Reproduce “Private traits and attributes are predictable from digital records of human behavior”

JING CHEN, University of Pittsburgh, Pittsburgh, PA, jic118@pitt.edu  
EASON, University of Pittsburgh, Pittsburgh, PA, \*\*\*@pitt.edu

KEYWORDS

data mining, private traits, facebook likes, SVD, machine learning

Abstract

The work we will do is reproducing research results in the paper “Private traits and attributes are predictable from digital records of human behavior” [1].

1 INTRODUCTION

The motivations we try to reproduce the result of this paper can be conclude as follow. First, following the paper instruction and reproducing the whole research process allows us to verify its result, whether the result is reproducible by the procedures given in the paper, how the difference of dataset can affect the result and wthether the methods they implemented in this paper is reasonable. Second, by reproducing the study, we can put data mining theory and technical skills learned in class into practice. Also, we will try to find whether we can make some adjustments that can improve the result and provide a detailed discussion on the adjustments. We believe this work can give us a deeper understanding of knowledge in this field. Last, this problem focus by the paper is meaningful and profound, nowadays, we are exposed to social media environment, there are a huge number of easily accessible digital records of behavior on such social media, which can be used to automatically and accurately predict a range of highly sensitive personal attributes. Data mining technique skills like SVD, linear regression model, logistic regression model, cross-validation, ROC, AUC, and Pearson product-moment correlation had been used in our work.

2 DATASETS

The dataset used in our work download from the myPersonality Project[[1]](#footnote-2) database. This dataset includes three csv files, users (110K), likes (1.58 million), and userslikes (10 million). Users has the information about 110K users, including userid, gender, age, political parties, openness, conscientiousness, extraversion, agreeableness, and emotional stability. Likes has two columns, likeid and name. Users-likes has two columns, userid and likeid.

3 FINISHED WORK

The first step is to preprocess users-likes datasets, according to the support information of this paper, that requires to remove Likes which associate with fewer than 20 users, as well as users with fewer than two Likes. After doing that, we got 105,262 users and 54,457 Likes. Then we created a sparse matrix (105262\*54457) which generated by a pakage called Matrix in R. This sparse matrix only has a small size of 87Mb (the dense matrix about 40Gb).

Next, dimensionality reduction technique SVD has been implemented to the matrix and got two matrixes (one has 100 top SVD components, one has 30 compoents).

For the next phase, we do cross validation on the dataset, in which, logistic regression model has been used for the dichotomous variables, linear regression model for the numerical variables.

2.3 Dynamic Measurements: BLS

BLS spectra of the thermal magnetic excitations were measured at room temperature in the back-scattering geometry by using a (3+3)-pass tandem Fabry–Perot interferometer. About 200 mW of monochromatic ( = 532 nm) laser were focused onto a spot having a diameter of about 30 microns. An external magnetic field with intensity ranging between −1.0 kOe < *H* < +1.0 kOe was applied parallel to the sample surface along the dots length and perpendicular to the incidence plane of light (Voigt geometry).

2.4 Ground-State Magnetization Determination and DMM Micromagnetic Simulations

*2.4.1 Determined.* The magnetization ground-states as well as the hysteresis loops were determined by using the OOMMF code. To reproduce the exact shape of the dots, a bitmap image of the basic unit of the bi-component dots was created from the SEM image of [Fig. 1](#fig1), and used as input for the simulations. Periodic boundary conditions have been applied to account for the chain arrangement of the Py/Co dots in the investigated sample.

*2.4.2 Micromagnetic.*For each micromagnetic cell the reduced magnetization takes the form where themagnetization (saturation magnetization) in the *k-*th cell; note that the saturation magnetization now depends on the ferromagnetic material through the index *k*. Hence, in a polar reference frame

(1)

where is the azimuthal (polar) angle of the magnetization (the time dependence is omitted). The second derivatives of the energy density depend on the micromagnetic cell indexes, and through them on the material index corresponding either to Py or Co. The expressions of *E*ext, *E*exch*, E*dmg and *E*ani are the same as the ones of the single-component system apart from the explicit dependence of the magnetic parameters on the given ferromagnetic material. Moreover, the uniaxial anisotropy energy density of Co is neglected.

It is possible to write the following periodicity rule valid for the dynamic magnetization of each collective mode, a version of the Bloch theorem, viz. Note that, exchange contribution is set equal to zero, because in each unit cell the two elliptical dots are separated. Moreover, the uniaxial anisotropy energy density of Co is neglected [Table 1](#tb1).

Table 1. Frequency of Special Characters

|  |  |  |
| --- | --- | --- |
| Non-English or Math | Frequency | Comments |
| Ø | 1 in 1,000 | For Swedish names |
| $ | 4 in 5 | Used in business |

(2)

Therefore one can observe either an in-phase (acoustic) or an out-of-phase (optical) character of the modes, with respect to the precession of the in-plane magnetization components in adjacent Py and Co dots.

We would like to mention that the DMM presents several advantages with respect to OOMMF for calculating the spectrum of magnetic eigenmodes for the following reasons: *a*) There is no need to excite the system by any magnetic field pulse, *b)* A single calculation llows to determine the frequencies and eigenvectors of all spin-wave modes of any symmetry, *c*) The spectrum is computed directly in the frequency domain, *d*) The mode degeneracy is successfully solved, *e*) The spatial profiles of the spin-wave modes are directly determined as eigenvectors and, finally, *f*) The differential scattering cross-section can be calculated accurately from the eigenvectors associated to each spin-wave mode. This is a clear indication that both the Py and Co sub-elements are in a single domain state where Py and Co magnetizations are all oriented with their magnetic moment along the chain and field direction. At point ** (*H* = −372 Oe) of the hysteresis loop, where the plateau is observed in the *M*-*H* loop, the dark and bright spots of the Py dots are reversed with respect to those of Co, accounting for an antiparallel relative alignment of magnetization.

3 RESULTS AND DISCUSSION

3.1 Magnetization Curves and MFM Characterization

The major hysteresis loop measured by MOKE, plotted in [Fig. 1](#fig1), displays a two-step switching process due to the distinct magnetization reversal of the Py and Co sub-elements, characterized by a different coercivity. As the field is reduced from positive saturation (upper branch of the M-H loop), a 100% remanence is attained. Within each bi-component unit (about 36%) in good agreement with experimental result (about 40%).

To directly visualize the evolution of the magnetization in the Py and Co subunits of our bi-component dots during the reversal process, we performed a field-dependent MFM analysis whose main results are reported in [Fig. 2](#fig2). At large positive field (*H*= +800 Oe, not shown here) and at remanence (** point of the hysteresis loop of [Fig. 1](#fig1)), the structures are characterized by a strong dipolar contrast due to the stray fields emanated from both the Py and Co dots.



Fig. 2. MFM images of the bi-component Py/Co dots for different values of the applied magnetic field which are indicated by greek letters along both the major and minor hysteresis loop.

This is a clear indication that both the Py and Co sub-elements are in a single domain state where Py and Co magnetizations are all oriented with their magnetic moment along the chain and field direction. At point ** (*H*= −372 Oe) of the hysteresis loop, where the plateau is observed in the *M*-*H* loop, the dark and bright spots of the Py dots are reversed with respect to those of Co, accounting for an antiparallel relative alignment of magnetization.

At relatively large negative fields (point ** *H*= −770 Oe) the magnetization reversal is completed and the magnetization of the two adjacent sub-elements are saturated in the negative direction. The ground state remains unchanged when the field is now reduced to zero, i.e. remanent state coming from negative saturation, as confirmed by the MFM image taken at point  of [Fig. 1.](#fig1)

We have also used MFM to measure the magnetic configurations along the minor hysteresis loop, described above. Once the AP ground state has been generated at *H*= −500 Oe, the applied field is increased in the positive direction. The MFM image taken at point **' of [Fig. 2](#fig2), remanent state of the minor loop (*H* = 0), shows that the AP state is stable and remains unchanged until the magnetic field is increased up to +300 Oe where the Py magnetization reverses its orientation and returns to be aligned with that of Co dots. On the basis of the above MFM investigation, one can say that the structures are always in a single domain state, while the relative magnetization orientation between the adjacent Py and Co elements depends on both the field value and the sample history.

3.2 Field Dependent BLS Measurements and DMM Calculations

[Fig. 3](#fig3) displays the frequencies of BLS peaks plotted as a function of the applied field magnitude starting from positive values. The field is then decreased and reversed following the upper branch of the hysteresis loop, shown in the same figure. Up to five peaks are measured in the spectra, as shown in spectrum measured at H = 0 Oe in the [Fig. 3](#fig3) inset, and their field evolution analyzed over the whole field range investigated. The detected modes are identified and labeled on the basis of their calculated spatial profiles, shown in [Fig. 4](#fig4) for *H*= 500 and −500 Oe.

Dependence of the magnetic eigeinmode wave frequency on the applied field strength.

Fig. 3. Dependence of the magnetic eigeinmode wave frequency on the applied field strength.



Fig. 4. Calculated spatial distribution of the in-plane dynamic magnetization.

They exhibit marked localization into either the Co or the Py dots, as stated at the end of the previous Section, were it was introduced the labelling notation containing the dominant localization region (either Py or Co) and the spatial symmetry (EM, F, DE, etc).

When the dots are in the P state, up to five modes were detected in BLS spectra. On the basis of the calculated profiles (right panel of [Fig. 4](#fig4)), we identified in the P state the two modes at lowest frequencies as the EM(Py) and the F(Py), with a very small spin precession amplitude into the Co dot. This is because for this material we are below the frequency threshold for the existence of spin waves. A similar effect has been observed in periodic array of alternating Permalloy and Co nanostripes

Note that the nodal lines present in the spatial profile of the F (Co) mode perpendicular to the long axis of the ellipse do not correspond to a real change of sign of the dynamic magnetization and are due to the partial hybridization of the F mode with higher-order modes having frequencies close to the one of the F mode. Interestingly, the frequency slope of modes localized into the Co dots is larger than that of Py modes, due to larger values of the Co magnetization and gyromagnetic ratio. An overall good agreement between the calculated (dotted curves) and measured frequency (full points) has been achieved (see [Fig. 3](#fig3)) even if some discrepancies are observed for the frequency of the EM and 1DE (Py) modes.

The corresponding spatial profiles of the modes are shown in the left panels of [Fig. 4](#fig4). Here one can see that the only mode which is purely localized in one dot is the EM of Co, because now it is sub-threshold for Py. A further reduction of *H*, which is sufficient to cause the Co magnetization reversal, produces a P state at negative fields and the frequency starts to increase.



Fig. 5. Full point are the frequencies measured along the minor hysteresis.

Again as a function of the applied field. In this field range the frequencies of modes in the Py dots monotonously increase in a way similar to that measured in the P state for positive field values while an abrupt change in the frequency of Co modes occurs.

Notice that if one stops increasing the negative field to about −300 Oe and comes back towards positive applied fields, BLS measurements can be performed following the minor hysteresis loop. This method permits to study, for example, the magnetization dynamics at remanence (without any external applied magnetic field) when the system is in the AP state (see MFM image ' in [Fig. 2](#fig2)), a configuration which cannot be achieved at remanence along the major M-H loop. In [Fig. 5](#fig5) we show the modes frequency measured along the minor loop (full points) and compare them with values measured along the major M-H loop (open points).

By inspection of the frequency slope of the modes, one can immediately understand the localization of modes into dots of different materials looking at their slope.

In particular, for three (two) modes we measure a negative (positive) frequency slope with an almost linear dependence on *H*. It is evident that modes with negative frequency slope are modes localized into the Py dot (EM, F and 1DE) while the two with positive slope are the F(Co) and the EM(Co) modes.

3.3 Analysis of the Dynamic Coupling as a Function of the Gap Size

One interesting point which emerges from analysis of [Figs. 3](#fig3) and [4](#fig4) is that the frequency values of the eigenmodes are not the same at +500 Oe and at −500 Oe. This is expected for modes localized into the Co elements, since the external field is either parallel or antiparallel to their magnetization. However, for those mode localized into the Py sub-element. One could have predicted to find the same frequency values at ±500 Oe, unless the dipolar coupling arising from the adjacent Co dot plays a significant role. In fact, as seen in [Figs. 3](#fig3) and [4](#fig4), reversing the field from +500 to −500 Oe, the frequencies of EM(Py) and 1DE(Py) modes increase by about 0.2 GHz and 0.6 GHz, respectively, while that of F(Py) decreases by 0.25 GHz. The reason of this complex behavior will be addressed in the following, analyzing the interplay of both static and dynamic dipolar coupling between the adjacent Py and Co dots [Table 2](#tb2).

This is a clear indication that both the Py and Co sub-elements are in a single domain state where Py and Co magnetizations are all oriented with their magnetic moment along the chain and field direction.

Table 2. Comparison of Coefficients from Atomistic

|  |  |  |
| --- | --- | --- |
| Atm | MS-CG | MS-CG/DPD |
| 1.78 | 14.32 | 1.74 (−2%) |
| 0.43 | 31.00 | 0.40 (−7%) |
| 0.062 | 15.61 | 0.048 (−23%) |
| 0.032 | 9.76 | 0.024 (−24%) |
| 0.020 | 4.66 | 0.015 (−25%) |
| 0.012 | 2.32 | -"- |
| 0.0076 | 0.016 | -"- |



Fig. 6. Calculated frequency evolution of modes detected in the BLS spectra.

In [Fig. 6](#fig6) the calculated frequencies of the most representative eigenmodes at +500 Oe (FM state) and – 500 Oe (AP state) are plotted as a function of the gap size d between the Py and Co sub units (please remind that in the real sample studied here, d = 35 nm). As a general comment, it can be seen that the frequencies for the system in the AP state are more sensitive to d than those of the P state. In particular, the lowest three frequency modes of the AP state (EM(Co), EM(Py) and F(Py)) are downshifted with respect to the case of isolated elements (dotted lines) and show a marked decrease with reducing d, while the two modes at higher frequencies (F(Co) and 1DE(Py)) have an opposite behavior even though they exhibit a reduced amplitude. In the P state (right panel), the modes concentrated into the Py dots exhibit a moderate decrease with reducing d, while an opposite but less pronounced behavior is exhibited by the F(Co) mode.

* 1. Never, ever use vertical rules.
* 2. Never use double rules.

4 CONCLUSIONS

In summary, we have performed both an experimental and theoretical study of the spin eigenmodes in dipolarly coupled bi-component cobalt and permalloy elliptical nanodots. Several eigenmodes have been identified and their frequency evolution as a function of the intensity of the applied magnetic field has been measured by Brillouin light scattering technique, encompassing the ground states where the cobalt and permalloy dots magnetizations are parallel or anti-parallel, respectively. In correspondence to the transition between the two different ground states, the mode frequency undergoes an abrupt variation and more than that, in the anti-parallelstate, the frequency is insensitive to the applied field strength. The experimental results have been successfully interpreted by the dynamic matrix method which permits to calculate both the mode frequencies and the spatial profiles.

A HEADINGS IN APPENDICES

The rules about hierarchical headings discussed above for the body of the article are di.erent in the appendices. In the appendix environment, the command section is used to indicate the start of each Appendix, with alphabetic order designation (i.e., the first is A, the second B, etc.) and a title (if you include one). So, if you need hierarchical structure within an Appendix, start with subsection as the highest level. Here is an outline of the body of this document in Appendix-appropriate form:

A.1 Introduction

A.2 Experimental and Computational Details

A.2.1 Sample Fabrication

A.2.2 Quasi-Static Measurements: MOKE and MFM

Component Structures

Magnetization.

A.2.3 Dynamic Measurements: BLS

A.2.4 Ground-State Magnetization Determination and DMM Micromagnetic Simulations

Determined.

Micromagnetic

A.3 Results and Discussion

A.3.1 Magnetization Curves and MFM Characterization

A.3.2 Field Dependent BLS Measurements and DMM Calculations

A.3.3 Analysis of the Dynamic Coupling as a Function of the Gap Size

A.4 Conclusions

A.5 References

ACKNOWLEDGMENTS

This work was partially supported by the MIUR-PRIN 2010–11 Project 2010ECA8P3 “DyNanoMag” and by the National Research Foundation, Prime Minister's office, Singapore under its Competitive Research Programme (CRP Award No. NRF-CRP 10-2012-03).

REFERENCES

[1] Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. Proceedings of the National Academy of Sciences, 110(15), 5802-5805.

[2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. 2002. Wireless Sensor Networks: A Survey. *Comm. ACM* 38, 4 (2002), 393–422.

[3] David A. Anisi. 2003. *Optimal Motion Control of a Ground Vehicle*. Master’s thesis. Royal Institute of Technology (KTH), Stockholm, Sweden.

[4] P. Bahl, R. Chancre, and J. Dungeon. 2004. SSCH: Slotted Seeded Channel Hopping for Capacity Improvement in IEEE 802.11 Ad-Hoc Wireless Networks. In *Proceeding of the 10th International Conference on Mobile Computing and Networking* (MobiCom’04). ACM, New York, NY, 112–117.

[5] Kenneth L. Clarkson. 1985. Algorithms for Closest-Point Problems (Computational Geometry). Ph.D. Dissertation. Stanford University, Palo Alto, CA. UMI Order Number: AAT 8506171.

[6] Jacques Cohen (Ed.). 1996. Special Issue: Digital Libraries. *Commun. ACM* 39, 11 (Nov. 1996).

[7] Bruce P. Douglass. 1998. Statecarts in use: structured analysis and object-orientation. In *Lectures on Embedded Systems*, Grzegorz Rozenberg and Frits W. Vaandrager (Eds.). Lecture Notes in Computer Science, Vol. 1494. Springer-Verlag, London, 368–394. DOI: http://dx.doi.org/10.1007/3-540-65193-429

[8] Ian Editor (Ed.). 2008. *The title of book two* (2nd. ed.). University of Chicago Press, Chicago, Chapter 100. DOI: http://dx.doi.org/10.1007/3-540-09237-4

Journal Name:ACM TRANS. GRAPH.

Volume:9

Issue:4

Year:2010

Month:March

Article Seq:39

DOI:10.1145/1234

Copyright Year:2017

Copyright Statement:rightsretained

ISSN:0730-0301

Price:$15.00

1. Data deposition: The data reported in this paper have been deposited in the myPersonality Project database (www.mypersonality.org/wiki). [↑](#footnote-ref-2)