV news broadcasting domain can already be scaled to 1000+, for which 101 concepts are defined in [41] and 834 in [27]; 491 concepts are detected in [39], 374 in [9] and 311 in [22].

However, the large effort in building concept detectors in the TV news broadcast domain cannot be applied directly to the domain of retrieval from everyday lifelog activities. Among the above-mentioned semantic concept sets, the Large-Scale Concept Ontology for Multimedia (LSCOM) is the most comprehensive taxonomy developed for standardizing multimedia semantics in the broadcast TV news domain [27]. As a framework, the LSCOM effort also produced a set of use cases and queries along with a large annotated data set of broadcast news video. However, in the lifelogging domain, many of the LSCOM concepts, for example weapon, government leader, etc., are never normally encountered. In this paper, we investigate the definition of everyday concepts and their automatic detection from visual lifelogs in order to satisfy the requirements for indexing everyday multimedia lifelogs. We also investigate comprehensive ontological similarities for the lifelogging domain.

The rest of the paper is organized as follows: In Section 2, related work on lifelogging research and automatic concept selection in multimedia retrieval is discussed. In Section 3, we select everyday activities in order to construct a semantic concept space for a passive-capture visual lifelogging domain. In Section 4, various semantic similarity measures are investigated on two mainstream semantic networks – WordNet and ConceptNet. Our algorithm for semantic density-based concept selection and ranking is discussed in Section 5, followed by an evaluation in Section 7 in a user experiment. Finally, we close the paper with conclusions.

## 2 Related Work

Many digital devices and sensors are now lightweight and computationally efficient, and can be used to capture the heterogenous contexts of wearers as part of automatic lifelogging. In wearable lifelogging, we can capture visual data using head-mounted cameras [17] or cameras mounted in front of the chest [35]. Using visual information we can infer contextual information like ‘who’, ‘what’, ‘where’ and ‘when’, and using digital cameras or camera-enabled mobile devices is a very attractive form of lifelogging. Visual lifelogging is the term used to describe both image-based and video-based lifelogging. Within the lifelogging community, cameras are often used as wearable devices to record still images [35] or videos [17, 25]. Example visual lifelogging projects are Steve Mann’s WearCam [24], the DietSense project at UCLA [31], the WayMarkr project at New York University [6], the inSense system at MIT [5] and the SenseCam [16] developed at Microsoft Research, Cambridge. Though these projects use various mobile devices for digital logging, they have the common feature of using cameras to capture still images or videos, taken from the first person view of the wearers. Camera-embedded mobile phones are employed in both the DietSense and WayMarkr projects for diet monitoring and experience recall whereas the SenseCam is a sensor-augmented wearable camera designed to capture a digital record of the wearer’s day by recording a series of images and other sensor data. Viewing SenseCam images has been shown recently to be effective in supporting recall of memory from the past for memory-impaired individuals [35]. Due to its advantages of multimodal context-awareness, light weight and unobtrusive logging with long battery life, we employ SenseCam shown in Figure 1, as the visual recording device in our work.



**Fig. 1** The Microsoft SenseCam (right as worn by a user).

The management of visual lifelogging data such as the SenseCam image streams, should involve semantic indexing and retrieval for which much preliminary work has already been done in other domains such as TV news broadcasting. Concept-based information retrieval has received much interest in the community due to its potential in filling the semantic gap and its semantic reasoning capabilities. In concept-based video retrieval, for example, there are known methods to expand query terms into a range of concepts and user judgments and feedback can be used to reveal correlations between concepts. Subjects can be asked to choose the concepts they think are appropriate for their queries. This kind of approach to searching, however, may be difficult for a user when the semantic “space” of concepts becomes large. Previous work has only tested how users use concepts in retrieval based on a small number of concepts and queries as in work by Christel *et al.* [11] for which two collections included 23 queries and 10 concepts together with 24 queries and 17 topics. We also tend to get low inter-annotator agreement, as described by [11]. Previous work on using semantic concepts explicitly from lifelog data has used the distribution of the occurrence of these semantic concepts to profile an individual’s activities and behaviour from a broad, generic viewpoint [12] and this has used a static and pre-defined set of concepts on which to base this.

Automatic approaches to selecting appropriate concepts for semantic querying fall into two categories: lexical and statistical [28]. Lexical approaches leverage the linguistic relationships among semantic concepts in deciding the most related and most useful concepts whereas statistical approaches apply occurrence patterns from a corpus to reveal concept correlations. Statistical approaches also make use of collection-specific associations driven by the corpus set while lexical approaches depend on global linguistic knowledge.

Semantic similarity can be used as a measure to rank the relevance of concepts to a given query text and WordNet [26] is a popular source of the lexical knowledge needed for this. One approach involves selecting concepts based on minimizing the semantic distance between concepts and query terms. WordNet-based semantic similarity between query terms and concepts can be calculated as the weight of concepts using semantic similarity scores and some of the work in the area goes back many years, e.g. [34] and [32]. In more recent work, the Lesk-based similarity measure [4] is demonstrated as one of the best measures for lexical relatedness and is employed in [13] for lexical query expansion. WordNet-based concept extraction is also investigated in [14] to evaluate the effectiveness of high-level concepts used in video retrieval where it achieved

results comparable to user-selected query concepts. The issue with concept selection when using a lexicon ontology such as WordNet is that the local similarities across branches are not uniform which could lead to incomparable similarity values obtained from local ontology branches, as argued in [42]. In work by Snoek [40], Information Content is used to calculate similarity in order to deal with the problem of similarity inconsistency caused by non-uniform distances within the WordNet hierarchy.

A large manual annotation effort in the TRECVID benchmarking activity for video retrieval [29], and in the LSCOM concept ontology for broadcast TV news [1], has enabled the analysis of static patterns for video retrieval. The groundtruth of hundreds of individual concepts and dozens of query annotations is used in comparing retrieval systems as well as selecting and analyzing the relevant concepts associated with particular queries. More recent work by Wei and Ngo [42] proposed an ontology-enriched semantic space model to cope with concept selection in a linear space. The ontological space is constructed with a minimal set of concepts and plays the role of a computable platform to define the necessary concept sets used in video search. This linear space guarantees the uniform and consistent comparison of concept scores for query-to-concept mapping [42].

### 3 Constructing an Event Semantic Space (EES)

One of the limitations in building automatic classifiers or concept detectors for images or videos is for them to reveal higher level semantics when they detect multiple concepts with high correlations. Since the concepts involved in lifelogging cover aspects of our daily lives and thus the range of concepts is very broad, interpreting lifelogging events thus demands a strategy which helps to select the most appropriate combination of concepts for event representation rather than just all possible concepts. We now elaborate the construction of a semantic space reflecting everyday event semantics.

#### 3.1 Everyday Activities: Exploring and Selecting

Patterns of everyday activities are investigated in areas such as occupational therapy and diet monitoring to improve physical and mental health by understanding how we use our time in various activities. Several investigations and surveys have shown that most of our time is spent on activities such as sleeping and resting (34%), domestic activities (13%), TV/radio/music/computers

(11%), eating and drinking (9%), which collectively count for nearly 70% of the time in a typical day.

In [19], the most frequently-occurring everyday activities are explored to rate their level of enjoyment when people experience these activities. 16 activities are investigated and ordered decreasingly by their enjoyment rating. The impact of everyday activities on our feelings of enjoyment also affect our health, which makes these activities important in an analysis of well-being and lifelogging. Similar patterns of activity are also shown in [2], [3] and [10] with sleep being the most dominant activity followed by housework, watching TV, employment/study, etc. [2] and [3] also show that the distribution of activities varies with age group. However, some activities achieve high agreement in participation among all subjects investigated for activities such as sleeping, eating and drinking, personal care, travel, etc.

In constructing an event semantic space for modelling activities, we select our target activities from those with the following criteria:

- **Time dominance:** A small number of activities occupy a large amount of our time and analysis of these activities could maximize the analysis of the relationship between time spent and our well-being.
- **Generality:** Even though the time spent on activities varies from age group to age group, there are some activities that are engaged in by all age groups. The selection of activities with high group agreement will increase the generality of our activity analysis.
- **High frequency:** It is important to select the activities which have enough sample data so we can build automatic classifiers.

With these criteria in mind, we selected the activities listed in Table 1 as targets:

### 3.2 Topic-Related Concepts

How to decide the set of concepts related to the event topics above is the focus of our work. In state-of-the-art everyday concept detection and validation [8], concepts are suggested by several SenseCam users after they have gone through and studied several days’ of their own lifelogs events. Then, being more familiar with their own lifestyles through reviewing their own lifelogs, concepts are discussed and filtered with the added criterion that the concept can be detected with satisfactory accuracy. During this procedure, concepts are not selected in a way that considers the related event topics so concepts may be selected that might not be helpful in interpreting specific event semantics. In addition, some concepts

**Table 1** Target activities for our lifelogging work

1 Eating	2 Drinking	3 Cooking
4 Clean/Tidy/Wash	5 Washing clothes	6 Using computer
7 Watching TV	8 Child care	9 Food shopping
10 General Shopping	11 Bar/Pub	12 Using phone
13 Reading	14 Cycling	15 Pet care
16 Going to cinema	17 Driving	18 Taking bus
19 Walking	20 Meeting	21 Presentation.give
22 Presentation.listen	23 Talking	

which might be helpful in recognizing and interpreting a specific event type may be ignored in the selection procedure. This limits the performance of event detection and semantic interpretation especially when particular concepts relevant to the event are missed. Given the fact that concept detection is always noisy, the situation is compounded when a non-relevant concept is selected to be used in a query, which will reduce the performance by incurring high noise in the query step.

To find a set of candidate concepts related to each of the activities described in Section 3.1, we carried out user experiments on concept selection where candidate concepts related to each of the activities above were pooled based on user investigation. Although individuals may have different contexts and personal characteristics, there is a common understanding of concepts that is already socially agreed and allows people to communicate about these according to [20] and [18]. This makes it reliable for users to choose suitable concepts relevant to activities. User experiments were carried out to discover candidate concepts which potentially have high correlation with activity semantics and details of the experimental methodology will be described in Section 7.1.

The user experiments gave us a set of candidate concepts with regard to the activities we explored in Section 3.1. These concepts were used to construct an event-based semantic space for every activity and the concept space was expanded by each concept as one dimension. Events are represented by groups of images which have their own concepts. We propose a novel semantic density-based concept selection algorithm to find the most useful concepts in the following sections because we believe that existing algorithms are not a

good match for the particular problems of detecting the most appropriate semantic concepts for lifelog events.

## 4 Investigating ESS Concept Relationships

An ontology is used to represent the concepts and concept relations within a domain. Usually ontologies are considered as graphs, where nodes represent concepts and edges represent relations between concepts and in this way the ontology structure captures the semantics for the domain. Ontology-based similarity or relatedness measures can exploit the ontology structure or additional information to quantify the likeness or correlation between two concepts.

### 4.1 Lexical Similarity Based on Taxonomy

Concepts are clustered according to their distribution in the semantic space. With a lack of features or coordinates in this semantic space, concepts can only be clustered in terms of their ontology relationships between each other. As a popular English lexical ontology, WordNet [26] is widely used as a semantic knowledge base. Synsets are the basic elements in WordNet representing the senses of words. The current version (3.0) of WordNet contains 155,327 words grouped into 117,597 synsets. The *is-a* relationship is modeled as hypernymy in WordNet where one concept is more general than another. Hyponymy represents the characteristic that one concept is more specific than another. The meronym or holonymy connection is the semantics representing a *part-of* relationship. This comprehensive coverage and explicit representation of concept relationships make WordNet useful in analyzing the concept relationships within the semantic spaces in our work.

Semantic similarity has been explored in previous research to define a matrix for concept relationship analysis. Rada [30] was first to develop the basis for edge-based measures for concept similarity by defining the distance in a semantic network as the length of the shortest path between the two concept nodes. Richardson and Smeaton [34] further refined the similarity measures. The Hirst and St-Onge [15] similarity measure, takes path direction into account and the idea is that the concepts are semantically close if their WordNet synsets are connected by a short path which does not change direction too often. Another similarity definition is proposed in [43] by Wu and Palmer for verb similarity calculation since most work is built upon noun concepts, and applied in machine translation. This was extended by Leacock and Chodorow [21] also as a path-

based similarity measure which determines similarity with regard to the maximum depth of the taxonomy.

Semantic similarity based on information content is also an important component of lexical relationship analysis. This approach relies on the hypothesis that the more information two concepts share, the more similarity they have. The informativeness of a concept is quantified by the notion of its Information Content (IC), which is calculated based on the occurrence probability of concepts in a given corpus. IC is obtained by negative likelihood of encountering a concept in a given corpus [32]. The basic intuition of using negative likelihood assumes that the more likely a concept appears in a corpus the less information it conveys.

Based on the IC formula, a concept will contain less information if the probability of its occurrence in a corpus is high. The advantage of using information content is that once given a properly constructed corpus, information content can be adapted in different domains because information content is included in a statistical way according to occurrences of the concept, its sub-concepts and subsumers.

In [33], Resnik applied information content to calculating semantic similarity using Most Specific Common Abstract ( $msca(c_1, c_2)$ ) as the amount of information that concepts  $c_1$  and  $c_2$  have in common. In this approach, only the *is-a* relationship is used because only the information of the subsuming concept of the two concepts being compared, is used. In [38], this similarity measurement is also employed by Quigley and Smeaton to compute word-word similarity in image caption retrieval.

### 4.2 Contextual Ontological Similarity and Relatedness

WordNet is a small ontology of primarily taxonomic semantic relations. ConceptNet extended WordNet to include a richer set of relations appropriate to concept-level nodes [23]. In the version of ConceptNet we use, the relational ontology consists of 20 relation types falling into categories like K-lines, Things, Agents, Event, Spatial, Causal, Functional and Affective.

In ConceptNet, all concepts are linked with the above-mentioned relations which can reflect the correlations between concepts. We apply a link-based relatedness measure to maximize the concept relations in measuring concept correlation. This differs from WordNet which uses mainly taxonomic relationships, while ConceptNet employs more context relationships. While WordNet similarities only consider subsumption relations to assess how two objects are alike lexically, relatedness

takes into account a broader range of relations which can be measured using ConceptNet.

The relations between concepts reflect the semantic correlation between two concepts. We assume that semantic relations are transitive so the more related two concepts are, the shorter paths they will have. The relatedness between two concepts varies inversely with the length of the shortest path between the two concepts. Conceptual relatedness is a monotonically decreasing function of path distance. Our approach takes into account the length of paths between two concepts. In ConceptNet, because the edges between concepts are directional, we combine the length of the path between concept  $c_1$  and  $c_2$  as well as path between  $c_2$  and  $c_1$ . The similarity between two concepts is defined as:

$$S_{CN}(c_1, c_2) = \max(AS(c_1, c_2), AS(c_2, c_1)) \quad (1)$$

where  $AS(c_1, c_2)$  represents the activation score of  $c_2$  starting from  $c_1$ , and vice versa. Activation score is computed by spreading activation in ConceptNet to find the most similar concepts with regard to a starting concept. The starting concept is initialized with activation score 1.0 and then the nodes connected with the starting concept with one link path, two link paths, etc. are activated. The activation score of connected node  $b$  with original node  $a$  is defined as:

$$AS(a, b) = \sum_{c \in Neighbor(b)} AS(a, c) \times d \times w(c, b) \quad (2)$$

where  $d$  is a distance discount ( $d < 1$ ) to give the concepts far from the original concept a lower weight and  $w(c, b)$  is the relation weight of the link from  $c$  to  $b$ . In this paper, we apply the same relation weight for activation scores. For any given concept  $b$ , the activation score related to  $a$  is the sum of scores of all nodes connected to it.

## 5 Concept Selection Based on Semantic Density

Our measure of semantic density relies directly on the semantic distance between concepts. If the distance measured between concepts is small, then the concepts have high density. The semantic distance is used as a measure by which concepts are clustered to represent event semantics. In our semantic topic-related concept selection, we deal with the research question by means of identifying the similarity between concepts as a linguistic problem. Processing consists of text pre-processing which consists of tokenization, POS tagging and stop-word removal, and this is followed by word similarity and phrase similarity calculations which we now describe.

### 5.1 Conjunctive Concept Similarity

In traditional text-based retrieval, a document or query is represented as a vector of term weights which are used in similarity comparison. The term “*vector*” can be regarded as a new and distinct compound concept and the concept reflected by a document is best described by ANDing the concepts represented by its index terms [30], which facilitates documents being treated as conjunctive concepts.

When concepts have several disjunctive meanings in WordNet synsets, we apply ‘disjunctive minimum’ [30] to obtain the similarity between the two concepts. That is, when a concept has alternative synsets because it is polysemous, we calculate the minimum conceptual distance between the synsets and the other concept as the final distance between the two concepts. Assume that we have two concepts  $c_1$  and  $c_2$  and  $c_1$  has three disjunctive synsets  $syn_1, syn_2, syn_3$ . In terms of ‘disjunctive minimum’, the conceptual distance between  $c_1$  and  $c_2$  will be given by:

$$d(c_1, c_2) = \min [d(syn_1, c_2), d(syn_2, c_2), d(syn_3, c_2)] \quad (3)$$

In calculating conjunctive concept similarity, we take into account all elementary concepts in the conjunctive concept. We regard the comparison of the similarity of two conjunctive concepts as finding the best assignment for a bipartite graph. On both sides of the bipartite graph, the nodes represent elementary concepts. As with solving the best matching problem, we apply the Hungarian algorithm to decide the maximum similarity matching between the two conjunctive concepts. An alternative to the computationally expensive Hungarian algorithm is to perform conjunctive concept similarity, defined as [38]:

$$sim(c_1, c_2) = \frac{1}{M \cdot N} \sum_{i=1}^M \sum_{j=1}^N sim(e_i, f_j) \quad (4)$$

where  $c_1$  and  $c_2$  are the compound concepts being compared and  $e_i$  and  $f_j$  are elementary concepts for  $c_1$  and  $c_2$  respectively. In this formula, the sum of pairwise elementary concept similarities is normalized by the product of the length of conjunctive concepts to reduce the bias of the number of elementary concepts [30]. Some other approaches to conjunctive concept similarity calculation can also be found in [38].

### 5.2 Density-based Concept Selection

In the concept set, each concept represents a semantic entity in the semantic space and the pairwise relationship between 2 concepts can be determined by their

semantic similarity, represented as an  $n \times n$  symmetric matrix,  $M$ . The most similar concept group can represent a subspace in the semantic space within which the concepts have high co-occurrence correlations.

Principle Component Analysis (PCA) is a useful tool in pattern recognition in high-dimensional spaces to reduce the number of dimensions without losing much of the information represented by the data. Although PCA can ensure the orthogonality of the bases, the representation of original data in terms of *feature vectors* is difficult to interpret and embed with semantics, a point made by Wei and Ngo in [42]. However, subsets of concepts which are clusters in semantic space represent specific domain semantics which should be as disjoint as possible to be selected as the bases in semantic space. Therefore, the number of clusters, that is also the number of bases selected by clustering, should be consistent with the number of *feature vectors* selected by PCA.

We apply PCA to help find the most appropriate number of clusters in density-based concept selection. The total number of clusters is decided by considering the inconsistency coefficient and PCA. In hierarchical clustering, the inconsistency coefficient was used to decide the appropriate number of clusters in the dendrogram. The inconsistency coefficient is defined to compare the height of a link in a cluster hierarchy with the average height of links below it and can be used to identify groups of concepts which are densely packed in certain areas of the cluster dendrogram.

To demonstrate how our approach works, we take ConceptNet contextual similarity as an example, as described in Section 4.2. Figures 2 and 3 are both demonstrated using the typical concept set (85 concepts) which we investigate in Section 7.3. In Figure 2 (in green), the number of clusters formed when inconsistent values are less than a specified inconsistency coefficient is shown. According to PCA, the cumulative energy content for the top  $k$  Eigenvectors is shown in Figure 2 (in blue). As described above, the number of orthogonal vectors represent disjoint semantics in the semantic space and we strive to group as many similar concepts as possible. The trade-off between PCA inconsistency coefficient is used to find a proper number of clusters for agglomerative algorithm. As shown in Figure 2, the intersection of PCA (blue) and inconsistency coefficient (green) curves is selected to decide the number of clusters. The number of clusters at the trade-off point can still keep the cumulative energy higher than 90% while the inconsistent coefficient is at a relatively low level.

A dendrogram generated by hierarchical clustering is illustrated in Figure 3. In the dendrogram, semantically related concepts are linked together within a clus-

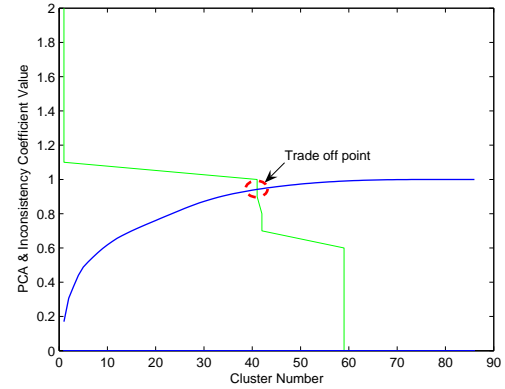


Fig. 2 Number of clusters.

ter. For example, ‘food’, ‘table’, ‘people’, ‘drink’ and ‘plate’ are grouped together, from which we see that these are more related to the activity ‘eating’ (shown as a dashed circle in Figure 3). More examples in Figure 3, such as ‘milk’, ‘water’, ‘cup’ are clustered for ‘drinking’ while ‘sky’, ‘path’, ‘tree’, ‘road sign’ and ‘road’ are clustered for ‘walking’. The semantic clustering facilitates the selection of topic-related concepts. As a concept, a given topic is also an instance which can be clustered in the concept space. Therefore, the concepts within the same cluster of a given topic are the concept candidates.

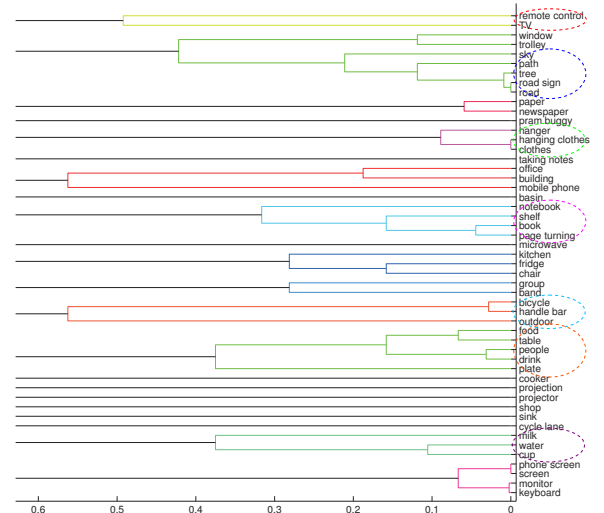


Fig. 3 Relationship dendrogram for concepts.

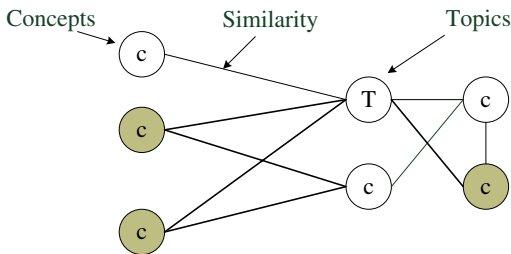


## 6 Leveraging Similarity for Concept Ranking

In the previous section we described a method for selecting candidate concepts in similarity matching based on clustering concepts in a semantic space. Although the selected concepts have high correlation with the given activity topic, there may still be other concepts missing which might be related to the topic. This is because the clustering algorithm only considers the local distance in the semantic space. Since the selected concepts have high semantic correlation with the given topic, they can be used as seeds in finding other related concepts. To leverage concept similarity in a global view, we employed the random walk model.

### 6.1 Concept Similarity Model

Random walk is a widely-used algorithm which uses links in a network to calculate global importance scores for objects which are connected in the network. It allows us to compute the probability of a random walker being located in each vertex performed as a discrete Markov chain characterized by a transition probability matrix. We model concept similarity as a graph  $G = (C, E)$ , where  $V$  is the concept set and  $E$  is a set of edges that link concepts. Each edge is assigned a given similarity value describing the probability that a random walker jumps among the concepts. As shown in Figure 4, concept sets and given topics can both be viewed as vertices in the graph, connected by similarity links. In last section, the concepts very similar to the given topic are selected as candidates, shown as the shaded concepts in Figure 4. However, the concepts which are similar to candidate concepts but have no direct similarity link with the given topic, are ignored. The random walk model ranks the concepts with candidate concepts as seeds from a global similarity view.



**Fig. 4** Concept similarity link.

### 6.2 Similarity Rank

Here we consider the process as a Markov chain where the states are concepts and transitions are similarity links between them. A random walker will start with a prior probability and surf on the graph, following similarity links. The similarity random walk is based on mutual reinforcement of concepts, that is, the score for a concept relative to a given topic influences, and is influenced by, the score of other concepts. We formulate the calculation of the score for  $c_i$  as:

$$x(c_i) = \sum_{j=1}^n Sim_{ij} x(c_j) \quad (5)$$

where  $Sim_{ij}$  is a normalized similarity value between  $c_i$  and  $c_j$ . Following the PageRank algorithm, we update the score of concepts by:

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \alpha \begin{pmatrix} Sim_{11} & \dots & Sim_{1n} \\ \vdots & \ddots & \vdots \\ Sim_{n1} & \dots & Sim_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + (1-\alpha) \begin{pmatrix} d_1 \\ \vdots \\ d_n \end{pmatrix} \quad (6)$$

where  $(d_1 \dots d_n)^T$  is a prior score vector, and  $\alpha$  is a decay factor. The equation can be formalized in a compact matrix form as:

$$\mathbf{x} = \alpha \mathbf{T} \mathbf{x} + (1 - \alpha) \mathbf{d} \quad (7)$$

In this formula,  $\mathbf{x}$  stands for the score vector and  $\mathbf{T}$  is the similarity matrix with the sum of each column normalized to 1. For each concept  $c_i$ , there is  $x_i = \sum_{j=1}^n \alpha \cdot Sim_{ij} x_j + (1 - \alpha) \cdot d_i$  for the score. To solve Equation 7, we convert it to:

$$\mathbf{x} = \alpha (\mathbf{T} + (1 - \alpha) / \alpha \cdot \mathbf{d} \cdot \mathbf{1}) \mathbf{x} \quad (8)$$

If we assume  $\mathbf{A} = \alpha (\mathbf{T} + (1 - \alpha) / \alpha \cdot \mathbf{d} \cdot \mathbf{1})$ , then  $\mathbf{x}$  will be the Eigenvector of  $\mathbf{A}$ . Although this leads to a direct solution for the formula, the iterative calculation converges fast enough and is usually employed. In our work, the iteration starts with  $\mathbf{x}$  initialized as  $\mathbf{0}$ .

## 7 Experiments and Evaluation

### 7.1 User Experiment

Our experiments start with a user investigation to determine a set of possible concepts for interpreting lifelog events. Respondents are chosen from among the researchers in our own lab, most of whom work in computer science and some also log their everyday lives



with the SenseCam so the group are sympathetic to, and familiar with, the idea of indexing visual content by semantic concepts. In total, 13 of our respondents then took part in our initial user experiment, 9 male and 4 female, whose ages are all in the range 20 – 40 years old. About half of the participants (7 in 13) are in the age group of 26 – 30 while 3 are in 21 – 25 and another 3 are over 30. There are 8 participants who are familiar with SenseCam and have worn it for various periods. Five participants carry out research using SenseCam and are engaged in different tasks like visualization, concept detection, medical therapy, etc. The demographic information for our participants is shown in Table 2.

**Table 2** Demographic information of participants

User	Gender	Age	Ever Used SenseCam ?	Working on SenseCam
1	Male	26-30	Yes	Yes
2	Female	21-25	No	No
3	Male	36-40	Yes	Yes
4	Male	26-30	Yes	Yes
5	Male	26-30	No	No
6	Male	26-30	No	No
7	Female	26-30	Yes	Yes
8	Male	26-30	Yes	No
9	Male	31-35	Yes	Yes
10	Male	21-25	Yes	No
11	Female	21-25	Yes	No
12	Female	26-30	No	No
13	Male	31-35	No	No

Participants were shown SenseCam images for samples of activities and were then surveyed on their understanding of images of SenseCam activity as well as the concepts occurring in those SenseCam images. The experiment was organized into three phases namely study, phrase and rating. In the study phase, target activities were first described to respondents to make them familiar with the activity concepts. Exemplar image streams for each activity were shown to the group and they were then asked to inspect the SenseCam images. In the pooling stage, participants were asked to go through images collected individually to list the possible concepts they thought might be helpful in order to retrieve the activities. The aim of the second phase is to determine a large concept set that might be helpful in analyzing SenseCam images in order to detect activities. In the final rating phase, the number of subjects who thought a concept was relevant to the given activity is calculated, for all target activities. The higher the number of “votes” each concept gets and the greatest agreement

among all subjects, the more importance we give to the concept for that activity.

In the pooling stage, subjects were asked to list as many concepts associated with each event topic while inspecting SenseCam images through controlled browsing. Note that the pooled event topics are all from the everyday activities we investigated in Section 3.1, as shown in Table 1. To provide cues for participants to find appropriate concepts, SenseCam images depicting different activities were shown. In our later experiment on evaluating concept selection in Section 7.4, we use the concept set obtained from this user experiment. The concepts investigated include 171 concepts in total.

## 7.2 Experimental Evaluation – Methodology

To evaluate our concept selection algorithm, the user experiment acts as the “oracle” result. In the user experiment, the ranked concepts are analyzed to determine the set of agreed ones, decided unanimously for the evaluation. Benchmarks are introduced to evaluate the algorithms from different performance points of views. These viewpoints are group consistency, set agreement and rank correlation [18].

In order to assess the generated clustering, we define *group consistency* to measure the degree of semantically related concepts to be clustered. When two related concepts are grouped in the same cluster by our algorithm, this should give a positive contribution to the overall consistency value, otherwise, a negative contribution should be given to overall consistency. To determine whether two concepts should be grouped together is a subjective decision hence the results of human experiments are used as an oracle evaluation. We formalize the notion of human judgement on concept group consistency as a binary function  $O$ :

$$O(c_i, c_j) = \begin{cases} 1 & \text{if } c_i \text{ and } c_j \text{ are under the same topic} \\ 0 & \text{if } c_i \text{ and } c_j \text{ not under the same topic} \end{cases} \quad (9)$$

Similarly, we define another binary function  $G$  to reflect the grouping result of two concepts by clustering as:

$$G(c_i, c_j) = \begin{cases} 1 & \text{if } c_i \text{ and } c_j \text{ are in the same cluster} \\ 0 & \text{if } c_i \text{ and } c_j \text{ not in the same cluster} \end{cases} \quad (10)$$

Note that these two binary functions are both symmetric which means  $O(c_i, c_j) = O(c_j, c_i)$  and  $G(c_i, c_j) = G(c_j, c_i)$ . Generating a set  $\mathcal{C}$  of ordered pairs  $\mathcal{C} = \{(c_i, c_j), 1 \leq i, j \leq |\mathcal{C}|, i \neq j\}$  from concept set  $\mathcal{C}$ , the overall group consistency for  $\mathcal{C}$  is defined based on these two functions and is formalized as:

$$GC = \frac{|\mathcal{C}| - \sum_{(c_i, c_j) \in \mathcal{C}} IC(O, G, c_i, c_j)}{|\mathcal{C}|} \quad (11)$$

where

$$IC(O, G, c_i, c_j) = \begin{cases} 1 & \text{if } O(c_i, c_j) \neq G(c_i, c_j) \\ 0 & \text{if } O(c_i, c_j) = G(c_i, c_j) \end{cases} \quad (12)$$

Group consistency reflects the performance of similarity-based clustering in the form of a pairwise grouping result. The ratio is computed as the fraction of the pairs for which the semantic clustering algorithm gives the same output as the user experiment. If there are no cases in which semantic clustering mis-groups a concept pair,  $GC$  is equal to 1. Conversely,  $GC$  is equal to 0 when no concept pairs are correctly grouped.

*Set agreement* is used to compare two concept sets without considering the ranking measurement. It defines the positive proportion of specific agreement between two sets [18]. The score of set agreement is equal to 1 when the two sets  $C_1 = C_2$ , and 0 when  $C_1 \cap C_2 = \phi$ .

Rank correlation is used to study the relationships between different rankings on the same concept set. We employ Spearman’s ranking correlation coefficient to measure the final output. According to the definition, the score is equal to 1 when agreement between the two rankings are the same, and -1 when one ranking is the reverse of the other.

### 7.3 Evaluation Setup

We recruited 13 persons in our user experiment for concept recommendation. Diverse concepts were suggested by our subjects as shown in Figure 5 showing that the number of concepts increases significantly when less agreement is achieved, from 13 votes to 2 votes. We ignore concepts with only 1 vote because one subject’s suggestion means very little in terms of a common understanding of concept selection.

We initially concentrate on a smaller concept set in which concepts are selected with *agreement*  $\geq 50\%$ . When too few concepts are selected for a topic, more concepts with a smaller agreement will also be selected in order to make each topic have at least 5 concepts. In this concept set, there are a total of 85 concepts.

To test the robustness of different similarity measures used in our density-based concept selection, we also carried out experiments on a larger concept set with less agreement among users (*vote*  $\geq 2$ ), forming a broader set of 171 concepts. The distribution of all 171 concepts across activities is depicted in Figure 6 with most activities having between 10 and 20 concepts, the

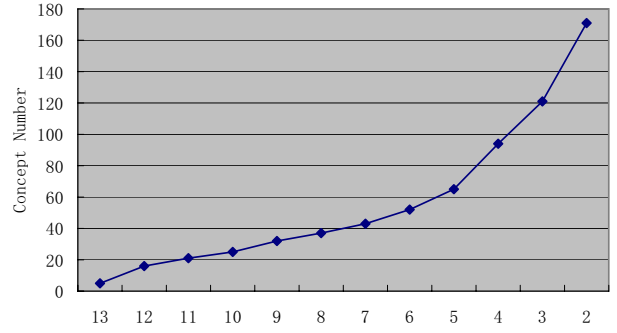


Fig. 5 Concept number vs. agreement.

overall average being 15. Among all activities, ‘Cooking’ has more relevant concepts selected as more visual concepts are involved and are helpful to identify the activity, such as various kitchen items and food which are very specific. Activities like ‘Using phone’, ‘Reading’, ‘Pet care’ and ‘Going to cinema’ tend to have relatively similar images within one single event sample, therefore have less concepts recommended.

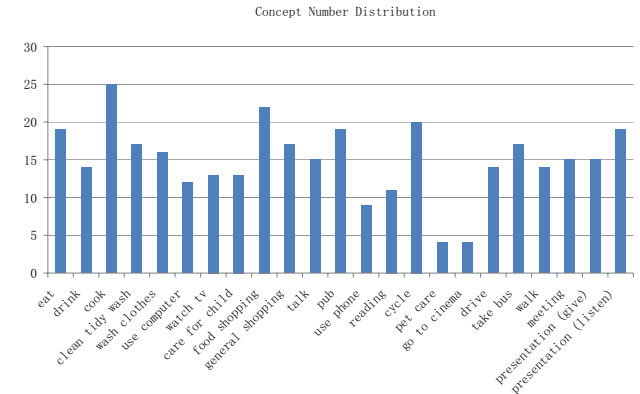


Fig. 6 Distribution of concepts.

To measure semantic similarity, we employed both taxonomic similarity and contextual similarity as discussed in Section 4 using the ontologies of WordNet and ConceptNet, respectively. For taxonomic similarity, we also compared 5 mainstream similarity measures, those of *Wu* and *Palmer*, *Leacock* and *Chodorow*, *Resnik*, *Jiang* and *Conrath*, *Lin* all of which were introduced and described earlier. Contextual similarity is obtained by spreading activation through ConceptNet links. After normalising by textual processing, the word-word semantic similarity is first calculated and then combined to get phrase-level similarity for conjunctive concepts composed of multi-words.

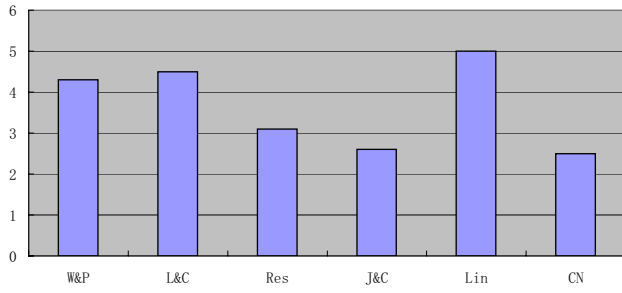


Fig. 7 Concept number per topic.

Concept-concept similarity and topic-concept similarity are both used in our density-based concept selection algorithm to cluster the most similar concepts in the same clusters with corresponding event topics. The output concepts from hierarchical clustering are first analyzed to show the diversity of result concepts by different semantic similarity measures. The average number of concepts selected per event topic is depicted in Figure 7. Though there is not much difference in the average number of concepts per topic, *Lin* selected more concepts (5.0) compared to *Jiang* and *Conrath* and *ConceptNet* which both select 2.6 and 2.5 concepts per topic respectively.

#### 7.4 Result of Evaluation

Our results are assessed to compare the performance of applications of the two ontologies, WordNet and ConceptNet, for semantic density-based concept selection in the lifelogging domain. Our density-based concept selection and re-ranking algorithm involves several steps including similarity calculation, agglomerative clustering, similarity ranking and so on, therefore we evaluate our results in manifold ways.

##### 7.4.1 Evaluating the Clustering Algorithm

We first apply clustering to group semantically related concepts based on similarity measurement. Group consistency is calculated for each ontology to assess the clustering performance of our agglomerative algorithm in capturing the semantic relationships in everyday life events. The comparison of all above referred ontological measures are shown in Figure 8.

The assessment is first carried out on a small concept set (85 concepts) as shown by blue bars in Figure 8 and as we can see, ConceptNet-based similarity shows more consistency compared to the other similarity measures. Using the same concept set and agglomerative clustering algorithm, this indicates that the similarity values

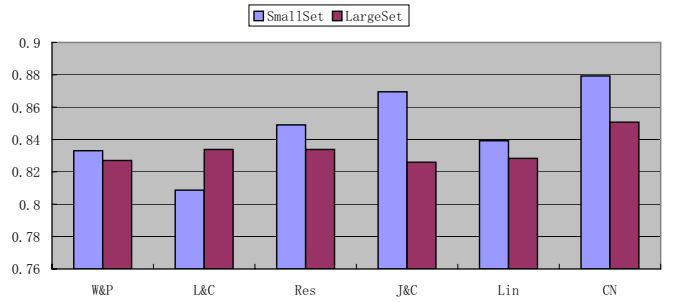


Fig. 8 Group Consistency comparison.

returned by our spreading activation from ConceptNet do reflect the semantics of everyday activities. We increased the testing concept set by using the larger concept set (171 concepts) shown by red bars and found that *ConceptNet* still outperforms the other similarity sources.

In the larger concept set, the semantic similarity calculation is also performed first for these 171 concepts and topics followed by an hierarchical clustering algorithm. Output concepts are compared on a topic basis against the groundtruth from the user experiment. Comparison is done on *Set Agreement* and *Rank Correlation* to evaluate the performance of different similarity measures. Because topics are not uniform in assessing performance, we do not average results over all topics. The results of our density-based everyday concept selection are presented in Figures 9 and 10. The performance of similarity measures is compared in Figure 9 using *Set Agreement* where we find ConceptNet-based concept selection has the highest median value and better quartile scores than WordNet-based measures. Among WordNet-based similarities, *Leacock* performs best on *Set Agreement* but does not show advantages on *Rank Correlation* as shown in Figure 10.

ConceptNet-based concept selection results in the highest median and quartile scores on *Rank Correlation* while *Jiang* has the best performance among WordNet-based similarities, but is outperformed by *ConceptNet*. We conclude that ConceptNet-based similarity performs best not only on the concepts selected (as implied by *Set Agreement*), but also on the ranking of these concepts (as implied by *Rank Correlation*). The contextual ontology is thus more suitable in everyday concept selection for the lifelogging domain.

##### 7.4.2 Similarity Ranking Assessment

Similar to group consistency, we define pairwise orderedness to evaluate ranking performance of our algorithm, as the following formula:

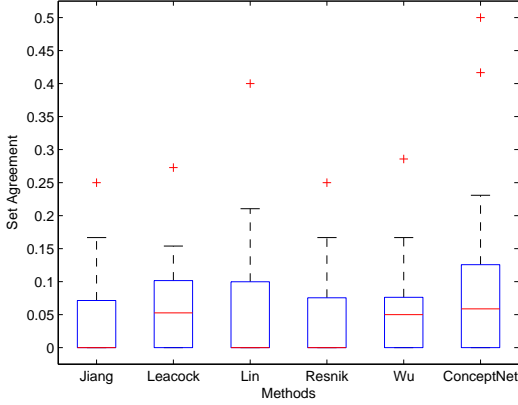


Fig. 9 Comparison based on Set Agreement.

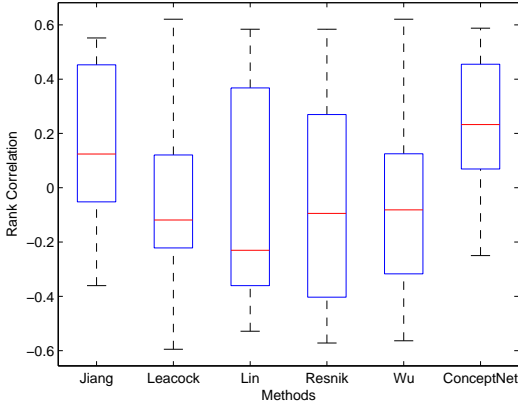


Fig. 10 Comparison of Rank Correlation.

$$PO = \frac{|\mathcal{C}| - \sum_{(c_i, c_j) \in \mathcal{C}} IC(O, R, c_i, c_j)}{|\mathcal{C}|} \quad (13)$$

where

$$IC(O, R, c_i, c_j) = \begin{cases} 1 & \text{if } R(c_i) \geq R(c_j) \& O(c_i) < O(c_j) \\ 1 & \text{if } R(c_i) \leq R(c_j) \& O(c_i) > O(c_j) \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

$O(c) = 1$  if concept  $c$  is selected as a groundtruth concept in the user experiment, otherwise,  $O(c) = 0$ .  $R(c)$  is the final score for concept  $c$  returned by the similarity ranking.

A comparison of ontology similarities using pairwise orderedness is shown in Figure 11 on the small concept set (85 concepts). *ConceptNet* similarity outperforms the other measures in most cases for which the curve of *ConceptNet* (CN) is above all the other curves (activities before ‘cook’). There are only four cases in which

*ConceptNet* shows worse performance than WordNet-based similarity measures, namely ‘cook’, ‘listen to presentation’, ‘general shopping’ and ‘presentation’. We also analyzed the poor performance of *ConceptNet* on these activity types. For ‘listen to presentation’ and ‘presentation’, *ConceptNet* did not perform well due to the lack of context information for the concept ‘presentation’. By looking up the ontology structure of *ConceptNet*, we find only two concepts that are contextually connected to ‘presentation’ with high correlation, ‘fail to get information across’ and ‘at conference’ and connected with ‘presentation’ by relationships ‘CapableOf’ and ‘LocationOf’ respectively. Thus it is hard to quantify related concepts in our concept set with a high similarity weight. In our experiment, ‘general shopping’ is introduced as a very general concept for which even humans find hard to decide the most related concepts.

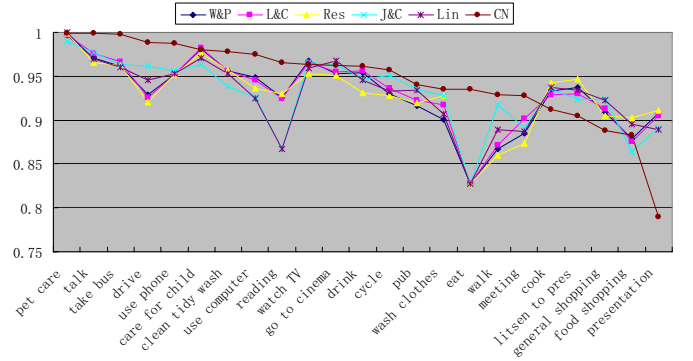
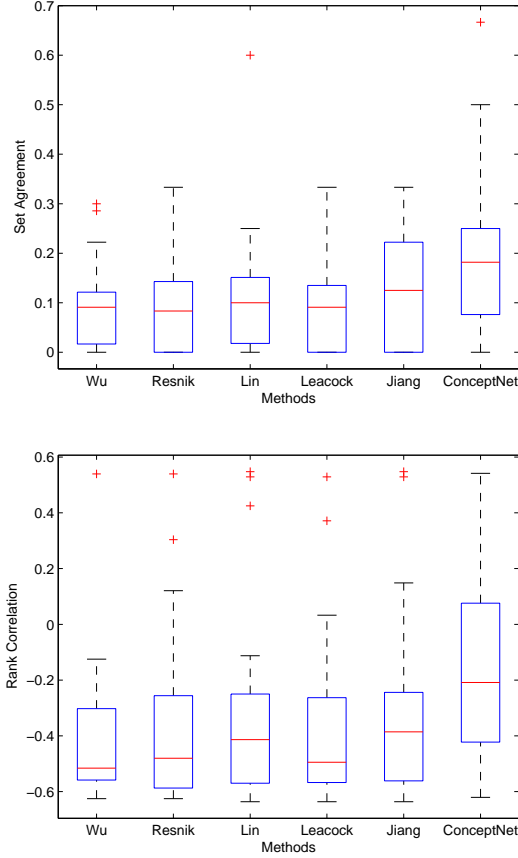


Fig. 11 Comparison of Pairwise Orderedness (small set).

An evaluation on pairwise orderedness is also carried out on the larger 171 concept. The comparison shows a similar result to Figure 11. *ConceptNet*-based semantic similarity still performs better than other similarity measures in most cases. In only three cases, *ConceptNet* does not perform as well as WordNet-based similarities, those three cases being ‘cook’, ‘presentation’ and ‘general shopping’. The reason for poor performance can be explained in the same way as when we were using the small concept set. Note that in the ‘cook’ topic, more procedures such as ‘washing’, ‘peeling potatoes’, ‘stir frying’, to name a few, are involved. The contextual diversity also makes it difficult for *ConceptNet* to return the contextual similarity correctly.

Ranked concepts based on semantic similarity are also compared using metrics of *Set Agreement* and *Rank Correlation*. To simplify the comparison, we perform an evaluation on the smaller concept set with the selection on the Top-5 and Top-10 concepts returned by the similarity rank algorithm. The performance of different se-

mantic similarity measurements are shown respectively in Figure 12 and Figure 13.

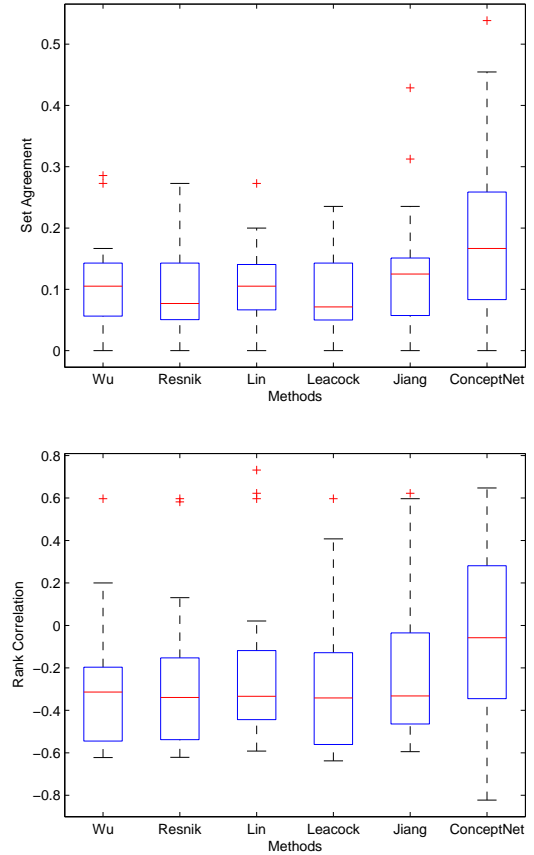


**Fig. 12** Comparison for Top-5 ranked concepts (smaller concept set).

As we can see from Figure 12, the advantage of using *ConceptNet* is more obvious as we select more concepts after similarity rank compared to very few concept seeds selected by clustering. In Figure 12, the *ConceptNet*-based algorithm outperforms the others not only in *Set Agreement* but also in *Rank Correlation*. The advantages of *ConceptNet* when Top-10 concepts are selected as depicted in Figure 13 show the robustness of our similarity rank algorithm which propagates the similarity network and gives higher weights to more relevant concepts based on the seeds selected by the clustering algorithm.

## 8 Conclusions

This paper investigates digital recording of everyday activities, known as visual lifelogging, and elaborates the selection of target activities for semantic analysis. A



**Fig. 13** Comparison for Top-10 ranked concepts (smaller concept set).

density-based approach to selecting semantic concepts is introduced to exploit and leverage concept similarity as reasoned from underlying ontologies. Suggested concepts are then re-ranked with candidate concepts selected by agglomerative clustering, used as seeds. In this paper, semantic reasoning on prevalent lexical and contextual ontologies are also discussed.

The efficacy of our concept selection algorithm is shown in the way we select and rank concepts used to represent everyday lifelogging activities, from a global view. We conclude that candidate concepts selected by clustering depend on grouping consistency. Usually, the similarity measures which correctly reflect semantic relationships between concepts can obtain better group consistency, as demonstrated by the good performance of *ConceptNet* similarity in lifelogging. The best performance of contextual similarity obtained from spreading activation on *ConceptNet* shows that contextual similarity is more suitable in reflecting semantics of everyday concepts in the lifelogging domain. *ConceptNet* similarity better reflects the relationship of everyday

activities and concepts because they are more contextually relevant in the lifelogging domain.

## References

1. <http://lastlaugh.inf.cs.cmu.edu/lscm/>. Accessed 26 November 2011
2. <http://www.statistics.gov.uk/statbase/ssdataset.asp?vlnk=7038&more=y>. Accessed 6 September 2010
3. <http://www.statistics.gov.uk/statbase/ssdataset.asp?vlnk=9497&pos=&colrank=1&rank=272>. Accessed 6 September 2010
4. Banerjee, S., Pedersen, T.: An adapted Lesk algorithm for word sense disambiguation using WordNet. In: *CICLing '02: Proceedings of the Third International Conference on Computational Linguistics and Intelligent Text Processing*, pp. 136–145. Springer-Verlag, London, UK (2002)
5. Blum, M., Pentland, A.S., Tröster, G.: InSense: Interest-based life logging. *Multimedia, IEEE* **13**(4), 40–48 (2006)
6. Bukhin, M., DelGaudio, M.: WayMarkr: Acquiring perspective through continuous documentation. In: *MUM '06: Proceedings of the 5th international conference on mobile and ubiquitous multimedia*, p. 9. ACM, New York, NY, USA (2006)
7. Bush, V.: As we may think. *The Atlantic Monthly* (1945)
8. Byrne, D., Doherty, A.R., Snoek, C.G.M., Jones, G.J.F., Smeaton, A.F.: Everyday concept detection in visual lifelogs: validation, relationships and trends. *Multimedia Tools Appl.* **49**(1), 119–144 (2010)
9. Chang, S.F., Hsu, W., Jiang, W., Kennedy, L., Xu, D., Yanagawa, A., Zavesky, E.: Evaluating the impact of 374 visual-based LSCOM concept detectors on automatic search. In: *Proceedings of the 4th TRECVID Workshop*. Gaithersburg, USA (2006)
10. Chilvers, R., Corr, S., Hayley, S.: Investigation into the occupational lives of healthy older people through their use of time. *Australian Occupational Therapy Journal* **57**(1), 24–33 (2010)
11. Christel, M.G., Hauptmann, A.G.: The use and utility of high-level semantic features in video retrieval. In: *CIVR'05: Proceedings of the International Conference on Video Retrieval*, pp. 134–144. Dublin, Ireland (2005)
12. Doherty, A.R., Caprani, N., O Conaire, C., Kalnikaite, V., Gurrin, C., O'Connor, N.E., Smeaton, A.F.: Passively recognising human activities through lifelogging. *Computers in Human Behavior* **27**(5), 1948–1958 (2011)
13. Haubold, A., Natsev, A., Naphade, M.: Semantic multimedia retrieval using lexical query expansion and model-based reranking. *IEEE International Conference on Multimedia and Expo* pp. 1761–1764 (2006)
14. Hauff, C., Aly, R., Hiemstra, D.: The effectiveness of concept based search for video retrieval. In: *Workshop Information Retrieval (FGIR 2007)*, pp. 205–212. Halle-Wittenberg, Germany (2007)
15. Hirst, G., St-Onge, D.: Lexical chains as representation of context for the detection and correction malapropisms. *WordNet: An Electronic Lexical Database (Language, Speech, and Communication)* (1998)
16. Hodges, S., Williams, L., Berry, E., Izadi, S., Srinivasan, J., Butler, A., Smyth, G., Kapur, N., Wood, K.: SenseCam: A retrospective memory aid. In: *Proc. 8th International Conference on Ubicomp*, pp. 177–193. Orange County, CA, USA (2006)
17. Hori, T., Aizawa, K.: Context-based video retrieval system for the life-log applications. In: *Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval, MIR '03*, pp. 31–38. ACM, New York, NY, USA (2003)
18. Huurnink, B., Hofmann, K., de Rijke, M.: Assessing concept selection for video retrieval. In: *MIR '08: Proceeding of the 1st ACM international conference on Multimedia information retrieval*, pp. 459–466. ACM, New York, NY, USA (2008)
19. Kahneman, D., Krueger, A.B., Schkade, D.A., Schwarz, N., Stone, A.A.: A survey method for characterizing daily life experience: The day reconstruction method. *Science* **306**(5702), 1776–1780 (2004)
20. Lakoff, G.: *Women, Fire, and Dangerous Things*. University of Chicago Press (1990)
21. Leacock, C., Chodorow, M.: Combining local context and WordNet similarity for word sense identification. *An Electronic Lexical Database* pp. 265–283 (1998)
22. Li, X., Wang, D., Li, J., Zhang, B.: Video search in concept subspace: A text-like paradigm. In: *Proceedings of the 6th ACM international conference on Image and video retrieval, CIVR '07*, pp. 603–610. ACM, New York, NY, USA (2007)
23. Liu, H., Singh, P.: Commonsense reasoning in and over natural language. In: *Proceedings of the 8th International Conference on Knowledge-Based Intelligent Information and Engineering Systems*. Springer, Wellington, New Zealand (2004)
24. Mann, S.: Wearable computing: A first step toward personal imaging. *Computer* **30**(2), 25–32 (1997)
25. Mann, S., Fung, J., Aimone, C., Sehgal, A., Chen, D.: Designing EyeTap digital eyeglasses for continuous lifelong capture and sharing of personal experiences. In: *Proc. CHI 2005 Conference on Computer Human Interaction*. ACM Press, Portland, Oregon, USA (2005)
26. Miller, G.A.: WordNet: a lexical database for english. *Communications of the ACM* **38**(11), 39–41 (1995)
27. Naphade, M., Smith, J.R., Tesic, J., Chang, S.F., Hsu, W., Kennedy, L., Hauptmann, A., Curtis, J.: Large-scale concept ontology for multimedia. *IEEE Multimedia* **13**(3), 86–91 (2006)
28. Natsev, A.P., Haubold, A., Tesic, J., Xie, L., Yan, R.: Semantic concept-based query expansion and re-ranking for multimedia retrieval. In: *MULTIMEDIA '07: Proceedings of the 15th international conference on Multimedia*, pp. 991–1000. ACM, New York, NY, USA (2007)
29. Over, P., Ianeva, T., Kraaij, W., Smeaton, A.F.: TRECVID 2005 – an overview. In: *Proceedings of TRECVID 2005* (2005)
30. Rada, R., Mili, H., Bicknell, E., Blettner, M.: Development and application of a metric on semantic nets. *IEEE Transactions on Systems, Man and Cybernetics* **19**(1), 17–30 (1989)
31. Reddy, S., Parker, A., Hyman, J., Burke, J., Estrin, D., Hansen, M.: Image browsing, processing, and clustering for participatory sensing: lessons from a dietsense prototype. In: *EmNets'07: Proceedings of the 4th workshop on Embedded networked sensors*, pp. 13–17. ACM Press, Cork, Ireland (2007)
32. Resnik, P.: Using information content to evaluate semantic similarity in a taxonomy. In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pp. 448–453. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1995)
33. Resnik, P.: Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. *Journal of Artificial Intelligence Research* **11**, 95–130 (1999)

34. Richardson, R., Smeaton, A.F.: Using WordNet in a knowledge-based approach to information retrieval. Tech. Rep. CA-0395, Dublin City University (1995)
35. Sellen, A., Fogg, A., Aitken, M., Hodges, S., Rother, C., Wood, K.: Do life-logging technologies support memory for the past? An experimental study using SenseCam. In: Proc. CHI 2007, ACM Press, pp. 81–90. New York, NY, USA (2007)
36. Sellen, A.J., Whittaker, S.: Beyond total capture: a constructive critique of lifelogging. *Communications of the ACM* **53**(5), 70–77 (2010)
37. Smeaton, A.F., Over, P., Kraaij, W.: High-level feature detection from video in TRECVID: a 5-year retrospective of achievements. In: A. Divakaran (ed.) *Multimedia Content Analysis, Theory and Applications*, pp. 151–174. Springer Verlag, Berlin (2009)
38. Smeaton, A.F., Quigley, I.: Experiments on using semantic distances between words in image caption retrieval. In: *Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR '96*, pp. 174–180. ACM, New York, NY, USA (1996)
39. Snoek, C.G.M., van Gemert, J.C., Gevers, T., Huurnink, B., Koelma, D.C., van Liempt, M., de Rooij, O., van de Sande, K.E.A., Seinstra, F.J., Smeulders, A.W.M., Thean, A.H.C., Veenman, C.J., Worring, M.: The MediaMill TRECVID 2006 semantic video search engine. In: *Proceedings of the 4th TRECVID Workshop*. Gaithersburg, USA (2006)
40. Snoek, C.G.M., Huurnink, B., Hollink, L., Rijke, M.D., Schreiber, G., Worring, M.: Adding semantics to detectors for video retrieval. *IEEE Transactions on Multimedia* **9**(5), 975–986 (2007)
41. Snoek, C.G.M., Worring, M., van Gemert, J.C., Geusebroek, J.M., Smeulders, A.W.M.: The challenge problem for automated detection of 101 semantic concepts in multimedia. In: *Proceedings of the 14th annual ACM international conference on Multimedia, MULTIMEDIA '06*, pp. 421–430. ACM, New York, NY, USA (2006)
42. Wei, X.Y., Ngo, C.W.: Ontology-enriched semantic space for video search. In: *MULTIMEDIA '07: Proceedings of the 15th international conference on Multimedia*, pp. 981–990. ACM, New York, NY, USA (2007)
43. Wu, Z., Palmer, M.: Verb semantics and lexical selection. In: *Proceedings of the 32nd annual meeting on Association for Computational Linguistics*, pp. 133–138. Stroudsburg, PA, USA (1994)