

'96 (Proceedings of EG Workshop on Animation and Simulation) / Edited by Edited by D. Thalman and M. van de Panne. – Wien: Springer-Verlag, 1996. – p. 61-76.

6. Braune, L. An initiation to SPH / L. Braune, T. Lewiner – Rio de Janeiro: PUC-Rio, 2013. – p. 1-7.

7. Sun F. Investigations of Boundary Treatments in Incompressible Smoothed Particle Hydrodynamics for Fluid-Structural Interactions / F. Sun, M. Tan, J. Xing // Recent Researches in Mechanics / Edited by N. Mastorakis – Corfu Island: WSEAS Press, 2011. – p. 92-97.

8. Muller M. Particle-Based Fluid Interaction / M. Muller, B. Solenthaler, R. Keiser, M. Gross // Eurographics/SIGGRAPH Symposium on Computer Animation / Edited by D. Breen and M. Lin. – Zurich: Federal Institute of Technology Zurich (ETHZ), 2005. – p. 1-7.

9. Randles P.W. Smoothed Particle Hydrodynamics: Somer recent improvements and applications / P.W. Randles, L.D. Libersky // Computer Methods in Applied Mechanics and Engineering – 1996 – № 139 – p. 375-408.

10. Janosi I.M. Turbulent drag reduction in dam-break flows / I.M. Janosi, D. Jan, K.G. Szabo, T. Tel // Experiments in Fluids – 2004 – № 37 – p. 219-229.

РАЗРАБОТКА POINT-OF-INTEREST РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ ПРИ ИСПОЛЬЗОВАНИИ LBSN СОЦИАЛЬНЫХ СЕТЕЙ

© Пахомова К.И.¹

Сибирский федеральный университет, г. Красноярск

В статье рассматривается проблема изучения жизненного цикла человека по средствам извлечения данных POI (Point-of-Interest) из LBSN (Location Based Social Network). На основе алгоритмов классификации и прогнозирования строится рекомендательная система, которая в свою очередь предугадывает поведение пользователя при смене локации в соответствии с его собственными привычками и традициями. На основе алгоритмов кластеризации строится модель рекомендательной системы, включающая временные, хронологические, географические и личные особенности пользователя.

Ключевые слова: POI (Point-of-Interest), LBSN (Location Based Social Network), рекомендательные системы, кластеризация, жизненный цикл человека, численный вероятностный анализ.

Жизнь современного человека весьма динамична, каждый день подобен циклу, а именно совершаемые процессы происходят последовательно один

¹ Аспирант кафедры Искусственного интеллекта.

за другим. Для того, чтобы организовать и дать точные рекомендации, необходимо определить эти циклы и извлечь именно интересную информацию из LBSN (Location Based Social Network). LBSN $\langle G, C \rangle$ состоит из графа социальной сети $G = \langle U, E \rangle$, где U – множество пользователей, $E = \{(u_i, u_j)\}$ отношения между пользователем u_i и u_j , $u_i, u_j \in U$, $u_i \neq u_j$ и множество, отмеченных локаций или комментарии $C = \{(u, l, t)\}$, где (u, l, t) представляет отмеченную пользователем u локацию l в определенное время t , $l \in L$. Локация l состоит из координат (x, y) – широта и долгота.

Для описания будущей модели рекомендательной системы, состоящей из нескольких компонентов, следует обратиться к определению LBSN [4-6].

Современные LBSMs позволяют распространять пользовательский и мультимедийный контент, совершать обмен сообщениями, определение текущей геолокации, с их помощью можно анализировать как собственную жизнь, так и жизнь, интересующих вас людей. Обычно, привычки людей не претерпевают изменений и человек сам того не осознавая существует по созданным им самим правилам, например, кафе-работа-магазины-дом. Практически у каждого существует аналогичный распорядок дня, а именно определенная подсознательно созданная жизненная модель. Чаще всего такая модель протекает в определенной локации и пользователю легко ориентироваться в своих предпочтениях, при этом затрачивая минимальное количество времени. При смене локации возникает смена правил и именно в этом случае проектируемая модель должна адаптироваться под текущее расположение и предложить пользователю аналогичные его предыдущим вкусам правила.

Цель данной работы – это разработать систему, которая бы помогала человеку адаптироваться за короткий промежуток времени, при этом не изменяя свои привычки. Исследовательская работа состоит из трех основных стадий: изучение и нормализация данных из социальных сетей, определение жизненной модели человека и разработка рекомендательной системы.

Для того, чтобы определить жизненную модель человека следует обратиться к теории графов и дефиниции LBSN, так как граф дает наглядное представление о любом циклическом процессе и представляет наглядное обоснование поставленной задачи.

Жизненная модель человека $G = (I, A)$ ориентированный граф, где $U = \{u_1, u_2, \dots, u_n\}$ – множество пользователей, $I = \{i_1, i_2, \dots, i_m\}$ множество интересов, и множество отмеченных пользователем $POI = \{UserID, Time, Location(x, y), Topic\}$, поэтому множество POI, посещенных пользователем u выражено как I_u .

Для построения жизненной модели человека были взяты данные из LBSN Foursquare, которой в настоящее время является одним из самых популярных социальных сетей, с ориентацией на локацию пользователя, в России данная социальная сеть действительна с 2010 года. После предвари-

тельной обработки данных с LBSN и опираясь выше изложенной определе- ние получаем граф жизненной модели человека, рисунок 1.

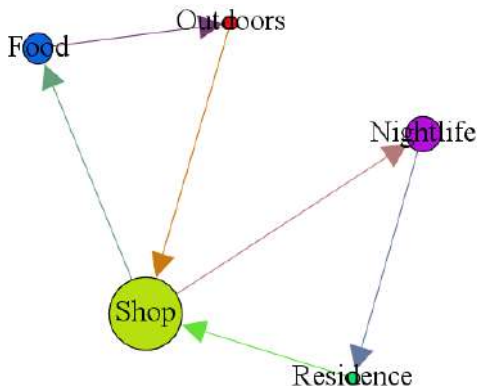


Рис. 1. Жизненная модель человека

На данном рисунке представлен жизненный цикл человека, где узлами являются интересы и ежедневные правила пользователя, именуемые как POI, а ребра показывают последовательность посещаемых мест. Данный граф иллюстрирует личные и распространённые интересы человека, классифицирует POI, а также сопоставляет и оценивает интересы текущего пользователя. Пользователь под номером 47 имеет 566 отмеченных им локаций, 27 июня 2010 года 9 отмеченных локаций составило 1.62 % из всего количества. Жизненная модель данного пользователя началась с 1:05:45 часов и этому времени соответствует POI – Nightlife (ночная жизнь). После 4:37:39 пользователь отметил POI как Residence (дом), которая имеет специальные координаты (x, y) ($x = 32.35869$, $y = -114.2258$) и собственное название (FIPS HQ). Далее в 14:08:15 перемещается в локацию магазин, в 15:02:49 Prospect Perk Café, тем самым цикл замыкается со стартовой локацией в 19:57:02 Boxers NYC Sports bar цикл замыкается. Размер узла графа отражает количество отмеченных POI, на рисунке 1 наблюдается градация POI от меньшего размера к большому, другими словами от часто посещаемого интереса, к менее посещаемому, например, в Июне 27 POI Shop соответствует 3 отметки – диаметр узла увеличенный, а POI Outdoors соответствует 1 отметка.

Помимо теории графов описать жизненный цикл человека возможно описать при помощи численного вероятностного анализа. В работах [1-3] используются метод гистограмм для детального изучения и анализа данных. На основе численного вероятностного анализа [1-3] предлагается концептуально-гистограммный подход, который применяется для разработки процедур представления и обработки информационных потоков, а также для численного моделирования и представления характеристик природных объек-

тов. Показывается, что применение разработанных процедур позволяет агрегировать данные, снижает уровень информационной неопределенности в данных и существенно повышает эффективность численных расчетов.

Рассматривая гистограмму как математический объект, который определяется как кусочно-постоянная функция P , определенная сеткой $\{z_i, i = 0, 1, \dots, n\}$, и на отрезке $[z_{i-1}, z_i]$ принимающая постоянное значение p_i , авторы определяют арифметические операции над гистограммами, операции вычисления максимума и минимума, возведения в степень [1-3], тем самым предлагая гистограммную арифметику над объектами, использующими как форму представления и исследования гистограмму.

Реализация арифметических операций над гистограммами основана на работе с $p(x, y)$ – совместной плотностью вероятности двух случайных величин x, y . Пусть p_z – гистограмма, приближающая плотность вероятности арифметической операции над двумя случайными величинами x^*y , где $*$ ∈ $\{+, -, \cdot, /, \uparrow\}$. Тогда вероятность попадания величины z в интервал $[z_i, z_{i+1}]$ определяется по формуле [1]

$$P(z_k < z < z_{k+1}) = \int_{\Omega_k} p(x, y) dx dy, \quad (1)$$

Основные принципы разработки гистограммных операций продемонстрируем на примере операции сложения. Пусть $z = x_1 + x_2$, и носители $x_1 - [a_1, a_2]$, $x_2 - [b_1, b_2]$, $p(x_1, x_2)$ – плотность распределения вероятностей случайного вектора (x_1, x_2) . Заметим, что прямоугольник $[a_1, a_2] \times [b_1, b_2]$ – носитель плотности распределения вероятностей $p(x_1, x_2)$ и плотность вероятности z отлична от нуля на интервале $[a_1 + b_1, a_2 + b_2]$. Обозначим $z_i, i = 0, 1, \dots, n$ – точки деления этого интервала на n отрезков. Тогда вероятность попадания величины z в интервал $[z_i, z_{i+1}]$ определяется по формуле

$$P(z_i < z < z_{i+1}) = \int_{\Omega_i} p(x_1, x_2) dx_1 dx_2,$$

где $\Omega_i = \{(x_1, x_2) \mid z_i \leq x_1 + x_2 \leq z_{i+1}\}$ [1]. И окончательно p_z имеет вид

$$p_{z_i} = \int_{\Omega_i} p(x_1, x_2) dx_1 dx_2 / (z_{i+1} - z_i).$$

Рассмотренный выше подход обобщается на случай большего числа переменных. Пусть требуется найти гистограмму p_z суммы

$$z = a_1 x_1 + a_2 x_2 + \dots + a_n x_n$$

и пусть $p(x_1, x_2, \dots, x_n)$ – плотность распределения вероятностей случайного вектора (x_1, x_2, \dots, x_n) . Тогда вероятность попадания z в интервал (z_i, z_{i+1}) соответственно равна [1]

$$P(z_i < z < z_{i+1}) = \int_{\Omega_i} p(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n,$$

где p_{z_i} имеет вид $p_{z_i} = \int_{\Omega_i} p(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n / (z_{i+1} - z_i)$.

Прогнозирование жизненной модели человека при изменении пользователя-тематикой локаций осуществляется при помощи алгоритма кластеризации GMM (Gaussian mixture model). Кластеризация способствует выявлению по координатам (x, y) текущей локаций пользователя и спрогнозировать POI при перемещении его на новое место и с новыми (x, y) . Данная модель представляет собой параметрическую функцию плотности вероятности как взвешенную сумму Гауссовых плотностей компонент. GMM, обычно используются в качестве параметрической модели закона распределения вероятностей непрерывных измерений или характеристики в биометрические системы. Параметры GMM оцениваются из обучающих данных с использованием итерационных Expectation-Maximization (EM), которая дает оценку из хорошо обученной априорной модели. На рисунке 2 показана гистограмма зависимости вероятности от времени суток, отмеченного пользователем POI, на вход гистограммы подавалась тестовая выборка. Красным пунктиром выделена функция, которая построена при помощи GMM алгоритма. Тем самым GMM неплохо справился с прогнозированием тестовой выборки, были выделены сходные группы POI объектов, в зависимости от времени суток. Стоит отметить, что результат кластеризации существенно зависит от метрики, выбор которой, как правило, также субъективен и определяется экспертом. Но стоит отметить, что есть ряд рекомендаций к выбору мер близости для различных задач.

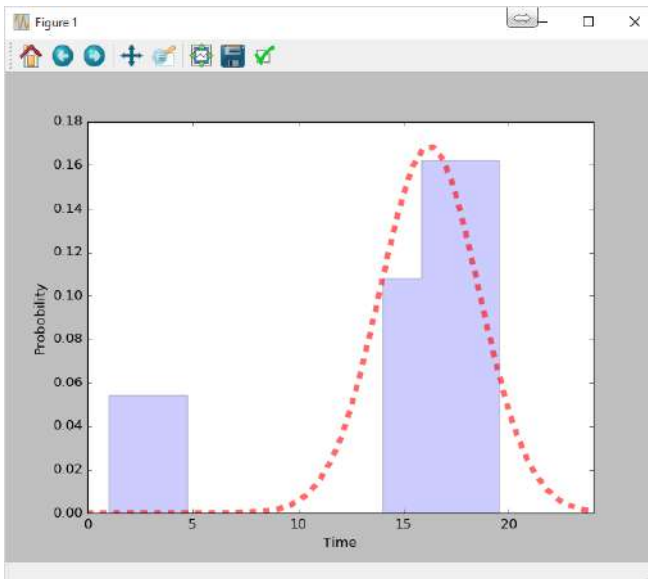


Рис. 2. Гистограммный анализ в совокупности с GMM

Список литературы:

1. Добронез Б.С., Попова О.А. Численные операции над случайными величинами и их приложения // Журн. СФУ. Сер. Матем. и физ. – 2011. – № 2. – С. 229-239.
2. Добронез Б.С., Попова О.А. Элементы численного вероятностного анализа // Вестник Сибирского государственного аэрокосмического университета. – 2012. – № 2 (42). – С. 19-23.
3. Добронез Б.С., Попова О.А. Гистограммный подход к представлению и обработке данных космического и наземного мониторинга // Известия Южного федерального университета. Технические науки. – 2014. – № 6 (155). – С. 14-22.
4. Yu Y., Chen X. A Survey of Point-of-Interest Recommendation in Local-Based Social Networks. Association for the Advancement of Artificial Intelligence, 2015.
5. Gao H. Personalized POI Recommendation on Location-Based Social Networks. Arizona state university, 2014.
6. Gao H., Tang J., Hu X., Liu H. Content-Aware Point of Interest Recommendation on Location-Based Social Networks. Association for the Advancement of Artificial Intelligence, 2015.