

おまかせ電子日記:行動履歴に基づいた電子日記自動生成システム

伊神 徹[†], 林 貴宏^{††}, 尾内 理紀夫^{††}

[†] 電気通信大学大学院電気通信学研究科情報工学専攻 東京都調布市調布ヶ丘 1-5-1

^{††} 電気通信大学電気通信学部情報工学科 東京都調布市調布ヶ丘 1-5-1

Automatic Electronic Diary:Automatic Generation of Diary Based on Action History

ToruIGAMI[†] TkahiroHAYASHI^{††} RikioONAI^{††}

[†] Department of Computer Science,Graduate School of Electro-Communications,The University of Electro-Communications 1-5-1 Chofugaoka,Chofu-shi,Tokyo, Japan

^{††} Department of Computer Science,The University of Electro-Communications 1-5-1 Chofugaoka,Chofu-shi,Tokyo, Japan

利用者の一日の行動履歴から電子日記を自動生成するシステム『おまかせ電子日記』を提案する。提案システムの特徴は利用者が日記生成のための特別な入力を行なうことなく、利用者の日常生活の中で生成される行動履歴を基に電子日記を生成する点である。提案システムでは電子メール、音楽の再生履歴、検索キーワード、デジタルカメラで撮影した画像、GPSによる位置情報といった行動履歴のデータから一日を特徴付ける情報の抽出を行ない、それらの情報をテンプレート文に当てはめることで電子日記の自動生成を実現する。また、生成される日記は過去に生成された日記に向けたリンクを持つ。リンクは興味対象の類似性、位置情報の類似性を基に生成される。

1 はじめに

近年のストレージ容量の増大やその価格の低下により、ライフログと呼ばれる映像や位置情報などの個人の行動履歴を一生分保存する試みが行なわれており、これを対象とした研究、開発が展開されている^{1) 2) 3) 4) 5)}。

日々保存されるライフログは大量のデータとなり、そのままの形では後に人が見返すことは困難となる。その為、それらのログについて要約などの再構成やタグ付けを行なうことが重要となる。そこで、本研究では利用者の行動履歴から電子日記の自動生成を行うのシステム『おまかせ電子日記』を提案する。『おまかせ電子日記』は日記を自動生成する機能と、生成した日記から過去に生成した日記に向けてのリンクを

自動生成する機能を主な機能として備えている。

以下、2章で『おまかせ電子日記』の機能である行動履歴からの日記の自動生成について述べ、3章でリンクの自動生成について述べる。4章で提案システムを評価する。そして、5章で関連研究を示す。

2 電子日記の自動生成

2.1 機能概要

日記の自動生成機能は、利用者の一日の出来事を自動的に日記として生成する機能である。本機能の特徴は利用者が日記生成のための特別な入力を行なうことなく、利用者の日常生活の中で生成される行動履歴を基に電子日記を生成する点である。日記生成の際に使用する行動履歴の種類は、

2007年10月21日



16時56分頃
神奈川県相模原市橋本六丁目9にて

天気は晴れ。
今日は清野達也に一番メールを送った。メール内容は、「明日、マウスを買おうと思うんだけど参考までに。コードレスのマウスって使い勝手違う?こないだ買ったというマウスどう?」などといった感じ。そういえば今日はコンピュータ周りのことが気になった。例えば、Microsoftとか。あと、プログラミングやwindowsも。そういえば、osなんかも。
そんな今日はKraftwerkのPocket Calculatorを聴きながら、14時15分に自宅出発。14:37:24頃に、神奈川県相模原市橋本二丁目1の辺りに行った。
KraftwerkのThe Voice Of Energyを聴きながら、19時59分に自宅到着。
といった一日でした

行った場所が似ている日記 興味持ったものが似ている日記

- | | |
|--------------|--------------|
| •2007年10月17日 | •2007年10月20日 |
| •2007年09月20日 | •2007年10月26日 |
| •2007年10月18日 | •2007年10月03日 |

Fig. 1 自動生成する日記例

- 送信した電子メール
- PC上やポータブルプレイヤーで再生した音楽
- Web上での検索キーワード
- デジタルカメラで撮影した画像
- GPSによる位置情報

である。これらのログから一日を特徴付ける情報を抽出する。本研究では一日を特徴付ける情報をキー情報と呼ぶ。それぞれの履歴から抽出したキー情報を基に電子日記の生成を行なう。自動生成される日記の例をFig.1に示す。

2.2 テンプレートを用いた日記生成

本研究において日記の文章の生成にはテンプレートを用いた文章生成手法⁶⁾を用いる。テンプレートを用いた文章生成では、あらかじめテンプレートのスロット(空欄)

- 1) 今日は friend に一番メールを送った。
- 2) メール内容は「mail」などといった感じ。
- 3) そういえば、category に関することが気になった。
- 4) 例えば、query とか。
- 5) あと、query や query も。
- 6) そういえば、query なんかも。
- 7) そんな今日は、summary といった一日でした。
- 8) time に place action。
- 9) time 頃に、place の辺りに行った。
- 10) music を聴きながら、

Fig. 2 テンプレートの一例:下線が引かれている部分はスロットを表す

で欲しい情報の属性を指定しておく。そして、文章生成の基となるデータからスロットに当てはまる情報を抽出し、それらのスロットを埋めていくことで文章を生成する。本システムで利用されるテンプレートの一例をFig.2に示す。図に示すように本研究では、スロットとして friend, mail, category, query, allaction, time, place, action, music を考える。以下、各スロットに当てはまる情報の属性について説明する。

- friend スロット

その日に最も多くメールを送信した相手の名前を当てはめる。この名前は、利用者の携帯電話やメールソフトのアドレス帳に登録されている、メールを送信した相手を示すメールアドレスに対応する送信相手の名前である。もし、送信相手の名前が登録されていない際には、送信相手のメールアドレスを friend スロットに当てはめる。

- mail スロット

一日の間に最も多くメールを送信し

メール内容は、
「研究室でうにようによしてるから。
じゃ、スーツで行きます。
雨でスーツ着る気持ちが萎えてた。」
などといった感じ。

Fig. 3 mail スロットへ情報を当てはめた例

た相手へのメールの本文からメールの内容を特徴付ける文を抽出し当てはめる。mail スロットに情報を当てはめた例を Fig.3 に示す。

- category スロット

その日に利用者が最も興味を持った分野の分野名を当てはめる。本研究では、この利用者が最も興味を持った分野を興味分野と呼ぶ。興味分野はコンピュータ、アート、ビジネス、ゲーム、健康の話、生活一般、メディア、ショッピング、趣味、科学、社会的な話、スポーツの12分野から選ばれる。これらの分野はOpen Directory Project(ODP)¹のカテゴリ分類を参考に決定した。ODPとは人手によりカテゴリ分類が行なわれているディレクトリ型検索システムである。

- query スロット

利用者がその日に使った検索キーワードの中から興味分野を代表する検索キーワードを抽出し当てはめる。

- time, place, action スロット

これらのスロットに当てはめる情報は一組のデータとして扱う。time スロットには「10時30分」、「23:30:42」などの時間情報を当てはめる。そして、place スロットには「自宅」、「友人宅」、「仕事場」などといった場所名や「東

9時32分に自宅を出発。
10時20分頃に東京都八王子市南大沢2丁目3の辺りへ行った。
13時12分に大学に到着。

Fig. 4 time, place, action スロットに情報を当てはめた例

The Beatles の Help! を聴きながら、
13時12分に大学に到着。

Fig. 5 music スロットに情報を当てはめた例

京都調布市調布ヶ丘1丁目5番地1」などの位置の住所を当てはめる。また、action スロットには「～に到着」、「～を出発」といった利用者の行動情報が当てはめられる。time スロットに当てはめる時間情報はaction スロットに当てはめる行動情報が起きた時刻である。time, place, action スロットに情報を当てはめた例を Fig.4 に示す。

- summary スロット

8),9)のテンプレート文をもとに生成した文を当てはめる。

- music スロット

action スロットに当てはめる行動が行なわれた時刻に再生された楽曲のアーティスト名+曲名や曲名などの楽曲情報を当てはめる。これは利用者がある行動をしていた時に聴いていた音楽は重要な情報になると考えるためである。music スロットに情報を当てはめた例を Fig.5 に示す。10)のテンプレート文は、8),9)のテンプレート文を修飾する文として用いる。

2.3 電子メールからのキー情報の抽出

送信メールの本文からのキー情報(mail)として本文内の重要文を抽出する。この処

¹ <http://www.dmoz.org/World/Japanese/>

理は a) 本文の分割, b) 重要度の算出, c) 重要文の選択, の三手順により実現する。

a) 本文の分割

送信メールの本文を文を単位として分割する。ここで文とは句点, 改行, 疑問符で区切られる文字列である。

b) 重要度の算出

a) において分割された各文に対し重要度を算出する。文の重要度は文に含まれる各単語の重要度の和を単語数で割った値と定義する。単語の重要度は tf-idf 値⁷⁾により定義する。

c) 重要文の選択

b) で算出された文の重要度の値が高いほうから最大 3 文を重要文としてキー情報として抽出する。

2.4 位置情報からのキー情報の抽出

本研究において, “位置”とはある緯度, 経度で示される一地点を表し, “場所”とはある位置を中心とした一定範囲の地域を表すものとする。また, “位置情報”とは緯度, 経度, 時刻の組を表し, “位置情報”の緯度, 経度に対応する場所の名前や住所を“位置情報”に付加した情報を“場所情報”と呼ぶ。

提案システムでは GPS により取得される全ての位置情報に対応する場所情報を日記に反映するのではなく, 一日を特徴付ける位置情報に対応する場所情報のみを利用する。そこで, まず GPS による位置情報の中から一日を特徴付ける位置の抽出を行なう。提案システムにおいて, 一日を特徴付ける場所とは,

1. 事前に利用者が登録した場所
2. 一日の間で長く滞在した場所

である。

2.4.1 事前に利用者が登録した場所

利用者は「自宅」, 「大学」や「会社」などの普段の生活において重要な場所と考える場所について, その場所の中心となる位置の緯度と経度, 場所の名前を事前に登録しておく。本研究ではこの事前に登録された場所をスポットと呼ぶ。このスポットへの到着時の場所情報と, スポットからの出発時の場所情報をキー情報として取り出す。これは, スポットに到着した時刻, スポットから出発した時刻は一日の特徴を認識する上で重要な情報になると考えるのである。

一日の移動情報の履歴の中で, スポットの緯度 ϕ_i , 経度 λ_i の示す位置 $\mathbf{x}_i = (\phi_i, \lambda_i)$ に対し距離 $d(\mathbf{x}_i, \mathbf{x}_j)$ が閾値以下となる位置 $\mathbf{x}_j = (\phi_j, \lambda_j)$ を持つ位置情報をスポットに滞在していたときの位置情報とする。

このスポットからの出発時の位置情報およびスポットへの到着時の位置情報に場所名を付加した場所情報をキー情報として抽出する。

2.4.2 一日の間で長く滞在した場所

通勤, 通学などを中心とした生活では, ある程度人が訪れる場所が限定されているため 2.4.1 節で述べたように事前に普段訪れるような場所を登録しておく事が出来る。しかし, 旅行地や休日の外出先のように普段は訪れないような場所をスポットとして事前に登録することは困難である。そこで, 一日の中で長く滞在した場所を GPS による位置情報の履歴から抽出し, その場所への到着時刻および, その場所からの出発時刻に対応する場所情報をキー情報として抽出する。

このような一日の中で長く滞在した場所の抽出は, a) 位置情報のクラスタリング, b) クラスタの選択, c) 特徴的な位置の決定, d) 位置の名前の取得, e) キー情報の抽出, といった五手順により実現する。以下, 各手順を説明する。

a) 位置情報のクラスタリング

まず, 同じ場所に居た時の位置情報を

まとめて抽出する事を目的して、GPSによる位置情報のクラスタリングを行なう。本研究では事前にクラスタ数を指定する必要の無い階層的クラスタリングの手法の群平均法 (group average method) を用いる。また、クラスタリングの終了条件は、「最も近いクラスタ対の距離がある閾値以上となること」とする。クラスタリング手法として、k-means法に代表されるような事前にクラスタ数を指定するクラスタリング手法の利用も検討した。しかし、抽出したい特徴的な場所というのはあらかじめ何箇所存在するかを知ることは出来ないため、この手法は本研究の目的に適さない。

b) クラスタの選択

a) において生成されたクラスタ集合 $\{D_i | i = 1, 2, \dots, n\}$ の中から、条件

$$\frac{|D_i|}{\sum_{i=0}^n |D_i|} > T \quad (1)$$

を満足するクラスタ D_i を選択する。ただし、 n はクラスタ数、式 (1) の $|D_i|$ はクラスタ D_i 内の要素数、 T はある閾値を表す。

c) 特徴的な位置の決定

b) で選択したクラスタ D_i の要素である位置情報に含まれる緯度、経度の重心を求める。この重心を特徴的な位置とする。この処理を b) で選択した全てのクラスタに対して行なうことにより、その日の特徴的な位置群を得る。

d) 場所名の取得

c) で抽出した特徴的な位置の緯度、経度に対応する住所を得る。そこで、c) で抽出した位置情報内の緯度、経度の組に対して逆ジオコーディング (inverse geocoding) を行なう。逆ジオコーディングとは、緯度、経度に対応する住所を取得する処理である。本研

究において、逆ジオコーディングは inv-geocoder² を用いて行なう。本研究ではここで得た緯度、経度、住所の組の情報を擬似スポットと呼ぶ。

e) キー情報の抽出

2.4.1 節で述べたキー情報の抽出と同様に、擬似スポットへの到着時刻における場所情報、および擬似スポットからの出発時刻における場所情報をキー情報として抽出する。

2.5 音楽の再生履歴からのキー情報の抽出

2.4 節において述べたスポットおよび擬似スポットへの到着時と、スポットおよび擬似スポットからの出発時にポータブルプレイヤーやPC上で再生されていた楽曲の曲名およびアーティスト名といった楽曲情報をキー情報として抽出する。

各スポットおよび擬似スポットに到着した時刻および出発した時刻の前後 10 分以内に再生されていた楽曲情報を抽出する。抽出された楽曲情報の中から各到着時刻、および出発時刻に再生時刻が最も近い楽曲情報をキー情報として抽出する。

2.6 検索履歴からのキー情報の抽出

利用者が一日で使用した検索キーワードから、利用者の興味分野、および興味分野を代表する検索キーワードをキー情報として抽出する。本研究では、ナイーブベイズ法 (naive bayes method)⁸⁾ を用いることで利用者の興味分野、および興味分野を代表する検索キーワードを抽出する。まず、興味分野を求める。興味分野の候補となる分野を $\{C_i | i = 1, 2, \dots, m\}$ (m は分野数; 提案システムでは $m = 12$)、利用者が一日で使用した検索キーワードを $Q = \{q_j | j = 1, 2, \dots, n\}$ (n は検索キーワードの種類数) と置くとすると、興味分野は式 (2) で表される事後確率 $P(C_i | Q)$ を最大化する

² <http://wiki.knya.net/wiki.cgi?page=invgeocoder>

ような分野 \hat{C} であると定義する.

$$\begin{aligned}\hat{C} &= \arg \max_{C_i} P(C_i|Q) \\ &= \arg \max_{C_i} P(C_i|q_1, q_2, \dots, q_n) \quad (2)\end{aligned}$$

ここで、ベイズの定理より式 (2) は

$$\hat{C} = \arg \max_{C_i} P(q_1, q_2, \dots, q_n|C_i)P(C_i) \quad (3)$$

となる. ここで各分野のもとで単語 q_j は互いに独立に生起すると仮定することで、式 (3) は

$$\hat{C} = \arg \max_{C_i} P(C_i) \prod_{j=1}^n P(q_j|C_i) \quad (4)$$

と変形する事ができる. 式 (4) より興味分野となる \hat{C} を求める. ただし,

$$P(C_i) = \frac{\sum_{q \in C_i} tf(q, C_i)}{\sum_{i=0}^m \sum_{q \in C_i} tf(q, C_i)}$$

とする. ここで、 $tf(q, C_i)$ は C_i に対応する ODP のカテゴリ内のインデックスページでの単語 q の出現頻度である. インデックスページとはカテゴリに分類された Web サイトへのリンクとその Web サイトについての説明文が記述されたページである. また、 C_i に出現する単語総数を N_i 、全分野での単語の異なり数を V とおき、

$$P(q_j|C_j) = (tf(q_j, C_i) + 1)/(N_i + V)$$

と定義する. この $P(q_j|C_j)$ の定義ではゼロ頻度問題に対応するためにラプラス法⁸⁾により単語の出現回数の補正を行った.

次に求めた興味分野を代表する単語を利用者が一日で使用した検索キーワードから抽出する. 検索キーワードと使用された単語 q_j について式 (5) から \hat{C}' を求める. この \hat{C}' が興味分野 \hat{C} と一致する単語 q_j を興味分野を代表する単語の候補単語として選択する.

$$\hat{C}' = \arg \max_{C_i} P(q_j|C_i) \quad (5)$$



16時56分頃
神奈川県相模原市橋本六丁目9にて

Fig. 6 キャプションが付加された写真例

候補単語 q_k を式 (6) によりスコアリングする.

$$\begin{aligned}s_k &= \frac{P(q_k|\hat{C})}{a_k} \quad (6) \\ a_k &= \frac{\sum_{i=0}^m P(q_k|C_i)}{m}\end{aligned}$$

このスコア s_k の値が高いほうから最大4つの単語を興味分野を代表する単語とする.

2.7 デジタルカメラで撮影した画像

撮影した写真にキャプションとして写真を撮影した時刻と撮影した場所を付加し、日記の先頭部分に掲載する. 掲載される写真は一日に撮影した写真からランダムに選択される. 写真の撮影時刻は写真画像の Exif(Exchangeable Image File Format) から獲得する. また、獲得した撮影時間と GPS により取得した位置情報の履歴を比較し、最も撮影時間に近い時刻における位置情報の緯度、経度における住所は逆ジオコーディングにより取得. 取得した住所を写真の撮影場所とする.

3 リンクの生成

3.1 機能概要

リンクの生成機能は、ある一日の日記に対し、その日の日記と似ている過去に生成した日記へのリンクを生成する機能である. 似ている日とは、

- a) 興味対象の似ている日:検索での検索対象やメールでの話題が似ている日

- b) 位置情報の似ている日:同じ場所に行き, さらにその場所での滞在時間が同程度である日

である. この a), b) の観点から, 2 種類のリンクの生成を行なう.

そこで, 日記を生成した日と過去のそれぞれの日の間の類似度を行動履歴をもとに算出する. 類似度の算出に利用する行動履歴の種類は,

- 送信した電子メール
- Web 上での検索キーワード
- GPS による位置情報

である. それぞれの観点から算出される類似度の高い日上位 3 日とその観点の下での似ている日と仮定し, 日記を生成した日からそれらの日に向けたリンクを生成する.

3.2 興味対象の類似性

興味対象の似ている日には同一の単語を Web 上での検索キーワードや電子メール内に利用している可能性が高いと仮定する. そこで, 各日の検索キーワードと電子メール内の単語集合に含まれる単語の tf-idf 値を要素とする特徴ベクトルを作成し, ある日 d_i の特徴ベクトル \mathbf{f}_i と別のある日 d_j の特徴ベクトル \mathbf{f}_j を比較することで, 二つの日の間の類似度 $T_{i,j}$ の算出を行なう. $T_{i,j}$ は式 (7) により定義する.

$$T_{i,j} = \frac{\mathbf{f}_i \cdot \mathbf{f}_j}{|\mathbf{f}_i| |\mathbf{f}_j|} \quad (7)$$

$$\mathbf{f}_t = \{tfidf(w_1, t), \dots, tfidf(w_m, t)\}$$

$$tfidf(w_s, t) = \frac{tf(w_s, t)}{W_t} \times \ln\left(\frac{K}{df(w_s) + 1}\right)$$

ここで, $tf(w_s, t)$ はある日 d_t において使用した検索キーワードと電子メール内での単語 w_s の出現頻度 (tf 値), $df(w_s)$ は日記を生成した全ての日において検索キーワードと電子メール内で単語 w_s を使用した日の日数 (df 値), W_t はある一日 d_t に用いられた単語数, K は過去に生成した日記数を表す.

3.3 位置情報の類似性

ある日と別のある日において同じ場所に行き, さらにその場所での滞在時間が同じであった場合, 高い類似度になるようにある日 d_i と別のある日 d_j における位置情報の類似度 $L_{i,j}$ を定義する. 具体的には, まず, d_i と d_j のそれぞれの GPS による位置情報から 2.4.2 節で述べた手順 a) により, d_i の GPS による位置情報についてクラスタ集合 $D_i = \{D_{i,s} | i = 1, 2, \dots, n\}$ を得る. ここで, n はある日 d_i の位置情報をクラスタリングし, 取得したクラスタ数を表す. 次に D_i を基に $L_{i,j}$ を式 (8) により算出する.

$$L_{i,j} = \sum_{s=0}^n |S_{i,s} - S_{j,\hat{t}(s)}| d(D_{i,s}, D_{j,\hat{t}(s)}) \quad (8)$$

$$S_{i,s} = \frac{|D_{i,s}|}{|\bigcup_{D_{i,k} \in D_i} C_{i,k}|}$$

ただし,

$$\hat{t}(s) = \arg \min_t d(D_{i,s}, D_{j,t})$$

であり, $|D_s|$ はクラスタ D_s に含まれる要素数である. そして, $d(D_{i,s}, D_{j,t})$ は階層的クラスタリング手法である群平均法において定義されるクラスタ間距離関数とする.

4 考察

提案システムを利用して取得した日記と手動で作成した日記とを比較して考察する. 本評価におけるシステムの利用者および手動による日記作成者は第一著者である. 利用者がある日の行動を思い出し記述した日記と提案システムにより出力された日記の一部を Fig.7, Fig.8, Fig.9, Fig.10, Fig.11, Fig.12 に示す.

4.1 電子メールからのキー情報の抽出

各日の送信メールからキー情報として抽出された文を見てみると, その日の特徴付ける文をある程度抽出できていると思われる.

.....
渡辺君に a) 「尾内先生が添削してくれ
た結果をよろしければ送って下さい」と
メールを打つ。送ってくれたが、スキヤ
ナの使い方が分らなかったようで結構苦
労する。申し訳ない。
.....

Fig. 7 その日を思い出し手動で記述した日記の一部:2007年10月7日の日記

しかし、重要文として選ばれなかった文の中に、提案システムにより重要文として抽出されるべきだと考える文がいくつかみられた。例えば、Fig.7では下線 a) で示される部分に「尾内先生が添削してくれた結果をよろしければ送って下さい」という文が記述されているが、Fig.8に示した日記にはこのような文は記述されていない。このような人名や組織名などが含まれる文はその日の話題の内容を想起しやすいという点において重要な文であると思われる。今後、この事を考慮した文の重要度の定義を検討する必要があると考える。例えば、その単語の品詞情報により単語の重要度を調節するなどといったことを現在検討している。

4.2 位置情報からのキー情報の抽出

「郵便局に寄った」という情報は Fig.9 に示した手動で作成した日記の下線 a) で示される部分に記述されているが、この場所情報は Fig.10 に示した提案システムにより出力した日記には反映されていなかった。「郵便局」という場所情報が抽出できなかった理由の一つにスポットとして登録されていなかったことがある。郵便局の様な利用者にとって訪れる頻度の低い場所は事前に登録しておく必要性が低いためである。そのため、2.4.1節で述べた手法ではキー情報としてそれらの場所情報を得ることは出来なかった。しかし、それらの位置への滞在時間が一定以上であれば2.4.2節

.....
今日は 渡辺君 (friend) に一番メールを送った。メール内容は、「そこに紙を載せてスキャンボタンを押せばスキャンできるはず。なのでその紙をスキャンしてメールで送ってほしいのです。ボタンを押すとゼミで使ってるパソコンでソフトが動くと思うから、そこでスキャンするを選べばよいと思います。」
(mail)
などといった感じ。
.....

Fig. 8 提案システムにより出力された日記の一部:2007年10月7日の日記; 下線が引かれている部分は「キー情報(スロット名)」を表す

で述べた手法によりキー情報として、それらの位置に関する場所情報を得ることが出来たと考える。実際はそれらの位置での滞在時間は十分長くなかったため、この手法ではキー情報としてそれらの場所に関する場所情報を得ることは出来なかった。

スポットとして事前登録が行なわれておらず、滞在時間も短い、普段の生活において訪れる頻度が低い場所を訪れた際にその場所の場所情報をキー情報として抽出することは重要であると考えられる。キー情報としてこのような位置情報を抽出する手法としては、過去のGPSによる位置情報から利用者の生活パターンを抽出しそれを考慮した形で、日記を生成する日のキー情報を抽出する手法を検討している。例えば、ある位置に対する利用者の存在確率⁴⁾を算出し、その存在確率の低い位置に利用者が訪れた場合、その位置に関する場所情報をキー情報として抽出するなどの手法である。

.....

昼間は3時からバイトに行く。2時間だけ働く。

そして、バイトが終わったら、家にまっすぐ帰りシャワー浴び、同窓会に行く準備をする。でも、眠くて寝る。

結果、起きたら同窓会の時間に間に合わない。

「まずいなあ」と思いつつも、財布の中にお金が無いので a) 郵便局へまず寄る。ATMはやっていた。

お金を補充して、同窓会の会場へ。7時開始だけでも30~40分位遅刻して到着。会場は.....

.....

Fig. 9 その日を思い出し手動で記述した日記の一部:2007年9月8日の日記

4.3 音楽の再生履歴からのキー情報の抽出

キー情報として抽出された楽曲情報が日記に付与されることで、この情報が無かった場合と比べて、その時々的情景をより深く思い出せたと考える。利用者がある行動をしていた時聴いた音楽の楽曲情報は、その時々を思い出す要素に成り得ると考えられる。

しかし、提案システムにより出力された日記を読むと楽曲情報が一部冗長に感じた。全ての行動に対し音楽の再生履歴からのキー情報を付与するのでは無く、このキー情報を付与するのに適切な行動を選択することが必要であると分かった。

4.4 検索履歴からのキー情報の抽出

提案システムによりキー情報として抽出された興味分野と利用者が判断した興味分野はある程度は一致していた。また、一致していない場合であっても、納得できる結果であることが多かった。しかし、提案システムでは利用者が興味分野を代表する単

.....

そんな今日は14時41分 (time) に自宅 (place) 出発 (action)。

14時44分 (time) に神奈川県相模原市相模原二丁目8(place) 到着 (action)。

18時17分 (time) に神奈川県相模原市相模原二丁目8(place) 出発 (action)。

18時27分 (time) に自宅 (place) 到着 (action)。

19時36分 (time) に自宅 (place) 出発 (action)。

POLYSICSのMaking Sense(music)を聴きながら、19時56分 (time) に神奈川県相模原市中央四丁目12(place) 到着 (action)。

.....

といった一日でした。

Fig. 10 提案システムにより出力された日記の一部:2007年9月8日の日記;下線が引かれている部分は「キー情報(スロット名)」を表す

語であると判断した検索キーワードの多くがqueryスロットへのキー情報として抽出されなかった。例えば、Fig.11に示した利用者が手動で作成した日記において利用者は興味分野はコンピュータであると考え、その興味分野を代表する単語として下線a),下線b)で示される“形態素解析”および“Sen”などが抽出されるべきであると考えた。しかし、Fig.12に示した提案システムにより出力された日記では興味分野は一致するものの、queryスロットに当てはめられる興味分野を代表する単語には利用者が検索キーワード中から抽出されるべきと考えた“形態素解析”および“Sen”などといった単語は抽出されていない。

この理由の一つにODPの各カテゴリにおけるインデックスページでの検索キー

.....
 今日はプログラミングをひたすらしていた。だけど、躓く。Javaでa)形態素解析がしたくてb)Senについて調べる。Senの設定方法を忘れてしまい、犬塚さんのサイトをみて確認。確認して快調に進むかと思いきや進まないのので、Mecab, b)Senでのa)形態素解析のやり方を調べる。調べていると、SPAMバーガーを食べた犬塚さんが

Fig. 11 その日を思い出し手動で記述した日記の一部:2007年9月20日の日記

.....
 そういえば今日はコンピュータ (category) に関することが気になった。
 例えば、java(query)とか。
 そんな今日は.....

Fig. 12 提案システムにより出力された日記の一部:2007年9月20日の日記; 下線が引かれている部分は「キー情報(スロット名)」を表す

ワード q の出現頻度の問題があると考えられる。利用者が代表単語になると判断した検索キーワード q の多くが分野 C_i に対して $tf(q, C_i) = 0$ となり、式(5)から求まる \hat{C}' が利用者の感覚とは異なる結果となるということが判明した。そのため、 $P(q|C_i)$ の定義における単語の出現数の補正手法の検討や $tf(q, C_i)$ の定義の検討などが今後必要である。

4.5 デジタルカメラで撮影した画像

一日に撮影した写真が複数枚ある場合、ランダムに一枚の写真を選択しその写真を掲載している。しかし、これでは一日を特徴付けるような写真が選択されるかは分か

らない。そのため今後、GPSによる位置情報などの行動履歴に基づき写真を選択する手法を検討する必要がある。

4.6 興味対象の類似性

興味対象の類似性に基づいて生成されたリンクのリンク元の日、リンク先の日のそれぞれにおいて使用された検索キーワードおよび電子メール内の単語集合を比較した。その結果、リンクで結び付けられた日の興味対象の類似性についてある程度確認できた。しかし、単語集合を見比べるとリンク先として選択されていない別の日の方がよりリンク元の日と類似していると感じることがあった。

これらの日では同一の単語は使用されていなかったが、それらの単語から連想される話題の分野が一致していた。このことから、興味対象の類似性に基づいたリンクを生成する場合には、同一の単語を利用しているかという点のみでなく、使われた単語の分野が一致しているかという事が重要であると考えられる。今後、この事を考慮した類似度の定義を検討し、効果を確認する必要がある。

4.7 位置情報の類似性

位置情報に基づいて生成されたリンクのリンク元の日とリンク先の日が同一の場所に訪れているとは限らず、似ているとはいえない日の一部みられた。

式(8)では、滞在時間と滞在地の一致度は同程度に利用者の各日が似ているか否かといった判断基準に影響を与えると考え定義した。しかし、実際は利用者は滞在時間の一致度よりも、滞在地が一致しているか否かということの方をより重視することが判明した。式(8)ではこのことが考慮されていないため、式(8)により算出された類似度では利用者の感覚からずれた結果となる場合がある。今後、この事を考慮した位置情報に基づいた類似度の定義を検討していく必要がある。

また、式(8)では利用者が滞在したの到着時刻および出発時刻を考慮していない。

しかし、到着時刻や出発時刻は重要な要素であると思われる。そこで上記で述べた内容に加えて、そのような時刻情報を考慮した位置情報に基づいた類似度の定義を検討し、効果を確認する必要がある。

5 関連研究

現在進められているライフログに関する研究としてMicrosoft社によるMyLifeBits Project¹⁾がある。MyLifeBits ProjectではSenseCamというデバイスにより撮影、保存される画像、GPSにより取得、保存される位置情報、通話履歴やe-mailの内容など個人に関する全ての情報をデジタル化し保存する。MyLifeBits Projectでは、これらの保存された情報を検索、閲覧するための機能がいくつか提案されているが、それらの情報から日記型のコンテンツを生成することは提案されていない。

Newmanら⁹⁾は『おまかせ電子日記』と同様に取得した利用者の行動履歴から日記を生成するシステムPEPYSを提案している。PEPYSでは日記の生成に利用者の一日の位置情報履歴を用いる。しかし、本研究と異なり複数種類のライフログを利用し日記を生成することは行なっていない。

沼ら⁵⁾は、個人の行動履歴に基づき日記の下書きを生成するシステムActionLogを提案している。ActionLogでは学術会議場といった閉じた空間中での行動履歴に基づいて日記型コンテンツを生成する。しかし、『おまかせ電子日記』では特定の空間に限らない利用者の日常生活での行動履歴に基づいた電子日記の自動生成を行う。

本庄ら¹⁰⁾はGPS機能付属携帯電話を用いた位置情報付きライフログ記録システムProfile Blogを提案している。Profile Blogでは、本研究と同様に蓄積されたライフログからWeblogコンテンツの生成を行なう。Profile Blogにおいて保存されるライフログは利用者が登録していく商品情報、メモ、買い物記録などである。『おまかせ電子日記』では、利用者が普段生活している中で生成される行動履歴に基づき電

子日記の自動生成を行ない、利用者に対し日記生成のための特別な行動、入力を求めない。

6 おわりに

本研究では利用者の一日の行動履歴から電子日記の自動生成を行うシステム『おまかせ電子日記』を提案した。電子メール、音楽の再生履歴、検索キーワード、デジタルカメラで撮影した画像、GPSによる位置情報からのキー情報の抽出手法およびリンクの生成手法について検討、考察を行った。その結果、各手法について更なる検討の余地があることが分かった。今後の予定として、4章で述べた各キー情報の抽出手法およびリンク生成手法の改良、そして多人数の被験者による評価実験を考えている。

参考文献

- 1) Gemmell Jim, Lueder Roger, and Bell Gordon, “The MyLifeBits Lifetime Store”, Proceedings of the ACM SIGMM 2003 Workshop on Experiential Telepresence, pp.80-83 (2003)
- 2) Mik Lamming, Mike Flynn, “Forget-me-not: Intimate Computing in Support of Human Memory”, INFRIEND21: International Symposium on Next Generation Human Interface, pp.125-128 (1994)
- 3) 河村竜幸, 福原知宏, 村田賢, 武田英明, 河野恭之, 木戸出正継, “対象物に『触れる』行為と記憶の遍在化による日常記憶支援”, 電子情報通信学会論文誌, Vol.J-88-D-I, No.7, pp.1143-1155 (2005)
- 4) Jun Rekimoto, Takashi Miyaki, Takaaki Ishizawa, “LifeTag: WiFi-based Continuous Location Logging for Life Pattern Analysis”, 3rd International Symposium on Location- and Context-Awareness, pp.35-49 (2007)

- 5) 沼晃介, 平田敏之, 濱崎雅弘, 大向一輝, 市瀬龍太郎, 武田英明, “ 学術会議における体験共有のための行動履歴に基づく Weblog システム ”, 情報処理学会論文誌, Vol.48, No.1, pp.85-97 (2007)
- 6) Inderjeet Mani, “ 自動要約 ”, 共立出版, pp.127-135 (2003)
- 7) 徳永健伸, “ 情報検索と言語処理 ”, 東京大学出版, pp.26-28 (1999)
- 8) 北研二, “ 確率的言語モデル ”, 東京大学出版会 pp. 46-48, 202-203 (1999)
- 9) William M. Newman, Margery A. Eldridge, and Michael G. Lamming, “ PEPYS: Generating Autobiographies by Automatic Tracking ”, Proc. 2nd European Conference on Computer-Supported Cooperative Work, pp.175-188 (1991)
- 10) 本庄勝, 森川大補, 山口明, 大橋正良, “ Profile Blog: Blog をベースとした想起的なライフログ検索の実現 ”, 情報処理学会 DICOMO2005 シンポジウム論文集, pp.461-464 (2005)