Inferring Person-to-person Proximity Using WiFi Signals

Today’s societies are enveloped in an ever-growing telecommunication infrastructure. This infrastructure offers important opportunities for sensing and recording a multitude of human behaviors. Human mobility patterns are a prominent example of such a behavior which has been studied based on cell phone towers, Bluetooth beacons, and WiFi networks as proxies for location. However, while mobility is an important aspect of human behavior, understanding complex social systems requires studying not only the movement of individuals, but also their interactions. Sensing social interactions on a large scale is a technical challenge and many commonly used approaches—including RFID badges or Bluetooth scanning—offer only limited scalability. Here we show that it is possible, in a scalable and robust way, to accurately infer person-to-person physical proximity from the lists of WiFi access points measured by smartphones carried by the two individuals. Based on a longitudinal dataset of approximately 800 participants with ground-truth interactions collected over a year, we show that our model performs better than the current state-of-the-art. Our results demonstrate the value of WiFi signals in social sensing as well as potential threats to privacy that they imply.
We test the performance of our allocation strategy using real data from over 600 peer feedback sessions and simulate the
effects of different allocation strategies. By comparing our method with a random allocation algorithm and a “super-
informed oracle” algorithm we demonstrate that we are able to allocate reviewers to submissions in such a way that all
submissions receive feedback of similar quality and that we are able to significantly outperform simple random allocation
of reviewers. Additionally we investigate the effect of pre-allocating reviews in comparison to allocating reviewers live
during the review process and show that live-allocation leads to better results. Our method is robust to reviews not being
completed and other real-life quirks and improves as more feedback data is collected.

General information
State: Published
Organisations: Department of Applied Mathematics and Computer Science, Cognitive Systems, Peergrade, Københavns
Universitet
Authors: Wind, D. K. (Intern), Jensen, U. A. (Ekstern), Jørgensen, R. M. (Ekstern), Hansen, S. L. (Ekstern), Winther, O.
(Intern)
Publication date: 2017

Host publication information
Title of host publication: Proceedings of 16th European Conference on eLearning
Main Research Area: Technical/natural sciences
Conference: 16th European Conference on eLearning, Porto, Portugal, 26/10/2017 - 26/10/2017
peer assessment, peer feedback, feedback, peer review, peer evaluation, peer grading, task assignment, reviewer
allocation

Bibliographical note
For ECEL2017 http://www.academic-conferences.org/conferences/ecel/
Source: PublicationPreSubmission
Source-ID: 138400287
Publication: Research - peer-review > Article in proceedings – Annual report year: 2017

Quantifying Feedback – Insights Into Peer Assessment Data
The act of producing content - for example in forms of written reports - is one of the most used methods for teaching and
learning all the way from primary school to university. It is a learning tool which helps students relate their theories to
practice. Getting relevant and helpful feedback on this work is important to ensure a good learning experience for the
students. Providing this feedback is often a time-consuming job for the teacher. An effective way to learn is to teach
others, and similarly give feedback on work done by others. One way to approach a combined solution to the above
challenges, is to use peer assessment in the classroom which as a learning method has become more and more popular.
In this paper we look at data collected using the web-based peer assessment system Peergrade. The dataset consists of
over 350 courses at more than 20 educational institutions and with a total of more than 10,000 students. The students
have together made more than 100,000 peer-evaluations of work by other students, and these evaluations together
contain more than 10,000,000 words of text feedback. A key problem when using peer assessment is to ensure high
quality feedback between peers. Feedback here can be a combination of quantitative / summative feedback (numerical)
and qualitative / formative feedback (text). A lot of work has been done on validating and ensuring quality of quantitative
feedback. We propose a way to let students evaluate the quality of the feedback they receive to obtain a quality measure
for the feedback. We investigate this measure of feedback quality, which biases are present and what trends can be
observed across the dataset. Using our measure of feedback quality, we investigate how it relates to various factors like
the length of the feedback text, the number of spelling mistakes, how positive it is and measures of the student's report-
writing skills.

General information
State: Published
Organisations: Department of Applied Mathematics and Computer Science, Cognitive Systems, Københavns Universitet
Authors: Wind, D. K. (Intern), Jensen, U. A. (Forskerdatabase)
Number of pages: 10
Publication date: 2017

Host publication information
Title of host publication: Proceedings of the 12th International Conference on e-Learning
Main Research Area: Technical/natural sciences
Conference: 12th International Conference on e-Learning, Orlando, United States, 01/06/2017 - 01/06/2017
Peer assessment, Peer feedback, Feedback, Peer review, Peer evaluation, Peer grading
Source: PublicationPreSubmission
Source-ID: 138400219
Publication: Research - peer-review > Article in proceedings – Annual report year: 2017
Inferring Stop-Locations from WiFi

Human mobility patterns are inherently complex. In terms of understanding these patterns, the process of converting raw data into series of stop-locations and transitions is an important first step which greatly reduces the volume of data, thus simplifying the subsequent analyses. Previous research into the mobility of individuals has focused on inferring ‘stop locations’ (places of stationarity) from GPS or CDR data, or on detection of state (static/active). In this paper we bridge the gap between the two approaches: we introduce methods for detecting both mobility state and stop-locations. In addition, our methods are based exclusively on WiFi data. We study two months of WiFi data collected every two minutes by a smartphone, and infer stop-locations in the form of labelled time-intervals. For this purpose, we investigate two algorithms, both of which scale to large datasets: a greedy approach to select the most important routers and one which uses a density-based clustering algorithm to detect router fingerprints. We validate our results using participants’ GPS data as well as ground truth data collected during a two month period.
Model Selection in Data Analysis Competitions

The use of data analysis competitions for selecting the most appropriate model for a problem is a recent innovation in the field of predictive machine learning. Two of the most well-known examples of this trend was the Netflix Competition and recently the competitions hosted on the online platform Kaggle. In this paper, we will state and try to verify a set of qualitative hypotheses about predictive modelling, both in general and in the scope of data analysis competitions. To verify our hypotheses we will look at previous competitions and their outcomes, use qualitative interviews with top performers from Kaggle and use previous personal experiences from competing in Kaggle competitions. The stated hypotheses about feature engineering, ensembling, overfitting, model complexity and evaluation metrics give indications and guidelines on how to select a proper model for performing well in a competition on Kaggle.
Link prediction in weighted networks
Many complex networks feature relations with weight information. Some models utilize this information while others ignore the weight information when inferring the structure. In this paper we investigate if edge-weights when modeling real networks, carry important information about the network structure. We compare five prominent models by their ability to predict links both in the presence and absence of weight information. In addition we quantify the models ability to account for the edge-weight information. We find that the complex models generally outperform simpler models when the task is to infer presence of edges, but that simpler models are better at inferring the actual weights.

General information
State: Published
Organisations: Department of Applied Mathematics and Computer Science, Cognitive Systems, Department of Informatics and Mathematical Modeling, Cognitive Systems
Authors: Wind, D. K. (Intern), Mørup, M. (Intern)
Number of pages: 6
Publication date: 2012

Host publication information
Title of host publication: 2012 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)
Place of publication: 978-1-4673-1025-3
Publisher: IEEE
ISBN (Print): 978-1-4673-1024-6

Series: Machine Learning for Signal Processing
ISSN: 1551-2541
Main Research Area: Technical/natural sciences
Complex networks, Weighted graphs, Stochastic Blockmodels, Non-negative Matrix Factorization, Link-Prediction
DOIs: 10.1109/MLSP.2012.6349745
Publication: Research - peer-review › Article in proceedings – Annual report year: 2012

String matching with variable length gaps
We consider string matching with variable length gaps. Given a string T and a pattern P consisting of strings separated by variable length gaps (arbitrary strings of length in a specified range), the problem is to find all ending positions of substrings in T that match P. This problem is a basic primitive in computational biology applications. Let m and n be the lengths of P and T, respectively, and let k be the number of strings in P. We present a new algorithm achieving time \(O(n\log k + m + a)\) and space \(O(m + A)\), where \(A\) is the sum of the lower bounds of the lengths of the gaps in P and \(a\) is the total number of occurrences of the strings in P within T. Compared to the previous results this bound essentially achieves the best known time and space complexities simultaneously. Consequently, our algorithm obtains the best known bounds for almost all combinations of \(m, n, k, A,\) and \(a\). Our algorithm is surprisingly simple and straightforward to implement. We also present algorithms for finding and encoding the positions of all strings in P for every match of the pattern.

General information
State: Published
Organisations: Department of Informatics and Mathematical Modeling, Algorithms and Logic, Computer Science and Engineering, Department of Applied Mathematics and Computer Science, Cognitive Systems
Authors: Bille, P. (Intern), Gørtz, I. L. (Intern), Vildhøj, H. W. (Intern), Wind, D. K. (Intern)
Pages: 25-34
Publication date: 2012
Main Research Area: Technical/natural sciences

Publication information
Journal: Theoretical Computer Science
Volume: 443
ISSN (Print): 0304-3975
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Statistical Models for Temporal Dynamics in Complex Networks
Department of Applied Mathematics and Computer Science
Period: 15/05/2014 → 18/07/2018
Number of participants: 3
PhD Student:
Wind, David Kofoed (Intern)
Supervisor:
Jørgensen, Sune Lehmann (Intern)
Main Supervisor:
Winther, Ole (Intern)

Financing sources
Source: Internal funding (public)
Name of research programme: Forskningsrådsfinansiering
Project: PhD